



# **UNIVERSITÀ' DEGLI STUDI DI TRIESTE**

## **XXVII CICLO DEL DOTTORATO DI RICERCA IN**

**SCIENZE DELL'INTERPRETAZIONE E DELLA TRADUZIONE**

### **VOICE QUALITY AND TV INTERPRETING: A PROPOSAL FOR A GESTALTIC EVALUATION**

Settore scientifico-disciplinare: L-LIN/07 LINGUA E TRADUZIONE – LINGUA SPAGNOLA

**DOTTORANDO  
GREGORIO DE GREGORIS**

**COORDINATRICE  
PROF. FEDERICA SCARPA**

**SUPERVISORA DI TESI  
PROF. HELENA LOZANO MIRALLES**

**CO-SUPERVISORA DI TESI  
PROF. ALESSANDRA RICCARDI**

**ANNO ACCADEMICO 2014 / 2015**



# VOICE QUALITY AND TV INTERPRETING: A PROPOSAL FOR A GESTALTIC EVALUATION

---

## INDEX

### INTRODUCTION

History and object of the research. Rationale, objective and hypothesis 17

### CHAPTER 1

#### LITERATURE REVIEW – QUALITY AND PROSODY IN SIMULTANEOUS INTERPRETATION

1.0 Introduction	25
1.1 Quality expectations and quality assessment in simultaneous interpreting	25
1.1.1 Questionnaire-based surveys on quality expectations in simultaneous interpreting	26
1.1.1.1 Bühler (1986)	26
1.1.1.2 Kurz (1989; 1993)	27
1.1.1.3 Kurz and Pöchhacker (1995)	28
1.1.1.4 Pöchhacker and Zwischenberger (2010)	29
1.1.1.5 Meak (1990)	30
1.1.1.6 Chiaro and Nocella (2004)	31
1.1.1.7 Kopczyński (1994)	35
1.1.1.8 Moser (1995)	36
1.1.2 Quality assessment questionnaire-based surveys in simultaneous interpreting	39
1.1.2.1 Gile (1990)	40
1.1.2.2 Ng (1992)	41

1.1.2.3 Marrone (1993)	42
1.1.2.4 Vuorikoski (1993)	43
1.1.2.5 Mack and Cattaruzza (1995)	44
1.1.2.6 Doerflinger (1993)	45
1.1.2.7 Garzone (2003)	46
1.1.2.8 Russo (2005)	48
1.1.2.9 Catana (2005)	49
1.1.2.10 Collados Aís et al. (2007)	51
1.1.2.11 García Becerra (2013)	53
1.1.3 General considerations on the results of reviewed studies on SI quality expectations and quality assessment.	54
1.2 Conclusion: the need for a new paradigm of quality assessment based on gestalt perception of SI	58
1.3 Approaches to the concept of quality in simultaneous interpretation	58
1.4 Prosody and simultaneous interpretation	63
1.4.1 Cognitive rhythm and speech production	63
1.4.2 Simultaneous interpreting prosody	66
1.4.3 Structure and rhythm in speech	70

## **CHAPTER 2**

### **ORENESIT CORPUS**

2.0 Introduction	73
2.1 The corpus of reference: CorIT ( <i>Corpus Italiano di Interpretazione Televisiva</i> )	75
2.1.1 Transcription conventions	76
2.2 The thematic corpus: ORenesit (Obama vs. Romney English español italiano) – the 2012 US Presidential Debate	81
2.2.1 Digital archive	81
2.2.2 Transcription of the corpus <i>ORenesit</i>	81
2.2.3 The corpus <i>ORenesit</i>	86
2.2.4 Corpus-based analyses	86
2.2.4.1 Analysis of rendition of face-threatening speech acts (Babetto 2013)	86
2.2.4.2 Analysis of strategies for the translation of culture-bound elements (Tomasetig 2014)	88
2.2.4.3 Analysis of self-repairs (Penso 2014)	89

2.2.5 TV interpreters, situation and working conditions (RTVE and RAInews24)	90
2.3 Conclusion	93

### **CHAPTER 3**

## **THEORETICAL FRAMEWORK FOR A NEW QUESTIONNAIRE FOR A GESTALTIC ASSESSMENT OF TV BROADCAST SIMULTANEOUS INTERPRETATION**

3.0. Introduction: prosody, voice and language	97
3.0.1 Prosody	97
3.0.2 Voice	100
3.0.3 Linguistic vs. paralinguistic (segmental vs. suprasegmental)	104
3.1. The gestaltic perception and production of speech: Bühler's <i>Theory of language</i> .	107
3.1.1 The axiom system	107
3.1.2 Axiom A: The model of language as Organon	109
3.1.3 Axiom B: The significative nature of language	111
3.1.4 Axiom C: Speech action and language work; speech act and language structure	113
3.1.5 Axiom D: Word and sentence. The S-F-System of the Type Language	114
3.1.6 The deictic field and the symbolic field of language	114
3.1.7 <i>Gestalt</i> and speech	119
3.1.8 Voice and expression	122
3.2 Rhythm as form (Benveniste 1966)	124
3.3 Rhythm as sense of discourse (Meschonnic 1982)	126
3.4 The acoustic face of words (Albano Leoni 2009)	129
3.4.1 Brief history of the phoneme	130
3.4.1.1 Pre-history	130
3.4.1.2 The psycholinguistic stage	130
3.4.1.3 Structuralism	131
3.4.1.4 The generativist and post-generativist stage	133
3.4.1.5 Cognitive phonologies	135
3.4.2 Conclusion	136
3.4.3 Criticism of segmental phonologies	137

3.4.4 The proposal for the acoustic face of words	139
3.5 Voice, audio-vision, television	143
3.5.1 The perception of <i>sound on screen</i> (Chion 1990/1994)	143
3.6 Voice as gesture: expression and perception (Fónagy 1983)	146
3.6.1 Voice as gesture	147
3.6.2 Prosody	147
3.6.3 Attitudes in voice	148
3.6.4 Vocal personality	150
3.6.5 Poetry and vocal art	156
3.7 Constancy of expressive traits and their symbolization	157
3.8 Proposal for a questionnaire to elicit a subjective gestaltic assessment of voice quality in television broadcast simultaneous interpreting	158
3.8.1. Deep structure of the questionnaire: categories related to the sources of the theoretical paradigm	158
3.8.2. Superficial structure of the questionnaire: items related to the properties of <i>Il volto fonico delle parole</i> (Albano Leoni 2009)	160

**CHAPTER 4**  
**PILOT SURVEY ON VOICE QUALITY OF A TV BROADCAST**  
**SIMULTANEOUS INTERPRETATION ASSESSMENT**  
**THROUGH GESTALT PERCEPTION**

4.0 Introduction	163
4.0.1 Television Interpreting and Film Interpreting	163
4.0.2 TV information, voice and TV broadcast simultaneous interpretation	165
4.0.3 Radiophonic information and credible voice	172
4.1 Rationale	175
4.2 Objective	176
4.3 Hypothesis	177
4.4 Method	177
4.5 Materials	178
4.5.1 Questionnaire	178
4.5.1.1 Process of construction of the questionnaire based on the development of the theoretical framework	181
4.5.1.1.1 First draft	181

4.5.1.1.2 Second draft	182
4.5.1.1.3 Third draft (pre-test)	183
4.5.1.1.4 Fourth and fifth draft	184
4.5.1.1.5 Sixth draft and on-line version	184
4.5.1.2 Results from pre-test	185
4.5.2 Video excerpts of simultaneous interpretations for the pilot survey	191
4.5.2.1 Selection of three video excerpts	191
4.5.2.2 Experimental variable on video 2	193
4.5.2.3 Analysis of video excerpts	197
4.5.2.3.1 Transcripts	197
4.5.2.3.2 Identification of linguistic units and indices of analysis	197
4.5.2.3.3 Video 1 (compared to videos 2 and 3)	203
4.5.2.3.4 Video 2 (original vs. experimental variable)	205
4.5.2.3.5 Video 2 (the performance of the dubbing actor)	207
4.5.2.3.6 Video 2 (compared to videos 1 and 3)	212
4.5.2.3.7 Video 3 (compared to videos 1 and 2)	214
4.5.2.3.8 Comparison of Videos 1, 2 and 3	216
4.6 Subjects	218
4.7 Results and analysis of the pilot survey	219
4.7.1 Responses	219
4.7.2 Responses to open-ended questions	219
4.7.2.1 Open-ended question on Video 1	220
4.7.2.2 Open-ended question on Video 2	221
4.7.2.3 Open-ended question on Video 3	222
4.7.2.4 Open-ended question on the whole questionnaire	223
4.7.3 Assessment of parameters	224
4.7.3.1 Modes and medians	225
4.7.3.2 Means	229
4.7.3.3 Similar response patterns (analysis of frequencies of ratings)	231
4.7.3.4 Interrelations among variables	237
4.7.3.4.1 Bivariate correlations of variables	237
4.7.3.4.1.1 Video 1	243
4.7.3.4.1.2 Video 2	244

4.7.3.4.1.3 Video 3	245
4.7.3.4.1.4 Videos 1-2, 1-3 and 2-3	246
4.7.3.4.2 Perceptual maps of variables by multidimensional scaling	246
4.7.3.4.3 Hierarchical clusters of variables	256
4.7.3.4.4 Validation of questionnaire (analysis of principal component)	264
4.8 Conclusion - synthesis of the pilot survey	285
4.8.1 Form of speech and gestaltic perception	285
4.8.2 Suggestions for a further development of the questionnaire and the survey	287
<b>CONCLUSIONS</b>	
1. Sound material and form	289
2. Synchresis	290
3. Expressivity and comprehensibility – the experimental variable	291
4. Effects on TV interpreting profession and didactics	292
<b>REFERENCES</b>	<b>293</b>
<b>APPENDICES</b>	<b>311</b>



## INDEX OF TABLES AND FIGURES

<b>Table 1.1.</b> Items of the questionnaire by Bühler (1986) with the respective preferences assigned to each item by respondents (in percentages) in a four-point labelled scale – my calculations based on the data in the paper.	27
<b>Table 1.2.</b> “Shows the significance attributed to the different criteria by the four groups of subjects”(Kurz 1993:16). Numbers indicating ranking positions (in brackets) are mine.	28
<b>Table 1.3</b> Comparative ratings of quality criteria by Kurz & Pöchhacker (1995: 352). Numbers in brackets indicating the ranking position are mine.	29
<b>Table 1.4</b> Relative importance of output-related quality criteria (in percentages).The criteria are presented in the same order as in the web-based questionnaire. Data from the present survey are shown in boldface, while Bühler’s results (1986) appear underneath in normal font (Pöchhacker and Zwischenberger 2010). Ranking numbers (in brackets) in the first column are mine.	30
<b>Figure 1.1.</b> “Distribution of the degree of importance given to each linguistic criteria” (Chiario and Nocella 2004: 287).	33
<b>Figure 1.2.</b> “Shepard diagram of linguistic criteria” (Chiario and Nocella 2004: 289).	34
<b>Figure 1.3.</b> “Interpreter’s image of linguistic criteria” (Chiario and Nocella 2004: 290).	34
<b>Figure 1.4.</b> “Requirements of interpretation”, from the opened-ended replies (Moser 1995: 8).	38
<b>Table 1.5.</b> “General presentation of results” is in Gile (1990: 70) – calculations of percentages are mine.	41
<b>Table 1.6.</b> Partial results by Marrone (1993, in Soler Caamaño 2006: 75).	43
<b>Table 1.7.</b> Results from quality assessment and ranking of criteria in Vuorikoski (1993, in Soler Caamaño 2006: 79).	44
<b>Table 1.8.</b> Comparison of ratings of quality assessment and expectations in Mack and Cattaruzza (1995, in Soler Caamaño 2006: 89).	45
<b>Table 1.9.</b> Results of the survey on SI quality expectations and quality assessment (Garzone 2003: 26-27).	47
<b>Table 1.10.</b> “Ranked average ratings (4-point scale) of quality [assessment] features by user group (based on Guardini 1995)” (Russo 2005: 12).	49

<b>Table 1.11.</b> “Comparison of quality-related preferences in film and conference interpreting (ranked average ratings of importance on a 4-point scale)” (Russo 2005: 15).	49
<b>Figure 1.5.</b> Proposal of graphic synthesis of the interdependence of “non-verbal parameters” and their “interrelation” with “verbal” parameters, as found in Collados Aís et al. (2007) – the figure is mine.	53
<b>Figure 1.6.</b> “Propuesta – un cuestionario de satisfacción” (Soler Caamaño 2006: 275).	57
<b>Figure 2.0.</b> Screenshot from WinPitch. The central largest window is the “analysis window”, where the pitch contour (melodic curve – in black), the intensity (volume – in blue) and the waveform (in green) are showed. On the top right of the image there is the video track window. Below the analysis window there are the layers containing the transcript.	77
<b>Table 2.1.</b> Transcription conventions for the CorIT project (Falbo).	79
<b>Table 2.2.</b> ORenesit audiovisual recordings.	82
<b>Table 2.3.</b> Excerpt of transcript from the thematic corpus ORenesit – first 2012 presidential debate (English-Italian from SkyTG24).	85
<b>Table 2.4.</b> ORenesit corpus.	86
<b>Figure 2.1.</b> Rendition of face-threatening attacks in the four interpretations of the three 2012 US presidential debates (ORenesit – 60%) – based on Babetto (2013: 96-97).	87
<b>Figure 2.2.</b> Rendition of face-threatening interruptions in ORenesit (60%) – based on Babetto (2013: 96-97).	88
<b>Figure 2.3.</b> Frequencies of rendition strategies for the translation of culture-bound elements in ORenesit (77%) (Tomasetig 2014: 79).	89
<b>Figure 2.4.</b> Occurrences of self-repairs in ORenesit (77%) (Penso 2014: 92).	90
<b>Figure 2.5.</b> RTVE interpreting booth equipment. (< <a href="http://manzanaresdigital.blogspot.it/2012/10/obama-hace-unos-discursos-muy-sensatos.html">http://manzanaresdigital.blogspot.it/2012/10/obama-hace-unos-discursos-muy-sensatos.html</a> >).	92
<b>Figure 3.1.</b> Complex casual connection of elements related to speech, through the mediation of interm-ediate elements (Bühler 1934/1990: 31). One of the interlocutors ( <i>efficiens</i> ) produces an acoustic phenomenon and acts as a stimulus on the other ( <i>effectus</i> ). The central element is the acoustic phenomenon perceptible by senses (cf. Bühler 1934/1983: 77).	109

- Figure 3.2.** Causal chain of primary phonic communication (Bühler 1934/1990: 32). 110
- Figure 3.3.** Organon model of language (Bühler 1934/1990: 35). The central circle represents the concrete acoustic phenomenon, the triangle the three moments of its perception, corresponding to three kinds of sign: symbol for representation, symptom for expression, and index for appeal. The triangle includes something less than the circle in virtue of the principle of abstractive relevance, and something more in virtue of the principle of non-perceptual integration. The three groups of lines symbolize the semantically defined relations of the linguistic sign. The broken lines symbolize the indefinite similarity between sounds and things. 111
- Figure 3.4.** “Intersecting dichotomies” of axiom C (Bühler 1934/1990: 58). 113
- Figure 3.5.** Gestaltic symbolization (Bühler 1934/1990: 250). 118
- Figure 3.6.** Gestaltic symbolization of proper names according to John Stuart Mill proposal (Bühler 1934/1990: 253). 119
- Figure 3.6.** Questionnaire on the vocal style of Gaby Morlay (G.M) and Simone Signoret (S.S.) (Fónagy 1983: 162-163) – part I. 151
- Figure 3.7.** Questionnaire on the vocal style of Gaby Morlay (G.M) and Simone Signoret (S.S.) (Fónagy 1983: 162-163) – part II. 153
- Figure 3.8.** Examples of synthesis of recurrent melodic pattern found by Fónagy (1983: 275, 302) in a poem read by the author himself. 154
- Figure 3.9.** Melodic curves compared (Fónagy 1983: 294). 155
- Figure 3.10.** Semantic tests compared (Fónagy 1983: 293). 156
- Table 3.1.** Proposal for a questionnaire to elicit a gestaltic assessment of voice quality in TV broadcast simultaneous interpreting. 159
- Figure 4.1** The actantial model by Greimas (1966; in Schleifer 1987:103). 166
- Figure 4.2.** Example of the response pattern of one of the closed questions from the web Platform Qualtrics (® 2015) – screenshot reduction. 179
- Figure 4.3.** Example (top) of the response pattern of the closed questions with the intensity bar where to indicate the of intensity of perception with a cross; and example (bottom) of grid used by the researcher to quantify the perception indicated by the respondent. 179
- Figure 4.4.** Structure of the questionnaire for the pilot study. 181

<b>Table 4.1.</b> Stages of the process of construction of the questionnaire flow for the pilot study.	186
<b>Table 4.2.</b> Frequencies of ratings in assessment of Question 2 (type of voice) in pre-test (0-7 point scale).	187
<b>Figure 4.5.</b> Compared frequencies of ratings of question 16 of pre-test questionnaire, thought to elicit perception as a unique form (Gestalt) of SI speech.	188
<b>Figure 4.6.</b> The three questions of the pre-test that obtained the widest range of different ratings (4 to 5 out of 8, in a 0-7 point scale) in the three videos.	189
<b>Figure 4.7.</b> The two question of the pre-test that obtained the narrowest range of different ratings (3 to 4 out of 8, in a 0-7 scale) in the three videos.	190
<b>Figure 4.8.</b> Screenshot from Audacity® reporting an image of the activity of cutting / lengthening of silent pauses, developed - through two functions of the program - on the imitation track, but using the English speaker’s track to check the modifications. At the end of the process the two tracks were merged into a single track to be later imported in Adobe Premier Pro CS3 for editing.	195
<b>Figure 4.9.</b> Screenshot from Adobe Premiere Pro CS3 showing an image of the last part of the editing session, when, after the sound-image synchronization between the original Italian video and the audio track with the imitation and the original English speech, the original soundtrack had been deleted.	195
<b>Table 4.3.</b> Transcripts of the video excerpts used for the pilot survey. The original version in video 2 is reported to show the differences with respect to the imitation.	196
<b>Table 4.4.</b> Values of linguistic units and indices in the three video excerpts of the pilot survey, plus the original video 2, reported to show the differences with the imitation.	202
<b>Figure 4.10.</b> Pitch contour of the interpretation in Video 1. The x-axis indicates the length of the audio excerpt (69 sec), while the y-axis reports, on the top, the tonal range (blue line), expressed in semitones (-12 to 24); and on the bottom the intensity (volume – blue graph) and the waveform (green graph).On the line corresponding to semitone -12 there are some traces of the original stereophonic track, converted into a monophonic track through Audacity®.	204

- Figure 4.11** Transcript of the original interpretation with the original notes written by the dubber after listening to the original interpretation and before the recording of his performance. The part highlighted in grey corresponds to the text of the Video excerpt used for the experiment, the first part was cut. 209
- Figure 4.12.** Pitch contour (from WinPitch) of the dubbing actor's imitation of the original interpretation recorded in stereophonic mode and later rendered monophonic through Audacity®. 210
- Figure 4.13.** Pitch contour (from WinPitch) of the dubbing actor's imitation, after the session of *décalage* settlement through Audacity®. Main silent pauses lengthened were highlighted through WinPitch function and then indicated in the pitch contour (high) and the waveform (low). 211
- Figure 4.14.** Pitch contour (from WinPitch) in video 2 (after the experimental modification). 211
- Figure 4.15.** Pitch contour (from WinPitch) of the original video 2 (speaker + original interpretation). 212
- Figure 4.16.** Pitch contour (from WinPitch) of the video 3 (speaker + original interpretation). 216
- Table 4.5.** Randomization of the videos 1, 2 and 3. 219
- Figure 4.17.** Social and demographic profile of the subjects of the pilot survey. 220
- Table 4.6.** Values of descriptive statistics (mode, median and mean) related to the assessment of parameters in the pilot survey. 226
- Figure 4.18.** Similar response pattern found in 2 out of 23 plots showing the ratings assigned to each variable for the three videos. 232
- Figure 4.19.** Similar response pattern found in 2 out of 23 plots showing the ratings assigned to each variable for the three videos. 233
- Figure 4.20.** Similar response pattern found in 2 out of 23 plots showing the ratings assigned to each variable for the three videos. 234
- Figure 4.21.** Similar response pattern found in 3 out of 23 plots showing the ratings assigned to each variable for the three videos. 235
- Figure 4.22.** Response patterns among *effective comprehension 1, 2 and 3*, found in 3 similar questions out of 23 plots showing the ratings assigned to each variable for the three videos. 236
- Figure 4.24.** Multidimensional scaling of variables in video 1 (SPSS – Alscal). 249

<b>Figure 4.25.</b> Multidimensional scaling of variables in video 2 (SPSS – Alscal).	251
<b>Figure 4.26.</b> Multidimensional scaling of variables in video 3 (SPSS – Alscal).	253
<b>Figure 4.27.</b> Dendogram of video 1 (SPSS – hierarchical cluster analysis).	257
<b>Figure 4.28.</b> Dendogram of video 2 (SPSS – hierarchical cluster analysis).	259
<b>Figure 4.29.</b> Dendogram of video 3 (SPSS – hierarchical cluster analysis).	261
<b>Table 4.7..</b> Composition of the clusters identified through the dendograms of the hierarchical cluster analysis.	263
<b>Table 4.8.</b> Table of coefficients extracted from video 1 (SPSS – principal component analysis).	265
<b>Table 4.9.</b> Table of coefficients extracted from video 2 (SPSS – principal component analysis).	266
<b>Table 4.10.</b> Table of coefficients extracted from video 3 (SPSS – principal component analysis).	267
<b>Table 4.11</b> Comparison among variables present in each component, for the three videos – from the principal component analysis through SPSS (tables 4.8 – 4.10).	269
<b>Table 4.12.</b> Variables in each component (from the principal component analysis – SPSS – tables 4.8, 4.9, 4.10) in each video.	271
<b>Figure 4.30.</b> Scree plot of video 1. The plot straightens out horizontally starting at component 3; therefore two components should be extracted for video 1.	275
<b>Figure 4.31.</b> Scree plot of video 2. The plot straightens out horizontally starting at component 6; therefore five components should be extracted for video 2.	275
<b>Figure 4.32.</b> Scree plot of video 3. The plot straightens out horizontally starting at component 6; therefore five components should be extracted for video 3.	276
<b>Table 4.13.</b> Synthesis from a principal component analysis of the three videos with selected number of components (SPSS – varimax rotation).	277
<b>Figure 4.33.</b> Principal component plot of video 1 (SPSS – two components – varimax rotation). The relevant variables, selected through the table of coefficients, are indicated in the plot (in red for the opposed variables along the component 1, blue for the variables	

of component 2 and violet for the variable relevant to the two components). 279

**Figure 4.34.** Principal component plot of video 2 (SPSS – two components – varimax rotation). The relevant variables, selected through the table of coefficients, are indicated in the plot (in red for the opposed variables along the component 1, blue for the variables of component 2 and violet for the variable relevant to the two components). 281

**Figure 4.35.** Principal component plot of video 1 (SPSS – two components – varimax rotation). The relevant variables, selected through the table of coefficients, are indicated in the plot (in red for the opposed variables along the component 1, blue for the variables of component 2 and violet for the variable relevant to the two components). 283

**Table 4.14.** Synthesis of the interpretation of the component plots – opposed variables common to the three videos. 284





## INTRODUCTION

### **History and object of the research. Rationale, objective and hypothesis**

When it comes to judge the quality of a television broadcast simultaneous interpretation, the voice is relevant; both listeners and speakers (interpreters) are aware of it. When the interpreter is visible, in the case of a consecutive interpretation, voice perception is related to gestures and facial expression; when s/he is not visible, as in the case of simultaneous interpretation, then expression it is all in the voice.

The present research study is an attempt to explain why this happens (**rationale**).

First (**chapter 1**), previous interpreting studies on the evaluation of quality of simultaneous interpretation were reviewed (De Gregoris 2014). They share more or less the same approach, based on the so called linguistic and paralinguistic (sometimes even extra-linguistic) evaluation criteria that led to similar results. The few studies on expectation and assessment of television or film interpreting confirmed the significance of voice. After a thorough analysis of each of the criteria adopted for the assessment, even through the use of experimental modifications of each of the parameters, it was proved that the linguistic and paralinguistic parameters are interrelated (ECIS Group – University of Granada). Afterwards, further experiments were carried out with artificial modification of parameters, and the interdependence of

parameters and their incidence on comprehension were confirmed (QuaSI project at the University of Vienna; Christodoulides and Lenglet 2014). Even before these experimental studies, a holistic evaluation of parameters was proposed (Garzone 2003; Soler Caamaño 2006). However, no research was carried out on the basis of perception of simultaneous interpreting speech, with the exception of Chiaro and Nocella (2004); in some cases (e.g. Collados Aís et al. 2007) the word “perception” could be found.

In a recent publication (Collados Aís and García Becerra 2015: 378) ECIS Group stated that “the division of the parameters as a function of whether they are verbal or non-verbal is difficult because of the many interrelationships that exist; nevertheless, we will keep this division for operational reasons”. The line of research of the group is further divided into the parameters of incidence of quality evaluation (voice, intonation, fluency, accent, diction, style, terminology, cohesion, complete and accurate meaning transfer), each of them is investigated by one or two researchers. The research carried out by the group is mainly experimental; however, in a recent publication (Collados Aís et al. 2011), each parameter was studied from a thorough multi(inter)disciplinary theoretical approach. As a consequence of the experimentally proven interrelation among parameters, future developments of the research by the ECIS group will concern a “vertical” approach, i.e. the study of “intraparameters” of each parameters (Collados Aís and García Becerra 2015). It may be supposed that within each parameter the related parameters will be considered.

In the above mentioned studies, notwithstanding the interdependence among parameters, a holistic theoretical approach has not been considered yet to study the perception of simultaneous interpreting speech from a formal perspective. For this reason, the present study, which is based on a gestaltic paradigm of quality evaluation, is supposed to be innovative. Moreover, considering the object of the study, that is TV interpreting, the audiovisual medium was also taken into account as a contextual aspect in the subjects’ situation of reception (**rationale**).

Goldman-Eisler’s experiments in psycholinguistics (1968), which included tests on spontaneous speech, reading and simultaneous interpreting, led the author to conclude that speech production is the result of a psychophysical activity where the entire body is involved. She found that the best way to define the structure of speech was the word “rhythm”, intended as an alternation of “hesitant periods” and “fluent periods”. However, when

she discovered that simultaneous interpreters created pauses also in “fluent periods”, she considered the speech production a result of a “totality of attitude”, a “specific neurophysiological set pervading the whole situation”. The conclusive proposal was that a “global tonigenic activation” stands on the background and engenders a “selective process” in speech production.

The second step of the research (**chapter 2**) was the building of a corpus of television broadcast simultaneous interpretation. Strictly related to the present study, one of the objectives of the corpus-building was to physically experience the perception and understanding of the phenomenon through the *praxis* of transcription.

The corpus built is made of Italian (2 versions) and Spanish (2 version: Spain and USA) interpretation of the 2012 US presidential debates. The transcription was executed by the researcher, in collaboration with Babetto (2013), Penso (2014) and Tomasetig (2014), who also analysed aspects of interpretation within their respective MA theses. The OReNesit corpus (Obama vs. Romney English español italiano) is made of 167,670 phonological words, while the digital archive includes 17 videos, for a total 1,147 minutes, or 19.4 hours).

Successively, (**chapter 3**), with the aim of proposing a holistic evaluation of television broadcast simultaneous interpretation, a theoretical paradigm was studied as a starting point to design a new questionnaire to test the hypothesis of a holistic assessment. The etymon of the word “prosodia” (i.e. modulation of aspects related both to words and sounds) and the anthropological, semiotic and linguistic evidences of the interrelation between words and sound, expression and representation, as well as the results of previous interpreting studies on quality evaluation, led to adopt an approach to speech perception from the point of view of the *Gestalt* perception. *Gestalt* is the visual perception of a whole as an interdependence of its components, which are not distinguished individually but that constitute a form, which in turn, in a second time, is fundamental to distinguish its components. The concept of *Gestalt* in psychology of perception was intertwined with the original notion of structure in linguistics, which designated the interrelation among linguistic elements considered as a unique system, where the whole is not the result of the sum of all its parts, but is something more. However, phonology abandoned very early the contribution from the psychology of perception to focus on the distinctive traits. Later on in the history of phonology, not only

the language sounds (phones), but also their mental representation (phonemes) became and still are segmental.

One of the few authors that applied the *Gestalt* perception to language was Karl Bühler in his Theory of language (1934/1983). Bühler considered language as an instrument functional to communication among human beings. Thus, he conceived an instrumental (*organon*) model of language, where the function of a language sign is threefold: it is a symptom (index) when it develops the expressive function, and is focused on the sender; it is a signal when it develops an appeal function, and is focused on the receiver; it is a symbol when it develops the representational function, and is focused on objects and states of affairs. The physiological and perceptive basic unit of spoken language is the syllable, and not the phoneme. The phoneme is a distinctive trait of word, but not for being immediately perceived in a distinctive way; however, receiver uses the phonemes only when the word is not recognized (for a damaged signal or distorted communication) through the gestaltic perception of its components in relation to the symbolic and the deictic field. One of the symbolic fields is the sentence, while the deictic field is the situation of the communicative act, actors included. A sentence is perceived as a word. Both words and sentences have a melody, which, in a gestaltic way, helps the receiver to recognize them; in fact, the sound shape of a word is not entirely determined by the expressive function, but also contributes to the symbolic value and the syntactic valence of the word. Word images are recognized primarily by their acoustic shape, and not only by the sequence of phonemes. Voice is deictic when indicates the person of the speaker (mainly through the timbre), it is symbolic when conveys physiognomic and pathognomic information about the speaker. Thus, it may be assumed that, according to Bühler, the main components of the acoustic face of words are voice-syllable-melody-words-sentence-context.

Fónagy (*La vive voix* 1983) thoroughly studied the physiognomic and pathognomic information conveyed by voice. He proposed that, depending on the psychic state of the speaker, s/he produces phones that in semantic tests with children were perceived as being consistent with that psychic state; hence the proposal of a phonation reflecting the subject's drives. He also studied physiologically and acoustically how attitudes are expressed through the voice, being the "ex-expression" a discharge of tension by the subject. The melodic curve is iconic of the physiological movements related to phonation. The continuous perception of attitudes in a speaker's voice may lead the

hearer to figure out a vocal style of the speaker through a sort of unconscious symbolization process; this process, from the vocal style may induce the hearer to imagine a vocal character. Artistic voice may condense semantic and prosodic patterns that in ordinary speech can also be contradictory; vocal artist's technique seem to add a third dimension to the melodic movement (made of duration and tone), that is, 'melodicity' (= a regular distribution of the fundamental frequencies among syllables). Poet's readings revealed "recurrent melodic structures" of the poem, confirmed by semantic tests, that, in the author's view, went beyond ordinary (linguistic) prosody. Prosody and articulation that are considered "expressive" can be different from ordinary ones; however, such deviation has to stay within the limits of comprehensibility, in order to acquire an aesthetic effect and thus be effective.

Benveniste (1966) was another scholar who studied formal perception, for example, through his philological proposal of the original etymon of the word "rhythm" as form. In fact, until the V century B.C., *ῥυθμός* (= rhythm) was used with reference to a form not fixed yet, but as it was perceived/produced by the subject in a moment of its development; and it was used with reference to atoms, institutions, moral formation, the display of letters in writing, human character, personality, etc. Afterwards, from Plato on, the word designated the durations and the intervals regulating the succession or the order of movements of human beings, their activities, their attitudes, or even things and events.

Drawing on Benveniste's (1966) archeology of the word "rhythm", and in line with the theory of enunciation, Meschonnic (1982) theorized the rhythm as form of a discourse. He defined the rhythm in language as the display, organization of the marks through which signifiers, both linguistic and extralinguistic (markedly in oral communication), produce a proper semantics, different from the lexical meaning, that he named *significance*, i.e. the specific values of a discourse. The marks of *significance* constitute a syntagmatics and a paradigmatics that neutralize the notion of linguistic level, since they may be classified at any level (prosodic, lexical, syntactic, etc.). The sense of a discourse is the product of the *significance*, and it is undefinable. The *significance* can be found in any mark, and then the rhythm is the organization of the sense in a discourse. Moreover, the rhythm includes the subject, because if the sense is an activity of the subject, and the rhythm is an organization of discourse, consequently the rhythm is the configuration of the subject inside the discourse. Meschonnic points out that his notion of rhythm is not

semiotic but semantic. In the end, in his proposal, sense, discourse, subject and rhythm almost coincide.

Albano Leoni (*Il volto fonico delle parole* 2009) recently criticized the segmental phonology on the basis of the ascertained variability and continuity of the acoustic signal. For these reasons, he adopted the *Gestalt* as a model of speech perception, in the way as it was proposed by Bühler. Albano Leoni proposed an actualization of Bühler's application of *Gestalt* to the speech perception, where the properties of his proposal of 'acoustic face of words' are: voice-syllable-prosody-sense-context-(linguistic) knowledge of the world. As Meschonnic, Albano Leoni points out that his model of speech perception is semantic more than semiotic, and is based on interpretation more than on decoding.

In the audiovisual field, Chion (*L'Audio-Vision* 1990/1994) adopted the concept of *Gestalt* perception to explain the "transsensorial" perception of audio-vision, which is not the perception of the sum of the two channels (image and sound), but it is something more, originated by the merging of the two components. To explain the audio-visual transsensorial perception, Chion uses the word "rhythm", while the *Gestalt* is mentioned with reference to the ear's perception: the ear in fact listens in brief slices, and what it perceives and remembers *already* consists in short syntheses of two or three seconds of the sound as it evolves; thus, perception (synthesis) follows the analysis. In audio-vision perception, gestaltic perception occurs in case of dubbing, post-synchronization and sound-effect mixing, that is, when a synchronized synthesis or "synchresis" of auditory and visual phenomena is required.

On the basis of the theoretical background synthesized above, and in line with the questionnaire-based survey on quality assessment of simultaneous interpretation, a new questionnaire was designed and tested in a pilot survey. First, a deep structure of the questionnaire was considered, taking into account: i) the properties of the 'acoustic face of word' (voice-syllable-prosody-sense-context-(linguistic) knowledge of the world); ii) the properties of *L'Audio-Vision* (synchresis, textual speech, theatrical speech and emanation speech); iii) and the properties of *La vive voix* (expressivity-comprehensibility-melodicity-vocal attitude-vocal personality). Secondly, a questionnaire flow was defined, made of 21 items, starting from the sound perception and ending with the effective comprehension of the audiovisual interpretation. The questionnaire-building consisted of five drafts, the third draft was used for a pre-test. Particular attention was paid to the phrasing and wording of the

questions, and their evaluation pattern, bearing in mind to administer the questionnaire, in a definitive study, to TV experts, a sample of ordinary TV viewers, actors and musicians.

For the survey (**chapter 4**), three video excerpts (1 min each) were selected from the Italian interpretation of the third US 2008 presidential debate, since the ORe nesit corpus had not been completed yet. An experimental variable was introduced by artificially modifying one of the three video excerpts: the interpreter's voice was replaced with that of a dubbing actor who (artistically) imitated the original Italian interpretation reading the original transcript in a recording radio studio. The recorded audio track was first modified by cutting and artificially lengthening silent pauses to create a *décalage* with the speaker's original English, then it was merged with the original speaker's track, and finally inserted in the original video track through a video editing software. The voice of the TV dubbing actor was used to test the expressivity/comprehensibility relationship of a telegenic voice to compare and contrast it with that in authentic televised interpretations.

The last draft of the questionnaire was built through the web-based tool Qualtrics, with the aim of administering the pilot survey through the web. However, this was not possible for legal reasons related to the modified voice of the original interpreter, which could not be made public through the web with the graphics of Italian state-owned TV channel Rainews24 superimposed on the video image. Consequently, the pilot survey was administered *in praesentia* to 101 BA and MA translation and interpretation students at the University of Trieste (SSLMIT). The sequence of the three videos (related to the same set of questions) administered was randomized by the researcher. Data obtained from the questionnaire were treated statistically through the methods of frequency of ratings, bivariate correlation of variables, multidimensional scaling, hierarchical cluster analysis and analysis of principal component. Results obtained from each method were analysed and discussed. After the analyses, a synthesis of the pilot survey was proposed.

In conclusion, it is worth to clearly state the **hypothesis** and the **objective** of the present research study. The aim is to propose a model of gestaltic assessment of TV broadcast simultaneous interpreting (or, in view of further developments, of other audio-visual material). The hypothesis is that a questionnaire built up for this purpose, on the bases of a sound theoretical paradigm, could be a useful tool.

A peer reviewer of an article of mine argued that the questionnaire I proposed is redundant (De Gregoris forthcoming). This is an undisputable argument, because the questions were created on the bases of a gestaltic theory. However, is there another way to elicit the gestaltic perception of the material in question? How can a questionnaire built on gestaltic perception be not redundant? Is not redundant the same concept of *Gestalt*? Is logic the gestaltic perception? A more strong argument would have been one against the fact that all the questions (excepted those on real comprehension) are closed questions. However, the evaluation pattern chosen (Osgood's semantic differential scales) is still applied in psychology, and is also used in the semantic tests of voice perception. In addition, the questionnaire to be tested for the first time is made of 20 closed questions: it is true that many of them are related to the same phenomenon (i.e. voice); however, a first test should be performed in order to see if few or many questions do not work. In the researcher's view, the main question is: can language perception be quantified? Is a quantitative approach to test language perception suitable? If so, can it explain the phenomenon better than linguistic or semiotic theory?



## Chapter 1

# LITERATURE REVIEW – QUALITY AND PROSODY IN SIMULTANEOUS INTERPRETATION

### 1.0 Introduction

This chapter includes a literature review of Interpreting Studies on the evaluation of quality in simultaneous interpreting (SI). The analysis was carried out with the aim of providing a proposal for a gestaltic evaluation, and the studies have been divided into two categories: quality expectations surveys (ideal expression of preferences) and quality assessment surveys (judgments after a real experience). Having considered the conclusions drawn from the studies reviewed, attention will be subsequently focused on studies on SI prosody.

### 1.1 Quality expectations and quality assessment in simultaneous interpreting

In surveys on quality expectations (ideal evaluation) of simultaneous interpretation, subjects tend to give more importance to those features that have been labelled as ‘content-related aspects’, like ‘transmission of the original sense’, ‘logical coherence’ and ‘terminology’, while in surveys on quality assessment (after a real experience of the phenomenon) subjects still give more importance to the so called ‘content-related aspects’. However, in this case they give higher ratings to those features that have been labelled as

‘form-related aspects’, like ‘fluency’, ‘voice’ and ‘rhythm’, which have an impact on the assessment of the ‘original sense’, ‘coherence’, ‘accuracy’ and on the evaluation of the ‘overall quality’ of an interpretation. This is particularly evident when perception is received through an audiovisual medium (Russo 2005), or when parameters are individually manipulated to measure the incidence of manipulation on other parameters (Collados Aís et al. 2007).

### **1.1.1 Questionnaire-based surveys on quality expectations in simultaneous interpreting**

In this section we will consider the studies on quality expectations of simultaneous interpreting (SI) through questionnaire-based surveys on the ideal evaluation of interpretation, or the expression of preferences of quality on the basis of pre-definite or spontaneous criteria. Quality criteria are linked to the items of the questionnaires, which require the respondents either to rate each of them on a numeric scale, to rank them in order of importance or to comment on them through open questions. Subjects are interpreters, delegates of EU institutions, and other users like medical doctors, engineers, etc.

#### *1.1.1.1 Bühler (1986)*

The first questionnaire on quality expectations in simultaneous interpreting (SI) was compiled by Hildegund Bühler (1986) and administered to 41 interpreters members of the *Association internationale des interprètes de conférence* (AIIC) and 6 members of the *Commission des admissions et du classement linguistique* (CACL) of AIIC. Quality criteria were grouped in “linguistic (semantic)” and “extra-linguistic (pragmatic)”. The “linguistic (semantic)” criteria were: *native accent, pleasant voice, fluency of delivery, logical cohesion of utterance, sense consistency with original message, completeness of interpretation, correct grammatical usage, use of correct terminology, use of appropriate style*; while the “extra-linguistic (pragmatic)” criteria were: *thorough preparation of conference documents, endurance, poise, pleasant appearance, reliability, ability to work in a team, positive feedback from delegates*. The study by Bühler (1986) revealed that the “linguistic” criteria received the highest ratings from respondents: *sense consistency with original message* was the first, followed by *logical cohesion of utterance, completeness of interpretation, use of correct terminology, correct grammatical usage, fluency of delivery* (1986: 232). The criterion *reliability* received the highest rating among “extra-linguistic” criteria, followed by *thorough preparation of conference documents* and

*ability to work in a team.* The criteria of *native accent*, *pleasant voice*, *use of appropriate style*, *endurance*, *poise* and *pleasant appearance* “were considered desirable in most cases but not essential” (1986: 233).

Items of the questionnaire (Bühler 1986)	highly important	important	less important	irrelevant
1. Native accent	23%	47%	28%	
2. Pleasant voice	28%	61%	9%	
3. Fluency of delivery	49%			
4. Logical cohesion of utterance	83%			
5. Sense consistency with the original message	96%			
6. Completeness of interpretation	47%	49%		
7. Correct grammatical usage	48%	50%		
8. Use of correct terminology	49%			
9. Use of appropriate style	7%	68%	15%	
10. Thorough preparation of conference documents	73%			
11. Endurance				
12. Poise				
13. Pleasant appearance				
14. Reliability	73%	22%		
15. Ability to work in a team	47%	49%		
16. Positive feedback from delegates				
Other criteria (please specify):				

**Table 1.1.** Items of the questionnaire by Bühler (1986) with the respective preferences assigned to each item by respondents (in percentages) in a four-point labelled scale – my calculations based on the data in the paper.

#### 1.1.1.2 Kurz (1989; 1993)

After Bühler, Kurz (1989) used a questionnaire-based survey using the first eight “linguistic (semantic)” criteria in Bühler (1986), but the questionnaire was administered to a specific group of users: medical doctors. Some years later the same questionnaire was administered to another two different user groups: engineers and delegates of the Council of Europe (Kurz 1993). The surveys conducted by Kurz (see below - Table 1.2) revealed that different user groups had different degrees of expectations of quality in simultaneous interpreting, or, in other words, quality criteria varied according to the user group. Interpreters had higher expectations than other groups, because they assigned higher rates to the criteria than other groups. Nonetheless, all user groups, on average, assigned more importance to *sense*

*consistency with original message*, followed by *logical cohesion*, *use of correct terminology* and *completeness of interpretation*; while the last positions of the ranking were occupied by *fluency of delivery*, *correct grammatical usage*, *pleasant voice* and *native accent*.

Criterion (rated on a 4-point scale)	Bühler '86 (in Kurz '93)		Kurz '89 (in Kurz '93)		Kurz '93		
	Interpreters (n=47)		Medical doctors (N=47)		Engineers (N=29)	CE delegates (N=48)	average
1. Native accent	2.9 (8)		2.3 (8)		2.2 (7)	2.08 (8)	2.365 (8)
2. Pleasant voice	3.085 (7)		2.6 (6)		2.4 (6)	2.396 (7)	2.6 (6)
3. Fluency of delivery	3.468 (4)		2.9 (5)		2.966 (4)	3.208 (5)	3.1 (5)
4. Logical cohesion	3.8 (2)		3.6 (1)		3.1 (3)	3.3 (4)	3.458 (2)
5. Sense consistency with original message	3.957 (1)		3.6 (1)		3.655 (1)	3.6 (2)	3.69 (1)
6. Completeness of interpretation	3.426 (5)		3.0 (4)		2.9 (5)	3.458 (3)	3.2 (4)
7. Correct grammatical usage	3.38 (6)		2.4 (7)		2.03 (8)	2.688 (6)	2.6 (6)
8. Use of correct terminology	3.489 (3)		3.4 (3)		3.138 (2)	3.729 (1)	3.4 (3)
Average	3.44		3.0		2.8	3.06	3.06

**Table 1.2.** “Shows the significance attributed to the different criteria by the four groups of subjects”(Kurz 1993:16). Numbers indicating ranking positions (in brackets) are mine.

### 1.1.1.3 Kurz and Pöchhacker (1995)

Kurz and Pöchhacker (1995) used Bühler’s eight “linguistic (semantic)” criteria for a questionnaire-based survey on quality expectations in television interpreting. This time, the questionnaire was administered to “a group of representatives of Austrian and German TV organizations” (1995: 351). In this case, TV professionals’ expectations for quality in “simultaneous interpreting for live television broadcasts” (1995: 350) were higher than conference participants’ (1995: 352).

Indeed, the ratings assigned to all quality criteria by TV experts were higher than those assigned by the average combined ratings of conference participants (interpreters, medical doctors, engineers, Council of Europe delegates) (see Table 1.3). Furthermore, even though TV representatives – just like conference participants – gave priority to *sense consistency with original*

*message*, followed by *logical cohesion*, they ranked *pleasant voice* as third most important criterion and *fluency of delivery* as the fourth, followed by *native accent*.

Criterion (rated on a 4-point scale)	TV professionals (n=19)	Conference participants (N=124) [“average” in Kurz (1993)]
1. Native accent	2.84 (6)	2.37 (8)
2. Pleasant voice	3.47 (3)	2.6 (6)
3. Fluency of delivery	3.32 (4)	3.1 (5)
4. Logical cohesion	3.68 (2)	3.46 (2)
5. Sense consistency with original message	3.84 (1)	3.69 (1)
6. Completeness of interpretation	3.53 (8)	3.2 (4)
7. Correct grammatical usage	2.79 (7)	2.6 (6)
8. Use of correct terminology	3.32 (4)	3.4 (3)
Average	3.22	3.06

**Table 1.3** Comparative ratings of quality criteria by Kurz & Pöchhacker (1995: 352). Numbers in brackets indicating the ranking position are mine.

#### 1.1.1.4 Pöchhacker and Zwischenberger (2010)

Within a web-based questionnaire survey on quality and role, Pöchhacker and Zwischenberger (2010) asked the respondents, among other things, to rate nine of Bühler’s “linguistic (semantic)” criteria, plus *lively intonation* and *synchronicity*, on a four-point scale. The questionnaire on simultaneous interpreting quality expectations was administered through Limesurvey to professional interpreters (AIIC members), and completed by 704 respondents.

As in Bühler’s study (1986), and in Kurz’s (1989; 1993), the criterion *sense consistency with the original* received the highest ratings, followed by *logical cohesion*. Unlike similar mentioned studies, in this case, the third most important criterion was *fluency of delivery*. The criteria *lively intonation* and *pleasant voice* received almost the same ratings, 28.2 and 27.5 respectively.

Criterion	Very important	Important	Less important	Unimportant	N=
Fluency of delivery	<b>70.7</b> 49 (3)	<b>28.6</b> 49	<b>0.7</b> 2	- -	<b>704</b> 47
Correct terminology	<b>61</b> 49 (4)	<b>38</b> 51	<b>0.9</b> -	<b>0.1</b> -	<b>703</b> 47
Correct grammar	<b>54.4</b> 48 (5)	<b>40.4</b> 50	<b>5.1</b> 2	<b>0.1</b> -	<b>701</b> 46
Sense consistency with original	<b>88.3</b> 96 (1)	<b>11.1</b> 4	<b>0.6</b> -	- -	<b>702</b> 47
Lively intonation	<b>28.2</b> (8)	<b>59.3</b>	<b>11.7</b>	<b>0.9</b>	<b>703</b>
Native accent	<b>14.1</b> 23 (11)	<b>42.1</b> 47	<b>39.7</b> 28	<b>4.1</b> 2	<b>701</b> 47
Logical cohesion	<b>74.8</b> 83 (2)	<b>24.8</b> 15	<b>0.4</b> 2	- -	<b>698</b> 47
Pleasant voice	<b>27.5</b> 28 (9)	<b>58.5</b> 61	<b>12.7</b> 9	<b>1.3</b> 2	<b>702</b> 46
Synchronicity	<b>15.3</b> (10)	<b>52</b>	<b>30.1</b>	<b>2.7</b>	<b>675</b>
Appropriate style	<b>36.2</b> 17 (7)	<b>55.6</b> 68	<b>7.4</b> 15	<b>0.9</b> -	<b>702</b> 47
Completeness	<b>47.7</b> 47 (6)	<b>45.7</b> 49	<b>6.3</b> 4	<b>0.3</b> -	<b>698</b> 47

**Table 1.4** Relative importance of output-related quality criteria (in percentages). The criteria are presented in the same order as in the web-based questionnaire. Data from the present survey are shown in boldface, while Bühler's results (1986) appear underneath in normal font (Pöschhacker and Zwischenberger 2010). Ranking numbers (in brackets) in the first column are mine.

#### 1.1.1.5 Meak (1990)

Meak (1990) devised a questionnaire to be administered to ten specialised medical doctors with a significant experience of participation in international conferences (1990: 8). The questionnaire was designed to find out what a specific audience (ten medical doctors) expected from a simultaneous interpretation in order to consider this information as feedback to “target” the interpretation to “the specific qualities of that audience” (1990: 8). Questions (see Appendix 1) were related to: (1) the ‘effectiveness’ of simultaneous interpretation in medical conferences; (2) the irritating aspects of the interpretation (incorrect terminology, incompleteness or omission of numbers and data); (3) the kind of information on the speaker required by the audience; (4) the kind of information that is considered necessary and therefore that the interpreter should select; (5) the possible irritating aspect of

speed of speech and its negative effect on comprehension; (6) the importance of the end of a presentation; (7) the importance of translating acronyms. The author maintained that the results of her survey did not have any statistical value (1990: 13). Responses and comments varied, but in some cases respondents gave similar answers: data and figures were expected to be accurate; descriptions of films needed to be translated; data shown on tables could be selected; a good translation of the end of a presentation as well as of final debates in conferences were also required. As to the speed of speech, a synthesis was not possible, since four respondents did not consider the high speed of speech an irritation factor, thus it did not have any impact on comprehension, while four respondents recognized that a high speed of speech negatively affected their comprehension.

#### 1.1.1.6 Chiaro and Nocella (2004)

Chiaro and Nocella (2004) used a questionnaire on quality expectations with the same “linguistic” criteria as those used by Bühler (1986) and partially the same “extra-linguistic” criteria from Bühler for the first web-based survey of its kind. The extra-linguistic criteria they used were: *preparation of conference documents*, *endurance*, *ability to work in a team* (also present in Bühler’s questionnaire), *concentration*, *physical well-being*, *mnemonic skills*, *encyclopaedic knowledge*, and *absence of stress* (not present in Bühler’s questionnaire). The authors pointed out that in their study “the criteria under scrutiny were not measured on an itemised-category scale but on a rank order scale”, because “Bühler’s results showed that interpreters found it difficult to point to unimportant factors” (2004: 283):

The difference between these two types of single-item scales is that respondents of an itemized-category scale must choose from one of several responses, options or categories, while on a rank order scale they are required to order a set of objects with regard to a common criterion (Chiaro and Nocella 2004: 283).

In addition, they decided to separate the nine “linguistic” and the eight “extra-linguistic” criteria “on two different ranking-order scales” in order to reduce respondents’ “mental effort” (2004: 283).

The questionnaire was sent to about 1,000 “interpreters belonging to several professional associations” in the world; 286 were returned, mainly from respondents “born” in Europe (49%) and America (56%); the majority of subjects were female (71%) with an average age of 45 (2004: 286).

Results showed that linguistic criteria were ranked in the following way: consistency with the original (1); completeness of information (2); logical cohesion (3); fluency of delivery (4); correct terminology (5); correct grammatical usage (6); appropriate style (7); pleasant voice (8); native accent (9). The “distribution of the degree of importance given to each linguistic criteria” is shown in Figure 1.1 (2004: 287).

As far as extra-linguistic criteria are concerned, the sum of the scores given to each criterion showed that *concentration* and *preparation of conference documents* were considered the most important, followed by (in this order): *ability to work in a team*, *endurance*, *physical well-being*, *mnemonic skills*, *encyclopedic knowledge*, *absence of stress* (2004: 288-289).

The authors also drew a “perceptual map” of the data on linguistic criteria, to have a graphic visualization of the respondents’ perception of the linguistic aspects assessed through the questionnaire (*ibidem*: 285).

In order to explore interpreters’ perception of linguistic factors affecting quality, multidimensional scaling (MDS) was performed using STATISTICA for Windows. MDS is a multivariate statistical technique which allows the researcher to develop perceptual and evaluative ‘maps’ (i.e. geometrical configurations) that summarize how people perceive various stimuli as being similar or different (see Hair et al. 1995) (*ibidem*: 285). MDS uses principal component analysis as the starting configuration of the similarity matrix. The program will then begin iterations under ‘steepest descent’ (Schiffman, Reynolds and Young 1981) with the goal of maximizing the goodness of fit (or maximizing ‘lack of it’).

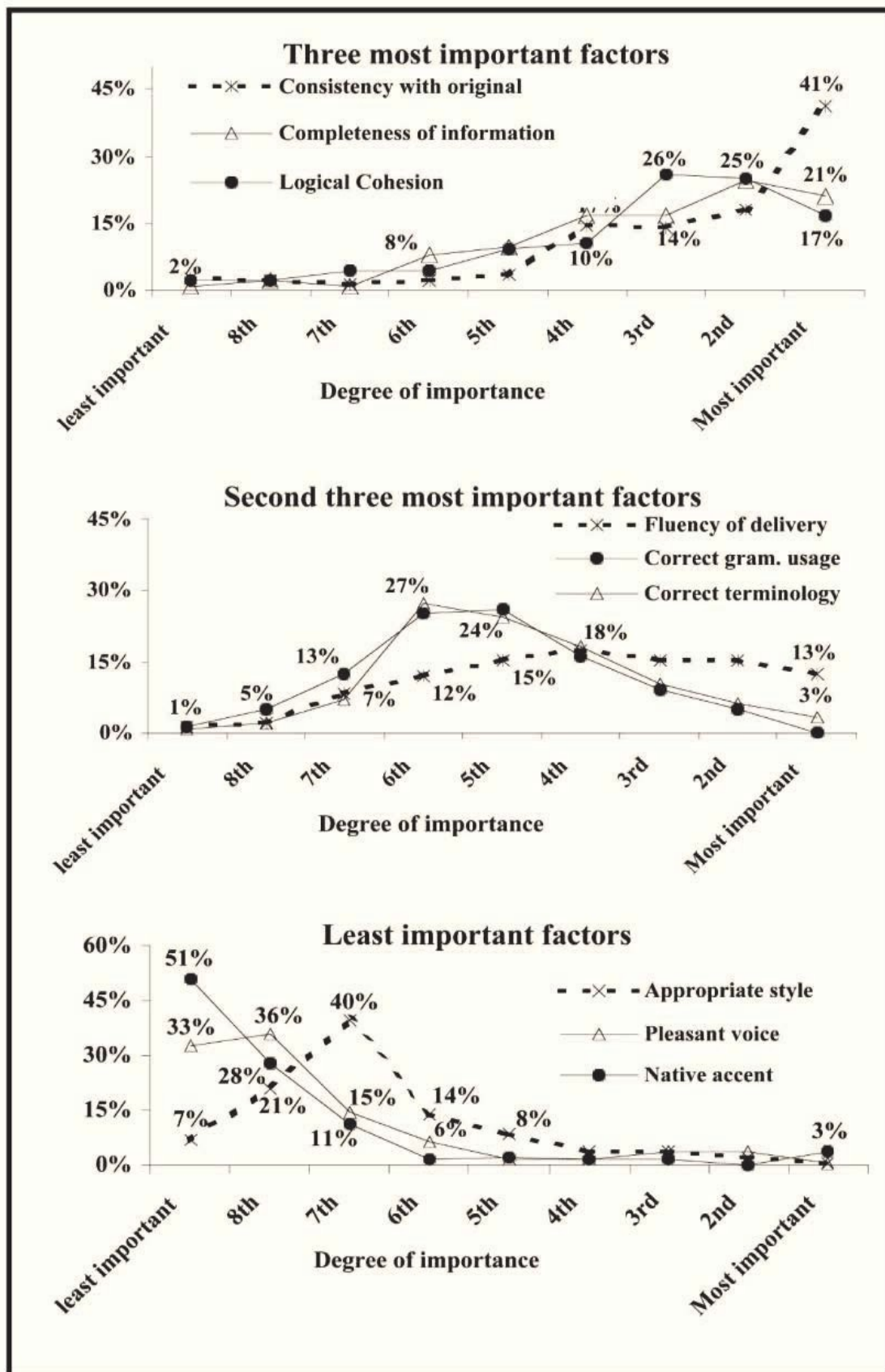
To this aim, first a “scree plot” and then a “Shepard diagram” were produced (*ibidem*: 289):

The ‘scree test’, which was performed plotting the stress value against a different number of dimensions, indicated that in our analysis two dimensions were the best solution for our data. Moreover, the Shepard diagram shows that reproduced distances fall close to the step-function (monotone transformation of the input data) and thus fit the model well.

---

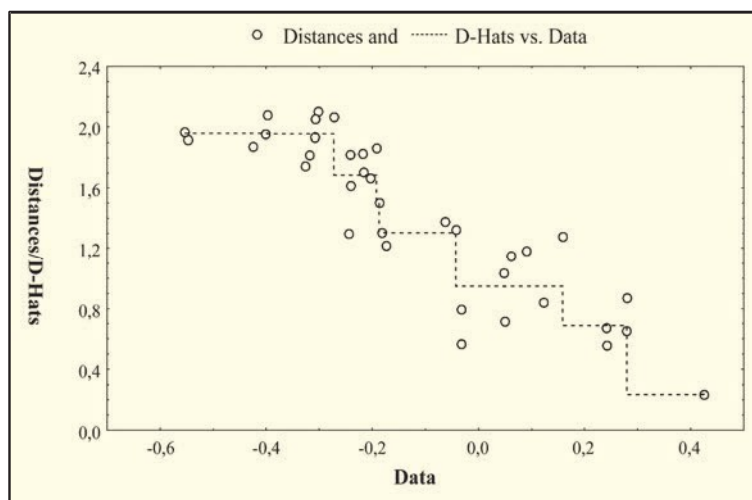
<sup>1</sup> “The scree plot is a test performed to decide how many dimensions are used in drawing the perceptual map” (Chiaro and Nocella 2004: 292).





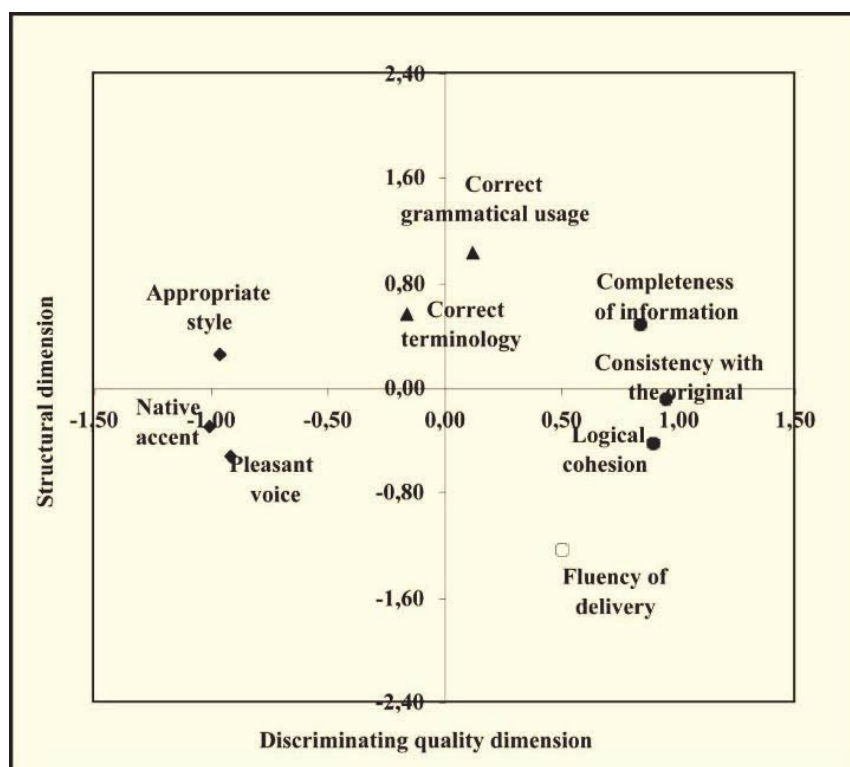
(Chiaro and Nocella 2004: 287).

The scree plot was not included by the authors in the article, while the Shepard diagram is reported below (Figure 1.2).



**Figure 1.2.** “Shepard diagram of linguistic criteria” (Chiaro and Nocella 2004: 289).

Thus, a perceptual map was drawn, called “interpreter’s image of linguistic criteria” (see Figure 1.3), where the horizontal axis was the “discriminating quality axis” and the vertical one was the “structural axis” (2004: 290).



**Figure 1.3.** “Interpreter’s image of linguistic criteria” (Chiaro and Nocella 2004: 290).

According to the authors, “this chart confirms and strengthens” their “previous findings”; therefore, the criteria *completeness of information*, *consistency with the original* and *cohesion* “score very closely on the right of the discriminating quality axis”, meaning that “they appear to be perceived in a more or less similar way”. The same is true for the criteria *voice quality*, *accent* and *style* on the left (negative) side of the horizontal axis, while the criteria *lexis* and *grammar* “score very closely and positively on the structural axis”. The criterion *fluency of delivery* “appears to stand alone and thus results in being dissimilar to any other features”. The authors proposed that such a position was due the fact that “intonation is considered by interpreters [...] on the interface of the two dimensions” (2004: 290):

On balance, considering that fluency in language plays the double role of both embellishment (i.e., in terms of speed, voice control and absence of hesitation) and structure (i.e., the supra segmental significance of stress, pitch and tone) then it would appear that this feature has indeed been placed where it would most obviously occur.

#### 1.1.1.7 *Kopczyński (1994)*

Kopczyński (1994) conducted a survey on quality expectations in different user groups: respondents involved in the humanities, experts in science and technology, diplomats. The questionnaire was administered to “people who attended international conferences as speakers or hosts, or who participated in negotiations in one or both of these roles” (1994: 91). Such a variability of subjects was due to the fact that according to the author “quality is not an absolute value, but rather contextually determined” (1994: 88). This means that the variables changed according to the communicative situation. Variables were: the “speaker”, his/her “attitude, status and intention”; the “interpreter” and his/her competence; the “receptors” and their “attitude”; the “message”, its “form” and “illocutionary force”; the “existing norms of interaction and interpretation of a speech community”; the “setting” (1994: 88). The questionnaire flow started with an open-ended question about what the respondents “considered to be the most important function of interpreting a conference” (1994: 92). The following question was the same as the first one, but this time respondents “were asked to grade the priorities alongside the proposed suggestions”: *rendering the general and detailed content of T1; terminological precision; style; grammatical correctness of utterances; fluency of delivery; diction; voice qualities*. The following two questions had the same pattern as the previous ones: the respondent was asked to “mention whatever s/he

considered as most irritating” (1994: 92) in an open-ended question first, and then suggestions to rank: *faulty terminology; ungrammatical sentences; stylistic mistakes; incomplete sentences; lack of fluency; poor diction; monotonous intonation and tempo; speeding up and slowing down; too general and too detailed rendition of content*. The remaining five questions concerned “the more or less active role of the interpreter (the ghost role vs. the intruder)” (1994: 92).

The results showed that the criterion *rendition of detailed content* was considered the most important and *terminological precision* the second most important both by “speakers” and “receptors”, while the third most important criterion was *fluency* for “speakers” (followed by *grammaticality* and *style*), and *style* for “receptors” (followed by *fluency*) (1994: 93). As to irritating aspects, both “speakers” and “receptors” agreed in considering *incorrect terminology* as the most irritating feature; the second source of irritation for the “speakers” was *exact rendition of the content*, while for the “receptors was the item *unfinished sentences*, followed by *grammaticality* (1994: 94). As regards “the role of the interpreter”, “all were in favour of empathy with the speaker and considered the ghost role of the interpreter as preferable” (1994: 96); according to the majority of respondents, “the interpreter should imitate the tempo and the intensity of voice of the speaker, but not necessarily the gestures”. Although all respondents preferred the ghost role of the interpreter, the majority of them also allowed “corrections of the speaker (with some reservation) and additional explanations”; among respondents, “speakers” accepted “to be corrected”, while “receptors opposed the idea” (1994: 96-97).

#### 1.1.1.8 Moser (1995)

Peter Moser (1995) reported a “Survey on Expectations of Users of Conference Interpretation” “entrusted” to “SRZ Stadt + Regionalforschung in Vienna/Austria” by the AIIC Research Committee (1995:1). The survey was designed taking into account the differences of expectations (i) between interpreters and users, and (ii) among different user groups (as defined by Kurz 1989, 1993). The survey also set out to consider the effect of the situational context (conference type) on the evaluation (1995: 4). It was an interview survey with questionnaire; interviews (over 200) were conducted by interpreters at 84 different meetings around the world, while the interviewees were all conference participants (listeners and/or speakers) (1995: 5-7).

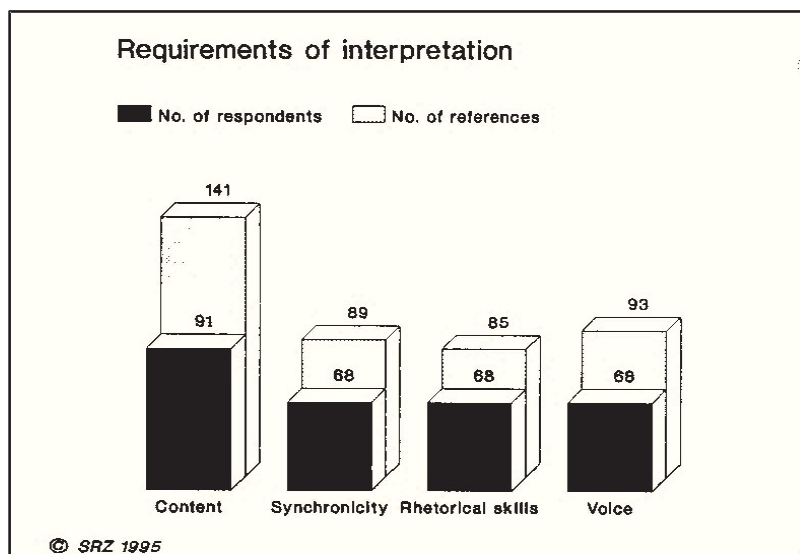
The questionnaire was made up of 33 questions (1995: 24-47). The first question about quality categories (number 2 in the questionnaire) asked the

respondents to rate on a five-point scale what criteria of an interpretation they considered most important among *completeness of rendition, clarity of expression, correct terminology* or *other* (1995: 25-26). The following questions asked the respondent whether – in the conference s/he has just attended – it “is more important that the interpreter focuses on the essentials or gives a complete rendition”, and “whether fidelity to the meaning of the original or the literal reproduction of what is said is more important” (1995: 26-27). The fifth question asked respondents to “indicate” (on a five-point scale) how important the following criteria were considered: (“the interpreter should...”) *speak in a lively and animated way; speak in complete sentences; interpret titles, names of functions, positions, offices held; interpret subtitles in graphs and tables on transparencies or slides; interpret abbreviations; anything else* (1995: 27-28). The sixth question was divided into two parts: the first part was an open-ended question about what most “irritates” the respondent; the second part suggested a series of criteria, asking “how irritating is an interpreter who”: *makes long pauses, lags far behind the original; speaks very quietly; speaks in a monotonous way; ums and ahs; has a foreign or regional accent* (1995: 28-30). An interesting aspect of the data processing is the following:

All the open-ended replies were first sorted into a highly differentiated category matrix. Each category was then allocated a code so that subsequently the open-ended replies could be statistically analyzed in conjunction with the replies to the closed questions (1995: 49).

In this way, if a respondent spontaneously mentioned the importance of one criterion two or more times, that number of references appeared in the statistics; therefore, there was no one-to-one correspondence between the number of interviewees and the number of references to a criterion.

The spontaneous quality criteria were grouped into four general categories: *content* (at least *faithfulness to the original*); *synchronicity* (referring to the need for a short *décalage*); *rhetorical skills* (*regular delivery, absence of hesitation, complete and grammatically correct sentences, clarity of expression*); *voice* (*lively, non-monotonous voice, clear enunciation, ‘a natural sounding voice’ or ‘an agreeable voice’*) (see Figure 1.4) (1995: 8).



**Figure 1.4.** “Requirements of interpretation”, from the opened-ended replies (Moser 1995: 8).

Data showed in Figure 1.3 show the aggregate factors of users’ expectations in open-ended replies; these data varied according to the conference type (“small/large general/technical meeting”), gender, level of respondents’ experience in conference interpreting (“newcomers, less experienced, very experienced”). From the analysis of open-ended replies it emerged that “participants in large general meetings mentioned voice quality in the interpreter almost as often (14 = 47%) as the need to be faithful to the original (16 = 53%) (ibidem: 9). However, in small general meetings, 11 participants (= 44%) spontaneously mentioned the interpreter’s rhetorical skills, while in large general meetings, rhetorical skills were mentioned only by 8 participants (= 27%) (ibidem). Therefore, the differences among the categories of quality were not related to the type of conference, but more probably to the characteristics of users (ibidem). Indeed, it was found that less experienced conference participants assigned more or less the same importance to the aspects related to faithfulness to meaning, synchronicity, rhetorical skills and voice; conversely, more experienced participants considered content match “a long way ahead of the other factors” (ibidem). Moreover, it was found that “overstated intonation or histrionic delivery” were not accepted by subjects; in fact, “the desirability of adopting a neutral tone in relation to the message being interpreted [was] explicitly mentioned 21 times” (ibidem: 11). Poor microphone discipline was mentioned 30 times as being “irritating” (ibidem). Subjects’ perception of interpreters’ professionalism was characterized by high qualification; interpreters were

expected to have a “broad grasp of current affairs and extensive general knowledge” (ibidem). Interpretation was seen as an activity requiring an “intense concentration without breaks” and the ability to convey “cultural and emotional content of the original” (ibidem: 12).

According to the author of the report, the above mentioned results from open-ended replies had to be compared to the analysis based on the replies to specific questions about aspects of quality (ibidem: 13). The author stated that these replies were divided into “form” and “content” and that such “distinction derive[d] from the empirical results of the survey” (ibidem: 19). From the replies related to the content, it emerged, among other things, that “concentration on essentials” was considered a more important aspect than “completeness of rendition”, across all conference types, more by experienced conference users than by less experienced users (ibidem: 15). It was observed that the marked preference for “faithfulness to meaning”, opposed to the “literal reproduction”, especially in “spontaneous debates”, rose with age (ibidem: 17-18). Among the formal expectations, “synchronicity” (i.e. “simultaneity between original speech and its interpretation”) was deemed to be the most irritating aspect, mainly due to “long pause” and “lagging behind”, especially for women and very experienced subjects (ibidem: 19). Among the “rhetorical skills” (another aspect of “form”), the most important feature was the “clarity of expression”, especially for women (ibidem: 21). Women proved to be more sensitive than men to the “ums and aahs”, or “other ways of filling in pauses” (ibidem). A “lively animated voice was considered an important or very important factor”, especially by more experienced subjects (ibidem: 22).

The type of conference proved to have an impact on quality expectations, for example, it emerged that “the larger the conference (number of participants) the greater the preference for interpreters to concentrate on essentials” rather than on “completeness of rendition” (ibidem: 24). In addition, “the less technical a conference, the weaker the requirement for completeness [...], the more technical the meeting the greater the preference for literal reproduction of papers and debates” (ibidem).

### **1.1.2 Quality assessment questionnaire-based surveys in simultaneous interpreting**

This section considers studies on quality assessment of simultaneous interpreting (SI), i.e. questionnaire-based surveys on the judgment of real

interpretations, on the basis of pre-definite or spontaneous criteria. These quality criteria are linked to the items of the questionnaires, requiring the subjects either to rate each of them on a numeric scale, to rank them in order of importance or to comment on them through open questions. Subjects are interpreters, interpreting students, delegates and other types of users. Some of the studies reviewed in this section, besides questionnaires on quality assessment also include questionnaires on quality expectations, administered either before or after questionnaires on assessment, with the aim of comparing the results of the two evaluation approaches. Only one of the studies reported (Collados Aís et al. 2007) also includes a “contextualization” questionnaire, designed to better elicit the respondent’s personal understanding of the criterion the subject was required to assess.

#### 1.1.2.1 Gile (1990)

Gile (1990) carried out a survey (a “case-study”) on quality assessment (judgement) of SI, with the aim of studying “the subjective perception of quality by delegates” (1990: 68). A bilingual questionnaire was devised to be administered to French and English speaking delegates at a conference. Eighteen French speaking delegates and five English speaking delegates returned the questionnaire (1990: 67). The quality was assessed according to the following criteria, rated on a five-point scale: *general quality*, *linguistic output quality*, *terminological usage*, *fidelity*, *quality of voice and delivery* (voix, rythme et intonation), *main weaknesses of interpretation* and *general comments on interpretation* (1990: 71).

The results showed that assessment of general quality was quite homogeneous, since it was considered “good” or “very good” by almost all the respondents. Homogeneity also characterised all the other criteria, with the exception of *voice*, which presented the most varied judgements (see Table 1.5, where the overall results are reported). Moreover, ratings assigned by English speaking delegates were higher than those assigned by French speaking delegates, since the former rated as “very good” all the criteria and defined the interpretation as “excellent”, without any “weakness” (1990: 67-68, 71). Six out of the eighteen French speaking delegates judged *voice* with the lowest ratings. However, this evaluation had no impact on the judgement of the *overall quality* of the interpretation, since the rating of this criterion was in line with those of other criteria (1990: 68).



CRITERIA	Very poor	Poor	Medium	Good	Very good
General quality of interpretation			9%	45%	45%
Linguistic output quality			14%	45%	41%
Terminological usage			18%	50%	32%
Fidelity			9%	50%	41%
Quality of voice and delivery		9%	22%	43%	26%

**Table 1.5.** “General presentation of results” is in Gile (1990: 70) – calculations of percentages are mine.

### 1.1.2.2 Ng (1992)

Ng (1992) designed a quality assessment questionnaire to be administered to Japanese native speakers, with the following working hypothesis: “appropriate use of speech levels is significant in English-Japanese interpretation at the conference level” (1992: 36), because speech levels could be “loosely defined” as “grammaticised social relationship indicators” (1992: 35), referring “to the choice of expressions which indicate the speaker’s social and psychological stance at a particular point in time” (1992: 36). “Speech levels” were otherwise referred to as “levels of politeness” (1992: 36, 37).

The questionnaire was administered to ten native Japanese speakers and submitted in two stages: in the first stage assessors were asked to indicate the general usability of the interpretation; in the second stage, they were asked to indicate if speech levels “interfered” with the content delivery. Specifically, in stage 1 subjects were requested to listen to the students’ prerecorded English-Japanese interpretations, executed by native English interpreters (not Japanese), and then answer three questions about: (1) whether or not the interpretation “could be followed”, and the message the interpreter was trying to convey could be grasped; (2) the general impression of the Japanese used by each interpreter; (3) which of the interpreters gave the best interpretation and which gave the worst, if possible explaining the reasons for the ranking.

Stage 2 consisted of an individual interview held approximately a week after Stage 1. After a short briefing on “the technical linguistic terms used to describe speech levels in Japanese”, with the aid of “examples”, the subjects listened to “short segments of the tapes” to “refresh their memory”. After listening, they were asked 4 questions on: (1) what they thought of the “inappropriate use of speech levels in interpreting”; (2) whether that “interfered with the content delivery”; (3) whether they found that “offensive”; (4) whether that made them feel uncomfortable. Lastly, the

subjects were asked to rate (on a five-point scale) the performance of each interpreter relative to “the control of Japanese speech levels” (1992: 37). The results showed that answers relative to Stage 1 “were general and covered a broad area”. However, the comments “fall into three broad categories of content, language and extralinguistic criteria”. “Comments on the language” [...] could be referred to the following “subcategories”: (2a) “naturalness, e.g. intonation, pronunciation and accent”; (2b) “grammatical structure”; (2c) “choice of vocabulary”; (2d) “speech levels”. “All the subjects were particularly concerned with whether the interpreters had grasped the meaning of the message”, therefore the interpreters “were criticised for giving obscure interpretations”.

In the end, the results from Stage 1 showed that “though speech levels were discussed, it was by no means the most important variable singled out by the subjects”. The same aspect characterised findings in Stage 2, where the subjects found it difficult to “isolate speech levels as a separate variable”; in some cases they “seemed to stray from the topic of speech levels to the more general problem of fluency”. The general result of the study was that for native Japanese speakers “appropriate use of speech levels” was “important”, but their “misuse” did not necessarily result in offending the audience” (1992: 37-40).

### *1.1.2.3 Marrone (1993)*

Marrone (1993) carried out a survey on “audience expectations and preferences” in consecutive interpreting, with particular attention to “definition and evaluation of interpretation quality” (1993: 35). The questionnaire was administered to an audience of 150 users (mostly students, with only a dozen researchers and professors) of consecutive interpreting who attended a lecture on constitutional law. The questions aimed at eliciting information about “audience’s preferences and reactions” (1993: 35). The questionnaire was made up of seven questions designed to elicit the respondents’ reactions to the interpretation of segments, the speed of delivery, the effect of the interpreter’s tiredness on performance, the translation of institutional terms, the interpreter’s role as a translator or a mediator.

The questionnaire was administered after the performance of a consecutive interpretation. However, from the wording of the questions it is not clear whether these asked the respondent to refer to the performance listened to for an ‘assessment’ or not: by reading the questions, it seems that

the performance served as an input to indicate “preferences” about the ‘ideal consecutive interpretation performance’.

Results showed that respondents “seem[ed] to attach far more importance to substance, fidelity, completeness of information than to the linguistic quality of prosodic features of interpretation”, even if “scores related to such features tended to fluctuate widely” (1993: 38). Moreover, responses “indicate[d] that it [was] appropriate” that interpreters “attempt[ed] a degree of cultural mediation”.

Parameters	Score
Complete transmission of the original message	216
Quality of style and correct terminology	165
Quality of intonation and delivery	137

**Table 1.6.** Partial results by Marrone (1993, in Soler Caamaño 2006: 75).

#### 1.1.2.4 Vuorikoski (1993)

Vuorikoski (1993) conducted a survey on quality perception of users of SI from English into Finnish with the aim of studying the effectiveness of communication mediated by interpreters in conferences held in Finland. A questionnaire was administered to the attendees of 5 seminars; 177 participants responded, some of whom were later interviewed by phone to complete the information provided in the questionnaires. The part of the questionnaire regarding quality assessment included absolute (yes/no) questions relative to the following aspects: the interpreter was well informed; the interpretation was coherent and easy to follow; it was accurate; rhythm was pleasant; it was fluent; terminology was correct. The following question asked respondents to rank the above-mentioned criteria.

The results showed that all the questions received a “yes” answer in a percentage equal to or higher than 50%. As to the ranking, the good preparation of interpreters was ranked first, coherent and easy-to-follow interpretation was ranked second and fluency of interpretation was ranked third (Vuorikoski 1993, in Soler Caamaño 2006: 77-79).

Criteria	Assessment	Ranking
Interpreter well informed	67%	22.7% (1)
Coherent and easy to follow interpretation	85%	19.7% (2)
Accurate interpretation	50%	16.9% (4)
Pleasant speech rhythm	64%	8.6% (6)
Fluency	50%	18.6% (3)
Terminology	60%	13.5% (5)

**Table 1.7.** Results from quality assessment and ranking of criteria in Vuorikoski (1993, in Soler Caamaño 2006: 79).

#### 1.1.2.5 Mack and Cattaruzza (1995)

Mack and Cattaruzza (1995) conducted a survey on user reception and expectations of SI based on the methodology adopted by Vuorikoski (1993), as in 2.4. According to this “multimodal” research, “data obtained through the questionnaire were integrated and partially checked by telephone and personal interviews as well as by non-reactive research in the conference situation” (1995: 38). The aim of this survey was to study “how quality [was] measured” in order to explore the possibility “to go beyond purely subjective judgment” (1995: 37). The questionnaire was distributed at 14 conferences, but only five of them were selected because their context was considered to match the criterion of “users’ high communication needs” (1995: 39). The number of participants who returned the questionnaire completed was 75; of these, only the 58 Italian participants were considered for analysis (1995: 41). Two central questions asked the respondents to evaluate – through a rating on a five-point scale – the simultaneous interpretation heard (assessment), and then to “indicate the importance of the criteria listed (their wishes and expectations)” (1995: 43). The criteria were the same as those used by Vuorikoski (1993) in her survey: the interpreter was well informed; interpretation was coherent and easy to follow; it was accurate; rhythm was pleasant; it was fluent; terminology was correct (cf. 2.4).

Findings in assessment (factual experience) showed that participants found that the interpretations were easy to follow, had a pleasant speech rhythm and that the interpreters were well informed – the mean rating of all these criteria was 4.2; the mean rating for fluency was 4.1, while both terminological correctness and accuracy received a rating of 3.8 (“slightly lower”). As to the results of quality expectations, terminological correctness received the highest mean rating (4.5), followed by the preparation of the

interpreter (4.3) and an accurate and easy-to-follow interpretation (both 4.1) (1995: 43-45).

All in all, the average values of the ratings moved in a rather narrow range (quality experienced: 3.8 to 4.2; expectations: 3.8 to 4.5). The largest deviation between experience and expectations appeared within the criterion of terminological correctness (-0.7), followed by accuracy (0.3) and informedness (-0.1), while the ratings for quality experienced were higher than those for expectations in the characteristics ‘easy to follow’ (+0.1), and ‘fluent and pleasant speech rhythm’ (+0.3) (Mack and Cattaruzza 1995: 45).

Criteria	Assessment / "experience"	Expectations
Informed	4.2	4.3
Easy to follow	4.2	4.1
Accurate	3.8	4.1
Pleasant rhythm	4.2	3.9
Fluent	4.1	3.8
Terminology	3.8	4.5
Average	4.05	4.1

**Table 1.8.** Comparison of ratings of quality assessment and expectations in Mack and Cattaruzza (1995, in Soler Caamaño 2006: 89).

#### 1.1.2.6 Doerflinger (1993)

Doerflinger (1993), on behalf of the Directorate General for Interpretation (SCIC) of the European Union, reported a survey on quality assessment of the conference interpreting service provided by the SCIC: “*qualité de l’interprétation et qualité globale du service fourni?*” (2003: 173, italics in original). The notion of *global quality* here referred to the interpretation as the *result* of a process starting with recruitment, and ending with planning, providing of documents etc. (2003: 173). The questionnaire was administered to interpreters and delegates in 80 conferences of different types. 800 delegates and 700 interpreters returned the questionnaire (2003: 175). The questions related to the following criteria (in brackets the percentage of respondents who answered ‘yes’, or who agreed with the questions): *l’interprétation est satisfaisante (91%)*; *le message passait bien (80%)*; *la terminologie utilisée par les interprètes était appropriée (78%)*; *l’expression des interprètes était agréable (voix – ton – débit – volume) (78%)*; *les sujets étaient bien maîtrisés (74%)*; *le professionnalisme des interprètes paraissait satisfaisant (81%)* (2003 : 175). According to

the author, interpreters criticized their performances more than delegates did. In addition, the survey revealed some problems related to interpreters' documentation and preparation before meetings. Finally, the degree of satisfaction by delegates was higher in meetings with a complete linguistic regime than in meetings with reduced linguistic regime, the latter being more technical (ibidem: 176).

#### 1.1.2.7 Garzone (2003)

Garzone (2003) carried out a survey-based research – a “small-scale pilot study” (2003: 25) – on both quality expectations and quality assessment. Sixteen subjects were administered a short questionnaire consisting of

four of the criteria of quality used by Kurz in her surveys (Kurz 1988; Kurz 1993 based on Bühler 1986), two concerning content and two concerning form: *fluency of delivery*; *pleasantness of voice*; *logical coherence of utterance*; *sense consistency with the original message*, which was reformulated as *fidelity to source text* in order to make it more readily comprehensible for the layman (Garzone 2003: 26).

The subjects were “eight doctors and other professionals (mostly engineers), operating in different technical fields” (2003:26). The respondents first had to rate the criteria according to their quality expectations, then listen to two interpreted versions of the same text to assess them according to the same criteria. One version, the “original” interpretation, was “correct and characterised by pleasantness of voice and good fluency” (2003: 27). The other version was artificially manipulated (“fabricated”), because it was re-recorded by another interpreter who “improved the rendition of the source text but added “a number of hesitations” (2003: 26) and altered the prosody to make it “somewhat erratic” (2003: 27). Therefore this version was “at least as correct [as the first one] but more objectionable formally” (2003: 27). For both expectations and assessment questionnaires “the rating scale was from 1 to 10, the easiest for Italians as it was used in Italian schools for assessing students' performances” (2003: 26). Questionnaire completion was followed by “short interviews” with the respondents (2003: 27).

The results showed that in the expectations questionnaire the criteria *pleasantness of voice* and *fluency of delivery* were considered “less important” than other criteria, while in the assessment questionnaire it was clear that they had a “marked impact on their assessment of other aspects of the two performances” (2003:27) (see Table 1.9).

EXPECTATIONS			
Criteria	Doctors	Other professionals	Average
Pleasant voice	6.12	6.25	6.18
Fluency of delivery	6.12	5.62	5.87
Fidelity of ST	9.00	8.87	8.93
Coherence of utterance	8.87	8.5	8.68
ASSESSMENT			
Criteria	Doctors	Other professionals	Average
<b>Tape A</b>			
Pleasant voice	8.50	7.87	8.18
Fluency of delivery	9.12	8.12	8.61
Fidelity of ST	8.50	8.75	8.62
Coherence of utterance	9.12	8.87	8.99
<b>Tape B</b>			
Pleasant voice	6.12	5.87	5.99
Fluency of delivery	4.5	4.25	4.37
Fidelity of ST	5.62	5.12	5.36
Coherence of utterance	6.00	5.50	5.57

**Table 1.9.** Results of the survey on SI quality expectations and quality assessment (Garzone 2003: 26-27).

In the light of these results, the author put forward two considerations. The first was that the interviews made after the completion of the questionnaires confirmed that the ratings assigned to voice quality and fluency were “ideologically biased”, because they were “based on the idea that after all form is not important, what really counts is content” (2003: 28). The second consideration was a consequence of the first one:

When one speaks in purely abstract terms the evaluation of each single criterion is given in isolation, while in real evaluation processes the different elements overlap and interfere with one another: the performance that is poor in terms of prosody and fluency is perceived as less correct and less coherent, even when in actual fact it isn't. This may be due either to objective difficulty in following the sense of an oral text which is not well presented, because it seems less comprehensible, or to a “psychological effect” which makes the interpretation in Tape B appear less reliable (Garzone 2003: 28).

#### 1.1.2.8 Russo (2005)

Russo (2005) for the first time studied both assessment and expectations of quality by different user groups in the simultaneous interpretation of films. The paper is based on the findings of the studies conducted by Guardini (1995) and Palazzini Finetti (2000). Both studies included questionnaire-based surveys based on Guardini's questionnaire, the questionnaire by Palazzini Finetti being "a modified and slightly shortened version of Guardini's, to allow comparison of results" (Russo, 2005: 7). Guardini's survey examined the simultaneous interpretation of films from English into Italian performed by professional interpreters, while Palazzini Finetti's survey examined the simultaneous interpretation of films from Spanish into Italian executed by interpreting students (Russo, 2005: 7). The questionnaire was administered to interpreting students, film critics and other attendees (clerks, lawyers, doctors, etc.) of two film festivals held in Italy who used the film interpreting service; in total 195 audience members returned the questionnaire. Questions relative to quality assessment constitute the first part of the questionnaire, while the second part is dedicated to the quality expectations (preferences). The quality assessment criteria were: *general quality, voice quality, formal and grammatical correctness, delivery quality, fluency, synchronisation (image-dialogue), dialogue completeness and expressiveness*. The question on quality expectations asked the respondents which of the following features of SI of film they considered important: *dialogue completeness with all details, rendition of the general dialogue content, synchronisation, acting, fluency, pleasant voice, adequate style, explanation of non-verbal elements, for ex. road and shop signs, written messages..., other* (Guardini 1995: 23-26, in Russo 2005).

In the assessment questionnaire relative to the survey conducted by Guardini (1995; in Russo 2005), which included the criterion *general quality*, the criteria that received the highest ratings by film critics was *general quality* (2.9), followed by *voice quality, grammatical correctness* and *word/image synchronisation* (all 2.8), and *fluency of delivery* (2.7). The students assigned the highest rating to *voice quality* (2.8), followed by *general quality* and *grammatical correctness* (both 2.4). The overall average of ratings shows that *voice quality* was the criterion mostly appreciated, followed by *overall quality* (2.8) and *style* and *fluency of delivery* (both 2.5) (Russo 2005: 12) (see Table 1.10).



Aspect of quality	Overall average (n=84-90)	Critics (n=8-10)	Students (n=17-18)	Other (n=48-53)
General quality	2.8	2.9	2.4	2.9
Voice quality	2.9	2.8	2.8	2.9
Style	2.5	2.5	2.1	2.7
Grammatical correctness	2.7	2.8	2.4	2.6
Fluency of delivery	2.5	2.7	2.1	2.6
Word/image synchronisation	2.4	2.8	2.1	2.3
Dialogue completeness	2.4	2.4	2.2	2.4
Expressiveness	2.2	2.2	2.1	2.3

**Table 1.10.** “Ranked average ratings (4-point scale) of quality [assessment] features by user group (based on Guardini 1995)” (Russo 2005: 12).

However, in both Guardini’s (1995) and Palazzini Finetti’s expectations questionnaires (in Russo 2005), completed after the assessment ones, the user groups’ average rating shows that in both questionnaires the criteria which received the highest ratings were *rendition of general dialogue content*, followed by *fluency of delivery* (Russo 2005: 15) (see Table 1.11).

Feature of quality	Guardini (1995)	Palazzini Finetti (2000)	Kurz (1993)
Rendition of general dialogue content	3.4 (1)	3.6 (1)	3.7
Fluency of delivery	3.2 (2)	3.2 (2)	3.1
Word/image synchronisation	3.1 (3)	3.0 (4)	-
Dialogue completeness with all details	2.8 (4)	2.5 (6)	3.2
Adequate style	2.7 (5)	3.1 (3)	-
Pleasant voice	2.7 (5)	2.8 (5)	2.6
Acting	2.6 (7)	2.5 (6)	-
Explanation of nonverbal elements	2.3 (8)	2.2 (8)	-

**Table 1.11.** “Comparison of quality-related preferences in film and conference interpreting (ranked average ratings of importance on a 4-point scale)” (Russo 2005: 15).

#### 1.1.2.9 Catana (2005)

Catana (2005) in her MA dissertation (unpublished) studied the impact of some specific features of voice in SI quality assessment. The objective of her study was to analyse how Italian intonation and diction influenced users’ judgments of SI quality. The questions in the survey were the following (rated on a five-point scale): *overall quality* assessment; *professionalism* of the interpreter; *credibility* of the interpreter; *sympathy/captivation* caused by the interpreter;

*pleasant listening*; *voice* (according to volume, speech rate, intonation, tone); *overall assessment of voice* (from “very harmonious” to “not harmonious”); aspects of *diction* (if any) that might have influenced assessment: open-ended question followed by rating of diction followed by a multiple choice (native accent, expressivity, intonation, stressed vowels; rhythm); self-evaluation of *comprehension* of text; ranking of the features considered for the overall evaluation of quality (native accent, pleasant voice and intonation, fluency, logical cohesion, correct grammar, transmission of general sense; correct terminology; style; correct diction and articulation); the last question (open-ended) was addressed to subjects who obtained a MA degree in interpretation and is about the qualifications of a conference interpreter. The questionnaire was administered to 30 MA interpreting students. The subjects were divided into 3 groups of 10 respondents, in order to evaluate 3 different versions of an interpretation (from Spanish into Italian): the first version with monotonous intonation (intonation was made monotonous by the speaker, who was not a real interpreter and read an interpretation) and correct execution of acute/grave accents of “e” and “o”; the second version with neutral intonation and correct execution of accents of “e” and “o” (read by a professional interpreter); the third version with neutral intonation but incorrect (reverse) execution of acute/grave accents of “e” and “o” (read by the same interpreter of the second version); a completely neutral control version, listened to by all the groups.

According to the results of the survey (2005: 117-202), the monotonous intonation negatively influenced the rating of the *overall quality* of interpretation, the *professionalism*, the *credibility* of the interpreter and his *ability to captivate the audience*. The overall judgment of *voice* and the *pleasantness of listening* were also affected by the monotonous intonation, as was the self-evaluation of *comprehension*. As to the indication (self-evaluation) of the aspects that influenced the assessment of the *overall quality*, for the first version (monotonous intonation) the features that most influenced the evaluation were fluency and correct terminology (ranked first, on average), followed by style; for the second version the ranking was the following: transmission of general sense, followed by fluency and correct terminology; for the third version the ranking was: transmission of the general sense, followed by pleasant voice and intonation and correct diction and articulation. The global evaluation showed that volume and speech were “medium”, intonation was “monotonous” and tone “neutral”. Nonetheless, the analysis of the individual

ratings assigned to each sub-parameter relative to the voice showed that the subjects recognised that the first version was more monotonous than the second and third. However, the third version was considered “melodious” compared to the second one. As to the *diction*, with respect to the first version 80% of respondents found no marked feature in the interpreter’s diction; in the second version, 90% of subjects maintained that there were marked features, since “intonation”, “speech rate” and “expressiveness” were considered “clearly perceptible”; in the third version, 100% of respondents found marked features, mainly the open vowels “e” and “o” (as was actually the case) (2005: 117-202).

#### 1.1.2.10 Collados Aís et al. (2007)

Collados Aís et al. (2007) published a research study carried out by the *ECIS* group (Quality assessment in simultaneous interpreting), coordinated by Collados Aís. This research followed the study by Collados Aís (1998) on the influence of intonation in the quality assessment of SI. The aims of the research were: (i) to analyse quality expectations of users on the basis of “eleven parameters”; (ii) to analyse “conceptualization and its incidence on the quality evaluation of a SI executed by specialised users” (2007: 6). The 11 parameters tested in both the expectations and the assessment questionnaires were: *accent*, *pleasant voice*, *fluency*, *logical cohesion*, *correct transmission of the original message*, *complete transmission of the original message*, *style*, *intonation*, *diction*, *correct grammar*.

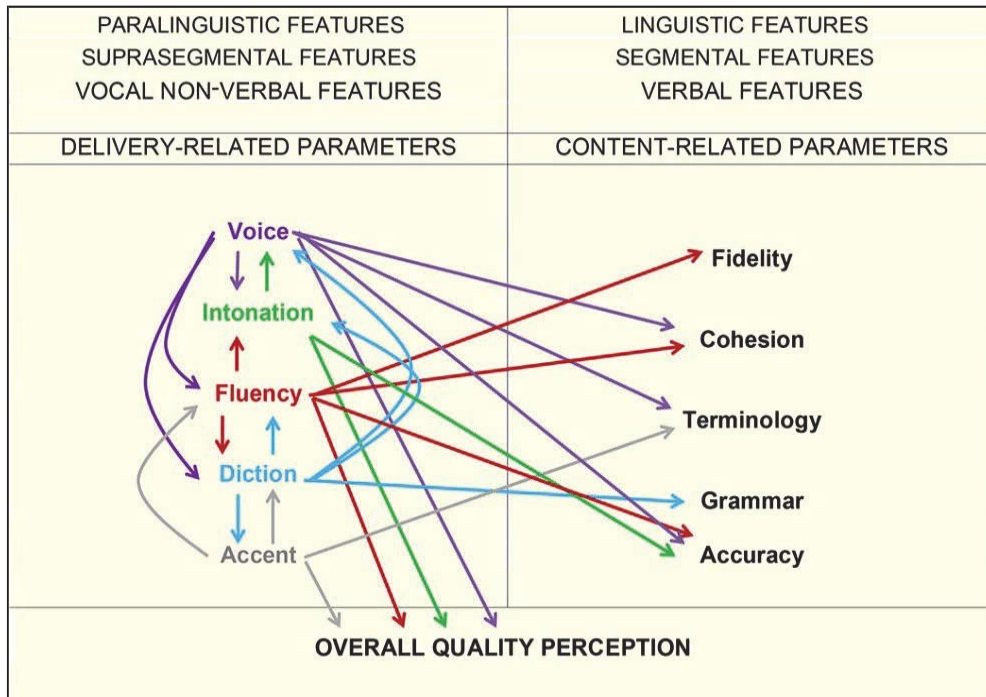
For the assessment session, each quality parameter was artificially manipulated, intervening in the text or in the performance of the interpretation. In total, 12 videos were created for the assessment questionnaire: one for each parameter plus one control video, where the interpretation was not manipulated. As to the manipulations: *intonation* was made “more monotonous”, *pleasant voice* “more nasal”, *style* “more guttural, then less concise”; in *diction* the “articulatory ability was reduced”; in *fluency* “speech rate was modified” and “false starts and self-repairs” were introduced; *accent* was made “native German”, and in *correct grammar* some “most common mistakes” made by “native German speakers speaking Spanish” were introduced; in *terminology* elements of “judicial phraseology” were “replaced with common language words”; in *logical cohesion* “logic discourse markers were removed” or logically “inverted”; in *correct* and *complete*

*transmission of the original*, “information units were distorted or removed respectively” (2007: 12-13).

The questionnaire for the assessment, in addition to the 11 questions relative to the parameters, also included questions on the *overall quality* of the interpretation, the evaluation of the *original text*, impressions of the *reliability* and the *professionalism of the interpreter* (2007: 14). The subjects were 197 university teachers of law for the expectations questionnaire and 164 equivalent (but not necessarily the same) teachers.

The questionnaire on “conceptualization” of parameters was made of 3 parts: the first part included two open-ended questions on one parameter; the second part had to be answered after the subject watched the video relative to the manipulated parameter, and asked the respondent to redefine the parameter after the screening and then rate it on a five-point scale; the third part contained a further evaluation, similar to that in the second part (2007: 16, 255). The subjects of this questionnaire were 32 university teachers of law (2007: 15).

The findings confirmed the hypothesis of the survey: expectations did not match experience (assessment). Manipulation of “verbal” parameters were not easily detected, while the ones relative to “non-verbal” parameters were all detected. Among the “verbal” parameters, the video where the *logical cohesion* had been manipulated received the lowest general score. In the video where the parameter *fluency* had been manipulated, the parameter *correct transmission of the original* received the lowest score (3). However, this score was lower than that received by *correct transmission of the original* in the video where the *correct transmission of the original* had been manipulated (4.33). *Fluency* also had an impact on *diction* and on *logical cohesion*; in both cases it was also the other way round: *logical cohesion* influenced perception of *fluency* and *logical cohesion*. *Fluency* also proved to be interrelated with *correct* and *complete transmission of the original message*. Other “combinations of interrelations” were: *style* and *intonation*; *style* and *pleasant voice*; *diction* and *intonation*; *diction* and *pleasant voice*; *complete transmission* and *cohesion*; *terminology* and *correct grammar*. The video where the parameter *style* had been manipulated received a higher general score than the control video (2007: 220-221). *Intonation*, as well as all other “non-verbal” parameters, influenced the perception of the *overall quality of the interpretation* (2007: 173).



**Figure 1.5.** Proposal of graphic synthesis of the interdependence of “non-verbal parameters” and their “interrelation” with “verbal” parameters, as found in Collados Aís et al. (2007) – the figure is mine.

#### 1.1.2.11 García Becerra (2013)

García Becerra (2013) in her doctoral thesis studied the impact of the first impression on quality assessment of SI. Both expectation and assessment questionnaires included questions about “formal aspects” (*diction, intonation, fluency, voice, etc.*); “content aspects” (*cohesion, style, terminology, etc.*); “fidelity aspects”; the ranking of those aspects; their variation as a function of the age, sex, etc. of the interpreter; other aspects that influence evaluation of expectations. The assessment questionnaire also included questions about a possible impression on the respondent of the interpretation and the interpreter (*competence, confidence, expressiveness, anxiety, pleasantness*) that might or might not have influenced the overall assessment of the interpretation (2013: 651, 654).

The subjects of the survey were university teachers, users of Facebook and Translation and Interpreting students. As to the expectations, university teachers and users of Facebook rated the parameters in a similar way as in Bühler (1986) and in Kurz (1989, 1993), while Translation and Interpreting students assigned higher rates to “formal aspects” (even higher than *fidelity*) (2013: 564). As to findings relative to quality assessment, ratings differed from

those assigned to the same parameters in the expectation questionnaire (2013: 564-565). Subjects confessed that their expectations could have been influenced by the “interpreter’s age, gender and vocal aspects”. In fact, the female interpreter obtained more positive opinions than the male interpreter from all participants, as well as the highest rates in assessment (2013: 566). University teachers and users of Facebook assigned higher rates to the “formal aspects” in the assessment questionnaire than they did in the expectation questionnaire (2013: 567). A “high proportion of subjects” admitted “that their first impression” influenced the “overall assessment of the interpretations”. This was confirmed by the “correlational analysis”, which suggested that the “formal aspects” were at the basis of the “impression” (2013: 567). This analysis also suggested that the administration of the survey might have influenced the subjects’ impressions, mainly the sequence of the three interpretations listened to, “comparison” and “tiredness” (2013: 569). The analysis of the “assignment of adjectives” confirmed the hypothesis on the “perception of the interpreters’ vocal features”. All three groups of subjects defined the male interpreter as “skilled”, “inexpressive” and “insecure”, while the female interpreter was defined as “skilled”, “expressive” and “self-confident”. The “perception of inexpressiveness” might be due to the “frequency range”, the “skillfulness” to the “speech rate” and “silence rate” (2013: 570). “Skillfulness” and “self-confidence” might have “positively influenced” the assessment of *fidelity*, *intelligibility*, *content* and even the *overall quality* of the interpretation, while the perception of “insecurity” might have negatively influenced the assessment of *content*, *fidelity* and *overall quality* (2013: 571). The author concluded that “it looks as if insufficient formal aspects could eclipse remaining parameters in the evaluation mechanism of subjects” (2013: 571).

### **1.1.3 General considerations on the results of reviewed studies on SI quality expectations and quality assessment.**

Studies on both quality expectations and quality assessment of SI mainly present the same “linguistic” criteria devised by Bühler (1986), sometimes with similar names, sometimes with the same criteria grouped into other categories, other times with other criteria adapted to the objective of the study (Soler Caamaño 2006: 101; García Becerra 2012: 55, 74, 84; Collados Aís and García Becerra 2015: 371) (see Appendix 1). “Linguistic” criteria have been divided into ‘content- and ‘delivery-related’ or ‘linguistic’ and ‘paralinguistic’

aspects. The results of expectation surveys show that subjects tend to assign higher ratings to content-related criteria like *sense consistency with the original message*, *logical cohesion*, *correct terminology* and lower ratings to form-related criteria, like *style*, *pleasant voice*, *accent* (Bühler 1986; Kurz 1989, 1993; Kopczyński, 1994; Chiaro and Nocella, 2004; Moser, 1995; Pöchhacker and Zwischenberger, 2010). In some of these cases (Kopczyński, 1994; Chiaro and Nocella, 2004; Pöchhacker and Zwischenberger, 2010) *fluency* was considered the most important among the form-related criteria, even though it ranked third after content-related criteria. The only case in which the criterion *pleasant voice* ranked third, after *sense consistency with original message* and *logical cohesion*, was in the quality expectations in television interpreting (Kurz and Pöchhacker, 1995: 352), and in the case where the subjects were TV professionals, because conference participants assigned *pleasant voice* a low rating, in which it ranked sixth. It is worth mentioning the ‘uniqueness’ of the AIIC survey reported by Moser (1995), where quality criteria were grouped in four categories (*content*, *synchronicity*, *rhetorical skills* and *voice*) and the statistical analysis was made considering all the references (mentions) to the single criteria made by respondents (including the spontaneous ones) throughout the questionnaire. In this case, the results show that, among the “requirements of interpretation”, references in the list of the category *content* ranked first (141), followed by references to *voice* (93), *synchronicity* (89) and *rhetorical skills* (89), *rhetorical skills* being the category that grouped *regular delivery*, *absence of hesitation*, *complete and grammatically correct sentences*, *clarity of expression* (Moser 1995: 8).

The general trend observed in the results of the SI quality expectation surveys also applies to the surveys on SI quality assessment (Gile 1990; Marrone, 1993; Vuorikoski, 1993; Mack and Cattaruzza, 1995; Garzone, 2003; Russo, 2005; Catana, 2005; Collados Aís et al., 2007; García Becerra, 2013). Nonetheless, among these studies, those including a quality assessment questionnaire plus a quality expectation questionnaire show that ratings assigned to form-related criteria in assessment are higher than the ratings assigned to the same ratings in expectations (Mack and Cattaruzza, 1995; Garzone, 2003; Russo, 2005; Collados Aís et al., 2007; García Becerra, 2013). This is even more true in the case of film interpreting (Russo, 2005), where the average rating assigned to voice quality (2.9) is even higher than that assigned to general quality (2.8) (Guardini, 1995 and Palazzini Finetti, 2000 in Russo, 2005). The results of the study by Ng (1992) show that the subjects found it difficult “to isolate speech levels [levels of politeness, in Japanese] as

a separate variable”, in some cases the respondents “seemed to stray from the topic of speech levels to the more general problem of fluency” (Ng, 1992: 39), and “all the subjects were particularly concerned with whether the interpreters had grasped the meaning of the message” (Ng, 1992: 37) even if they interpreted from their mother tongue (English) into Japanese (Ng, 1992: 40). In the study by Mack and Cattaruzza (1995) one of the quality criteria used was *easy to follow interpretation*, a parameter that goes beyond the boundaries between content- and form-related aspects, what may be termed a ‘supra-parameter’. In assessment this criterion is assigned a rating (4.2) which is the same rating assigned to pleasant rhythm (4.2) and quite close to that assigned to fluent interpretation (4.1) (Mack and Cattaruzza, 1995 in Soler Caamaño, 2006: 89). The fact that the manipulation of aspects like *voice*, *intonation* and *fluency* have an impact on the assessment (rating) of *cohesion*, *accuracy*, *terminology*, and *overall quality* in Collados Aís et al. (2007), may be due to the fact that in actual perception the distinction between content- and form-related parameters is not clear or well defined. This is also evident in the study by Garzone (2003), where the manipulation of *fluency* and *intonation* had a negative impact on perception, and therefore on the rating of *fidelity of source text* and *coherence of utterance*. According to Collados Aís and García Becerra (2015: 371-372):

From non-verbal vocal parameters, one can draw deductions about the speaker’s geographic or social origin, emotion or state of mind, personality, competence or credibility, and these deductions largely determine later judgments. These attributes fall within the category of unconscious evaluations that predetermine subsequent consciously-reasoned evaluations.

Soler Caamaño (2006), in her doctoral thesis on quality in specialised interpreting training, studied the evaluation criteria of an examination board of a postgraduate course in medical interpreting (a “case study”). The author analysed the board’s spontaneous (oral) deliberations (transcribed) and identified 67 different quality indicators. After dividing evaluations between “positive” and “negative”, it was found that 13 indicators were used only in negative evaluation:

la falta de reacción al factor sorpresa, la imprecisión en los datos, la repetición de los errores del orador, la mala actitud física en la cabina, la falta de idiomática, la falta de rapidez o agilidad mental, no captar nombres propios (fármacos, siglas), mostrar fallos de cultura general, cometer errores en los datos, respirar de forma que transmita estrés, cometer *Lapsus linguae*, no finalizar las frases, o no hacer un uso correcto del micrófono (2006: 278-279).



By contrast, only 2 indicators were used in positive ones: “*bacerse suyo el discurso original*” and “*ser fiel al contenido*” (2006: 278); the indicator “*voz*” was the only one more frequently mentioned in its positive sense (2006: 279). All the indicators were grouped in six categories:

Por orden de importancia y según frecuencias: prestación, control de la situación, conocimientos específicos, conocimientos generales, competencia traductora, y tácticas y estrategias. Estas categorías no sólo mantienen una relación de vasos comunicantes, sino también una relación jerárquica y de dependencia, de parte y condición de la categoría superior (2006: 280).

On the basis of these categories, the author formulates her proposal for an evaluation protocol for postgraduate examinations, adding a “briefing” for evaluators (2006: 273). According to the author, more than 70% of the comments by the evaluators could be classified according to the Effort Model proposed by Gile: of this percentage, about 70% are related to the production effort, about 15% to the listening effort and 14% to the coordination and memory effort (2006: 281).

At the end of her study, Soler Caamaño also formulates a proposal for a questionnaire of satisfaction, containing the following items: “el mensaje llega bien”; “los intérpretes demuestran un buen dominio del tema”; “la profesionalidad de los intérpretes es satisfactoria”; “la expresión utilizada es agradable (voz, tono, ritmo)”; “la terminología utilizada es adecuada”; “recomendaría/volvería a contratar a los intérpretes”; “explique brevemente el porqué” (2006: 275). This questionnaire is almost the same as Doerflinger’s (2003), as the author herself recognises (Soler Caamaño 2006: 275).

PROPUESTA DE CUESTIONARIO DE SATISFACCIÓN					
		1	2	3	4
1.	El mensaje llega bien	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.	Los intérpretes demuestran un buen dominio del tema	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.	La profesionalidad de los intérpretes es satisfactoria	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.	La expresión utilizada es agradable (voz, tono, ritmo)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.	La terminología utilizada es la adecuada	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
				<b>SÍ</b>	<b>NO</b>
6.	Recomendaría / volvería a contratar a los intérpretes que han hecho el servicio	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>
7.	Explique brevemente el porqué de su respuesta anterior: . .....				

Figure 1.6. “Propuesta – un cuestionario de satisfacción” (Soler Caamaño 2006: 275).

Nonetheless, at the end of her doctoral thesis the author maintains that a study on quality in conference interpreting should be carried out from a “holistic perspective” (Soler Caamaño 2006: 283).

According to Garzone (2003), it is possible that:

in the actual assessment of real instances of interpretation there might be interferences and interdependence between the different criteria separately submitted to, and evaluated by, respondents (Garzone 2003: 25).

Similarly, Iglesias Fernández (2013: 59) proposes that “quality criteria do not seem to be processed separately, but holistically, in clusters of features”.

## **1.2 Conclusion: the need for a new paradigm of quality assessment based on gestalt perception of SI**

Considering the conclusions by Garzone (2003: 25) concerning the “interference” and “interdependence between the different criteria”, and by García Becerra (2013: 571) concerning the possibility that “insufficient formal aspects could eclipse remaining parameters in the evaluation mechanism”; and taking into account the proposal by Iglesias Fernández (2013: 59) regarding the possibility of a holistic processing of quality categories, and the proposal by Soler Caamaño (2006: 283) regarding the need for a “holistic perspective” in carrying out a study on quality assessment in conference interpreting, a theoretical paradigm for such a study has been proposed (see below – chapter 3). It seems that the need for a “holistic perspective” is also due to the fact that questionnaires used in the previous studies on quality evaluation (both expectations and assessment) have not changed substantially over time, as the chronological and contrastive table in Appendix (1) shows.

## **1.3 Approaches to the concept of quality in simultaneous interpretation**

Taking into account the conclusions reached after the review of studies on the quality of SI (see above – sections 1.1.3 and 1.2), a general overview of the approaches to the concept of quality in SI was considered, in order to understand the points of view from which the quality-related aspects were seen. To this aim, a recent article by Collados Aís and García Becerra (2015) was considered as the main source for this section.

Surveys on quality expectations and quality assessment of interpretations based on questionnaires built on quality criteria constitute the majority of recent empirical studies on quality in Interpreting Studies. This line of research, initiated by Hildegund Bühler (1986) has been fruitful (cf. Collados Aís and García Becerra 2015: 376).

The author of the first survey on quality expectations (Bühler 1986) herself raised the question of who should judge the quality of an interpretation: interpreting teachers, professional organizations (through exams), professional interpreters, employers or end-users (*ibidem*: 231). In her study the author assumed that the system adopted at that time by the AIIC, based on the evaluation by professional interpreters, “also reflect[ed] user expectations and needs in its evaluation standards” (*ibidem*: 231). Danica Seleskovitch (*ibidem*: 236), commenting on this approach, maintained that an interpreter’s performance should be assessed by a precis-writer of the minutes of a meeting where interpretation was fundamental to understanding its proceedings. The difference between expectations and assessment appeared blurred. The majority of surveys on quality expectations (see above – section 1.1.1) were responded to by end-users (Kurz 1989, 1993; Meak 1990; Kopczyński 1994; 1995); others were addressed to interpreters (Chiaro & Nocella 2004) or professional interpreters (Pöchhacker and Zwischenberger 2010); one study was addressed to end-users that were also expert intermediaries of the service: this is the case of the television experts for a survey on quality expectations on a television interpreting service (Kurz and Pöchhacker 1995). As to the studies on quality assessment, the majority of them had end-users as assessors (Gile 1990; Ng 1992; Marrone 1993; Vourikoski 1993; Mack and Cattaruzza 1995; Garzone 2003); one had both end-users (delegates) and interpreters as assessors (Doerflinger 1993); two had end-users and interpreting students (Russo 2005; García Becerra 2013); and one had law professors as judges of quality (Collados Aís et al. 2007).

In the approaches to quality in simultaneous interpreting, Cartellieri (1983; in Collados Aís and García Becerra 2015: 369) proposed that “the perception and evaluation of quality often depend on the subjective judgment of the evaluator according to the situation and the premises”. Gile (1990, 1991; in Collados Aís and García Becerra 2015: 369) proposed the distinction between quality and success (or failure) of an interpretation. Pöchhacker (1994; in Collados Aís and García Becerra 2015: 368) proposed the question “what quality, for whom” as a starting point to approach a concept of quality,

which therefore depended on the “circumstances”. According to Gile (1995a; in Collados Aís and García Becerra 2015: 369), quality expectations of different users could not be considered the only way to define the quality of the message transfer; and that quality also depended on the role of the interpreter (“Sender’s alter ego” or “facilitator of communication”). Moser (1995; in Collados Aís and García Becerra 2015: 369) considered only two approaches to evaluation: that based on the standard imposed by interpreters themselves, and that based on the opinions, needs and expectations of users. Moser-Mercer (1996: 46; in Collados Aís and García Becerra 2015: 368-369) proposed the concept of quality as a function of the purpose of the communication. Thus, it can be evaluated from different perspectives: that of interpreters, employers, users, intermediaries and trainers. Garzone (2003; in Collados Aís and García Becerra 2015: 370) argued that the interpreter is the only guarantor of quality, being the only one responsible for the result of her/his interpretation. On the contrary, Kalina (2005; in Collados Aís and García Becerra 2015: 369, 370) maintains that the quality of result depends on collaboration among all actors involved in the interpretation. However, the interpreter, by analyzing expectations and demand of different groups, has to select priorities: fidelity or intelligibility, correct language usage or precision, elegant style or complete transmission of the message.

Quality of interpretation has also been related to rules. According to Shlesinger (1989:112; in Collados Aís and García Becerra 2015: 370), definitions of quality tend to reflect the rules internalized by interpreters, since they are taught in training programs, or constitute the ethical codes and the professional standards adopted by interpreter associations (e.g. AIIC) or institutional services (e.g. SCIC). Nevertheless, these norms can hardly be elicited in the context of conference interpreting. Harris (1990; in Collados Aís and García Becerra 2015: 370) argues that norms vary according to the interpreting setting; however, the interpreter has to be an “honest spokesperson”. Straniero Sergio (2003: 171; in Collados Aís and García Becerra 2015: 370) considers “norms”, in media interpreting, the “emergency strategies” (Straniero Sergio 2003: 140) adopted by professional interpreters, recurring in the performances of an interpreter, as translational patterns. However, “interpreters’ performances are closely dependent on the working conditions, which, to a large extent, determine the quality of a given SI” (ibidem: 171). In conclusion, the author proposes a concept of “achievable quality” of an interpretation, depending on the “conditions in which the

interpreter has to translate a particular text” (*ibidem*). The incidence of the context, mainly the working conditions in the resulting quality of television interpreting has been noted by different authors (Straniero Sergio 2003, 2007: 541; Kurz and Pöchhacker 1995; Mack 1999).

The incidence of contextual aspects has also been the object of studies on quality as a process; for example, the problem triggers in the original speech (Gile 1995a; in Collados Aís and García Becerra 2015: 373), the visual contact (Balzani 1988; in Collados Aís and García Becerra 2015: 373), the length of turns in the booth (Moser-Mercer, and Korac 1998; in Collados Aís and García Becerra 2015: 373), and the availability of the original text (Lamberger-Felber 1998, 2001; in Collados Aís and García Becerra 2015: 373).

Among the personal aspects having an impact on the interpreting process are the optimum balance between the different efforts (listening, memory and production) according to Gile’s model (1988: 15ff; in Collados Aís and García Becerra 2015: 373), or the use of strategies (Riccardi 2003).

The “first empirical study of quality as product” was Bühler’s survey on users’ expectations (Collados Aís and García Becerra 2015: 374). The author defined a series of parameters (used in the survey – see above, section 1.1.1.1) “in which she proposed criteria that could influence the quality of an interpretation” (Collados Aís and García Becerra 2015: 371). The same criteria have been used in subsequent studies of either expectations or assessment (see above – section 1.1), with few modifications (addition or elimination of criteria) made according to the purpose of the study; for example, intonation and diction were added, while pleasant appearance was eliminated, together with other aspects not relevant to the simultaneous interpretation, which was the mode more frequently used for this kind of survey (*ibidem*).

Collados Aís (1998) initiated a line of research based on the study of one parameter (intonation) under experimental conditions (“controlled interpreting environment”) (Collados Aís and García Becerra 2015: 375, 376). Later, this methodology was further developed and extended to other parameters (e.g. Garzone 2003; Pradas Macías 2003; García Becerra 2006; Collados Aís et al. 2007; Holub 2010; Rennert 2010; in Collados Aís and García Becerra 2015: 376). Subsequently, other researches have been conducted “to identify users’ perceptual standards”, which have an impact on the quality of interpretation, “and possible interaction among different parameters” (Collados Aís and García Becerra 2015: 376). “In light of the apparent contradiction between expectations and evaluation in previous

works, vertical studies of the parameters have been undertaken” (Pradas Macías 2003, 2006; Ahrens 2005; Collados Aís et al. 2007; Collados Aís et al. 2011; in Collados Aís and García Becerra 2015: 376; Zwischenberger and Behr 2013). In these studies, in order to isolate criteria to better study their impact on perception of quality, researchers used corpora of actual interpretations (e.g. Ahrens 2005), provided instructions to interpreters (e.g. Collados Aís 1998; Pradas Macías 2003), or manipulated them through software programs (e.g. Holub 2010; Rennert 2010) (in Collados Aís and García Becerra 2015: 376). In some of these empirical researches, ‘comprehension of interpretation’ was introduced as an aspect having an incidence on quality (Holubb 2010; Rennert 2010; Cheung 2013; in Collados Aís and García Becerra 2015: 377; Christodoulides & Lenglet 2014). In other researches, other aspects related to quality have been studied, among them, the influence of English as a lingua franca on the effectiveness of interpretations (Reithofer 2013; in Collados Aís and García Becerra 2015: 377).

The authors that represent the main source for this section maintain that “one of the most significant results” of research on interpreting quality is that “the formal criteria (see above – section 1.1.1.1; figure 1.5) for interpreting have a greater weight in the evaluation of a specific interpretation than a priori reasoning would suggest” (e.g. Collados Aís 1998; Garzone 2003; Pradas Macías 2003; Collados Aís et al. 2007; in Collados Aís and García Becerra 2015: 377). Such influence of formal criteria is fundamentally “evidence” for the “perception” of quality, and this perception is represented by the “interaction” or “overlapping” of parameters; hence, the “indivisibility of form and substance in the majority of criteria, far from the rigid distinction between the verbal and the non-verbal” (Collados Aís and García Becerra 2015: 377-378). Results of recent studies show that formal aspects also have an incidence on the comprehension of an interpretation (Rennert 2010; in Collados Aís and García Becerra 2015: 378).

Viezzi (1996: 79-105) proposed a theoretical model for evaluation of quality in interpretation, based on translation and interpretation studies, pragmatics and textual linguistics. The quality criteria defined in this model are the objectives of the interpretation activity, which stem from the conception of interpretation as a communicative act resulting from an intercultural, interlinguistic and text production activity. The criteria are “equivalence”, “accuracy”, “appropriateness” and “usability”. The parameters equivalence

and accuracy concern the relationship between source and target text, while appropriateness and usability concern the relationship between target text and receivers of interpretation (considered within the specific communicative situation). More specifically, equivalence refers to the fulfillment of the communicative intention of the source text; accuracy refers to the complete transmission of the information in the source text (through the interpreter's activity of selection of the information relevant to the communicative goal of the text); appropriateness refers to the interpreter's ability to adapt the target text to the audience and the context; usability refers to the interpreter's ability to facilitate the comprehension of the interpretation through logic, cohesion, concision, fluency, and use of voice. The four parameters are interrelated, and usability could be considered the synthesis of all the aspects of quality.

In conclusion, it is worth noting that, according to Viezzi's proposal, the criterion that synthesizes all aspects of quality is precisely the only one that concerns both form and substance; the same conclusion drawn by Collados Aís and García Becerra (2015: 377-378) after years of empirical research.

## **1.4 Prosody and simultaneous interpretation**

Considering the incidence of aspects like voice, fluency, accent, etc. on the assessment of other aspects of quality, or overall quality of SI, attention was paid to empirical studies on simultaneous interpreting prosody and its correlation with other aspects of speech.

### **1.4.1 Cognitive rhythm and speech production**

With the aim of better understanding the incidence of prosodic features on the total perception of speech, the psycholinguistic experimental researches by Goldman-Eisler (1968) were studied. The author approached speech production from a physiological and cognitive point of view, considering the involvement of the entire human body in the speech activity. Therefore Goldman-Eisler's contribution was also considered relevant to comprehending speech perception from a gestaltic point of view.

Goldman-Eisler (1968) conducted a series of "Experiments in spontaneous speech". She considered human speech as an activity where "physiological and logical operations" are linked, since it is the result of the "co-operation" of "organs the original function of which is to serve vital biological needs" (i.e. "breathing, eating and crying in emotion") and "highest levels and latest organs of the nervous system". In the end, "the entire body"

is involved in the speech act production. The organs that “were developed at an earlier stage of evolution for life serving activities” then “turned to a new use” to achieve human speech. Nevertheless, even after the development of speech, these organs “continue to serve the original vital functions”; therefore, speech “must compete with these functions in the use of these organs”. In addition, the speech act is executed under “voluntary control”, and so its movements are “controlled from the cortex”, while the operations of “swallowing, gagging, vomiting, suckling, breathing, chewing, laughing or sobbing, etc.” are “activated at the cerebellar level”.

If Goldman-Eisler’s approach to human speech could be defined as evolutionary, her approach to language and, more specifically, to phonology is structural and temporal. In fact, the development of the “process of symbolic representation” derived from a “continuous stream of primitive sound” (the one of “primitive cries and calls”, for example) that over time became “separated into highly controlled discrete elements of well-defined and distinct structure easily to be recognized”. The product of a speech act is the sequence of such structures of sounds, organized “in such a way to convey a message and a meaning”.

Not only does the organization of sounds of language respond to a structural conception, but also the entire process that produces speech was conceived as a complex structure of phenomena that are temporally integrated. The phenomena are: i) the “linguistic systems” (related to “phonemic, lexical and grammatical, i.e. syntactic operations”); ii) the “physical, physiological and neuro-physiological systems” (related to “respiratory, muscular and electrical activity”); and iii) the “temporal phenomena, which include durations of activity by any of these systems and of the gaps of inactivity which interrupt and alternate with activity”.

The temporal aspect then is the one that integrates all the others, and is the one that Goldman-Eisler mainly used to observe spontaneous speech through her experiments, perhaps because it could be measured; in fact, the main variables considered were breathing, pauses, hesitations, rhythm and muscular tension.

The author conducted her experiments on spontaneous speech (“conversations, discussions and indirect interviews and verbal assignments”), and not on “linguistic productions evoked in the laboratory under controlled conditions” (ibidem: 10), because she wanted to observe the speech more as “a sum of movements”, with an “internal unity of signification”, and less an



isolated “gesture”. Moreover, in the speech “organized on the spur of the moment”, “the discourse was left to follow a natural course” (ibidem: 10).

Nevertheless, Goldman-Eisler in her experiments also used reading and simultaneous interpreting to compare different cognitive activities, but all related to speech production.

Among Goldman-Eisler’s experiments concerning simultaneous interpretation, relevant to this study, there is the study of hesitations in speech (ratio pauses/speech) related to the structure of sentences (quantified through a “subordination index”). The results showed that both in spontaneous speech and in simultaneous interpreting, the two above-mentioned parameters seem “independent” (ibidem: 78). In simultaneous interpreting, “syntactical operations” (i.e. the simplification or complication of the sentence structures by the interpreter with respect to the source text) “were not reflected in the time of hesitation pauses”, and “any increase of pause time was due to cutting loose from the sentence structure of the received input and generating a different one” (ibidem: 78).

By analysing samples of spontaneous speech, readings and simultaneous interpretations (both having a read text and a spontaneous speech as source texts), the author reached the conclusion that rhythm is a function of overall fluency. She analysed the temporal patterning of these speeches in terms of the significance of the relation between pausing time in the hesitant periods and the time of speaking in fluent periods. The speech samples having a significant relationship (chi-square test,  $p < 0.001$ ) between hesitant and fluent periods were defined as “rhythm”, while those with a negative relationship were defined as “non-rhythm”. In all the types of above-mentioned samples she found that, for a period to be defined as rhythm the speaking time had to be more than about twice the time of pausing (ibidem: 82-86).

By analysing reading in particular, the author found that the reading of the same text by different readers, or even by the same reader at different speeds, had different rhythmic patterns, based on the percentage of pauses in speech, in similar proportions to those found in spontaneous speech. The author recognized that the cognitive activity involved in reading was not as intense as in spontaneous speech. However, difference in rhythm structures found in readings led the author to infer, according to her approach, that a cognitive activity was also present in reading, due to the interpretation of the text, even if this had an inherent rhythm (ibidem: 86-87).

In the analysis of simultaneous interpretations (ibidem: 87-89), no relation was found between the temporal rhythm of output (interpretations) and input (source text). However, a “highly significant relation” was found when temporal rhythm of interpretations was related to the difference between output and input rates. In the case of interpretations with a temporal rhythm, this was due to interpreters that most extended pause time with respect to that in source texts, these having a more rapid speed, among all others source texts. At this point, pause percentage was calculated separately in both passages classified as high pausing (more than 50% of pauses) and high fluency (less than 15 % of pauses) periods; as a result, in both of these periods, pausing turned out to be highly related to rhythm.

At this point, the author had to revise her psycholinguistic hypothesis, which considered pausing related to “planning” and speech related to “achievement” or “execution”, because it was found that pauses were created even in the most fluent periods. This phenomenon was then related to a “feedback-control” over the speech process, playing an “inhibitory” role, while from the behavioural point of view, it was considered the manifestation of a “totality of attitude”, a “specific neurophysiological set pervading the whole situation” of speech production. The conclusive proposal was that “global tonigenic activation” lies in the background and engenders a “selective process” in speech production (ibidem: 90-93).

#### **1.4.2 Simultaneous interpreting prosody**

In order to understand similarities and differences between spontaneous and simultaneous interpreting speech production and perception, Interpreting Studies focused on prosody were considered in this section.

Shlesinger (1994) perceptually noted that intonation in simultaneous interpreting speech was different from that of spontaneous speech. Since it was not possible to compare spontaneous speech with interpreting speech from the same speakers, she selected ten simultaneously interpreted passages (six from English into Hebrew and four from Hebrew into English), transcribed them, and then asked the interpreters to read their respective passages aloud, in order to use them for a perceptual survey on the effect of interpretational intonation on comprehension and recall. The reading was executed at least three years after interpreters had executed their interpretation, only two of the ten interpreters recognized the passage. To study the difference between SI intonation and reading-aloud intonation,

Shlesinger used four categories, taken from Halliday (1967; in Shlesinger 1994): tonality, tonicity, tone, duration and speed. With reference to tonality she perceptually measured the “pauses within constituent” and the “tentative pause in final position”; for tonicity, the “stress incompatible with semantic contrast” (Brown 1974; in Shlesinger 1994: 231); for tone, the “nonstandard matching of pitch movement and discourse pattern: low-rise, non-final pitch in final position”; for duration and speed the “lengthening/holding of segment” and “acceleration”. Whenever these salient features were also detected in the read-aloud passage, they were not counted in the equivalent interpreted passage.

According to the findings, the interpretations had a notable number of “pauses within grammatical structures” or in “unnatural positions”, while “pauses at sentence boundaries”, which were present too, “tended to be tentative rather than final, and were often coextensive with a low-rise intonation” (*ibidem*: 229). Occurrences of “stress incompatible with semantic contrast” were also found in a remarkable number, while the occurrences of “lengthening/holding of segment” were few and only one acceleration was found in ten passages analysed. Two groups of subjects, “matched for fluency in the languages concerned, and for familiarity with interpreting” (*ibidem*: 228) listened to the recordings of three texts: Group A (8 subjects) listened to the interpreted versions and Group B (7 subjects) listened to the read-aloud ones. Both Groups, after listening to the passages answered the same questions (3 for each passage, a total of 9 questions) “on recall and comprehension”.

The author assumed that, since the speakers (either as interpreters or as readers) and also the texts as well as the “format (a recording)” were the same, “the sole distinguishing factor would be the intonation”. She also hypothesized that the “tentative pauses” and the preponderance of the low-rise non-final pitch would have negatively affected comprehension, because they were against the listener’s expectations and, consequently, perception would have been “temporarily disrupted” (Cutler 1987; in Shlesinger 1994: 231). The correct answers given by Group A represented 20% of the total, while those given by Group B represented 41% of the total; thus, the subjects who listened to the read mode had a better comprehension and recall. Thus, the hypotheses were confirmed. However, Shlesinger pointed out that the “perception of conference participants, who are aware of the frame and

familiar with the background, is considerably higher”, but her subjects did not have it (Shlesinger 1994: 234).

Williams (1995) acoustically analysed examples of anomalous stress produced by a professional interpreter at a live conference (Swedish-English) in relation to stress patterns in the speaker’s input. She found that words stressed in SI did not appear to be directly related to semantic or pragmatic features in the incoming message.

Ahrens (2005) acoustically analysed the prosody of a SI corpus (72 min, English-German). She found that the German interpretation, with respect to the original English text, had a low number of pauses, but with a higher duration; information units in the SI were also more numerous and shorter (frequently made of one or two words); in the SI “almost every single word was stressed” (Ahrens 2005: 72) and the proportion of rising, level and rise-level contours in final pitch movement was high. Frequent pauses and non-final pitch contour in SI observed by Ahrens (2005) confirmed Shlesinger’s (1994) results.

Collados Aís (1998) carried out an experiment on the impact of monotonous intonation and sense consistency with the original message on overall quality assessment of a simultaneous interpretation (German-Spanish). Both intonation and sense consistency with the source text were artificially manipulated. Results confirmed the initial hypothesis, i.e. the monotonous intonation had a significantly negative impact on overall quality assessment of the interpretation, while sense inconsistency did not.

Following the same method adopted by Collados Aís (1998) and using the same materials, Pradas Macías (2006) conducted an experiment on the impact of silent pauses on the assessment of fluency, other aspects, and overall quality of interpretation (German-Spanish). Results showed that the two texts with additional (artificial) pauses received a lower mean rating with respect to the control text. Moreover, the low fluency had a negative impact on the assessment of correct rendition of sense and on the perception of intonation, but not on the overall quality of the interpretation, because the control video received the lowest rating for overall quality. Differences between two experimental conditions and one control condition did not attain statistical significance.

Tohyama & Matsubara (2006) conducted an experiment on the relationship between listener-friendliness and the length of silent pauses in SI (English-Japanese). Results showed that in simultaneous interpretations with

slow speech rate, short pauses had a notable positive impact on the evaluation of interpretation, while in those with a high speech rate, “the influence of pause length on the listeners’ impression was small” (Tohyama & Matsubara 2006: 896) - the evaluation was based on the easiness/difficulty of listening. In both cases of high and low speech rate, the interpretations that received high evaluations “had the characteristic that the speak-stop state [...] was stable and rhythmic” (ibidem: 896). In conclusion, the perceptual incidence of silent pauses on the listeners’ impression was small “if the interval and the distribution of those pauses are stable” (ibidem).

Christodoulides (2013) analysed similarities and convergence of speech rate, mean pitch and pitch range between speaker and interpreter, in a corpus of simultaneous interpretations (English-French) of the European Parliament (committee meetings, plenary sessions and press conferences), finding that interpretations had longer and less frequent silent pauses, with a more variable speech rate and a narrower pitch range than source texts.

Christodoulides and Lenglet (2014) analysed prosodic correlates of perceived quality and fluency in a simultaneous interpretation (German-French) and a reading of the same interpreted text (after transcription) by the same interpreter. Results showed that interpretations had longer silent pauses, more frequent filled pauses, more reformulation-related disfluencies, more variable articulation rate and less lively intonation than the read speech. All these aspects, together with more frequent pauses in non-syntactic boundaries, had a negative impact on the perception of fluency which, in turn, affected the perception of accuracy. Results from the listening comprehension test (cf. Christodoulides and Lenglet 2014) were also related to the difference between the two groups of subjects; in particular, the more fluent interpretation was perceived as more accurate by translation students than by economics students; this means that for economics students the poor fluency had a slighter impact on perception of accuracy. However, translation students scored better than economics students regarding objective comprehension. The authors hypothesized that, on the one hand,

translation students processed the speech at a more superficial level and hence, were more affected by perturbations of the prosodic structure of the speech. The students of economics could use their prior knowledge to process the speech content at a deeper level and make inferences to compensate for disturbing prosodic variations (ibidem).

On the other hand, the better comprehension by translation student could be due to their ability to “capture the gist of speeches in their notes thanks to an elaborate notetaking” (ibidem).

### **1.4.3 Structure and rhythm in speech**

Considering the results from the above-mentioned studies, it could be assumed that Goldman-Eisler’s (1968) proposal of structural activity of “linguistic”, “physical”, “physiological” and “neuro-physiological systems”, plus “temporal phenomena” (duration of activity vs. inactivity) (Goldman-Eisler 1968: 6-10), all involved in speech production, is also true for speech perception. According to Goldman-Eisler, the structural activity manifests itself in temporal sequences, hence the proposal of “cognitive rhythm” (Goldman-Eisler 1968: 6-10; 90-93). Now, the word ‘rhythm’, in its current definition, means a succession of accents according to a common pattern, therefore, it is related to the notion of ‘temporal aspect’. However, according to the philological proposal by Benveniste (1966: 333), the original meaning of the word ‘rhythm’ was ‘form’, and it was not related to the temporal aspect, but rather to an ‘organization’, ‘configuration’ or ‘display’ of parts in a discourse.

This proposal is consistent with the original meaning of the terms ‘structure’ and ‘system’ with reference to language, where all the aspects are interrelated. The term ‘system’ was frequently used by De Saussure (Albano Leoni 2009: 155), while the term ‘structure’ officially entered in the linguistic terminology through the Prague Theses in 1929 (Trubeckoj 1929; in Albano Leoni 2009: 90, 155). The meaning of ‘structure’ and ‘system’ is not so far from that of the word ‘Gestalt’, since they focus on the relationship between the parts and the whole, where the whole is not the sum of individual parts, but something more, where the parts are determined by the whole and the mutual relationships among them (Albano Leoni 2009: 155-156). In fact, in the studies on the birth of structuralism and its theoretical foundations, the Gestalt was “evoked”; and in psychology, ‘Gestalt’ and ‘structure’ have a very similar usage (Albano Leoni 2009: 156).

In the light of what has been discussed, Goldman-Eisler’s “cognitive rhythm” should be marked by a formal, and not temporal aspect. However, from Plato on, the meaning of rhythm changed from spatial disposition into ordered sequence of fast and slow movements, and the notion of measure entered in its definition (cf. Benveniste 1966: 334-335). Moreover, the

structural (systemic or gestaltic) character of speech perception concerned only the first part of the structuralist stage of the history of phoneme, inherited from the previous psychologicistic stage (cf. Albano Leoni 2009: 86-109). The successive development of phonology, up to the current cognitive stage, has focused on the segmental paradigm, which has dominated not only the field of the signifier (i.e. the linear succession of discrete sounds), but also that of the signified (i.e. the linear succession of discrete signs) (cf. Albano Leoni 2009: 110-126).

The theoretical paradigm (see below – chapter 3) proposed in this research study recovers the original meanings of the words ‘form’ and ‘structure’, at least as far as the analysis of speech perception is concerned, as will be better explained in chapter 3.





## Chapter 2

### ORenesit CORPUS

#### 2.0 Introduction

Within the Translation Studies, the Corpus Linguistics (CL) developed thanks to the diffusion of Descriptive Translation Studies (DTS), which initiated the research of new methodologies and empirical paradigms of analysis with the aim to formulate new theories (Valentini 2009: 33).

The descriptive approach to translation considers the translated text (or target text) as an independent text (from the source text); thus, translation has a proper function in the historical, social and cultural context that originates it, where the translator's role is significant due to her/his choices that may depend on his/her perspectives, expectations and ideology (Venuti 1995; in Valentini 2009: 33). This approach became the model of the Skopos Theory (Vermer 1996), according to which each translation can be evaluated on the basis of the objectives it is supposed to achieve in a given context and for a specific receiver (in Valentini 2009: 33). DTS are based on the need for a rigorous and systematic analysis of data, the rejection of definite categorizations and abstract theories, the complexity of linguistic phenomena, an emphasis on interdisciplinary approach and the use of both qualitative and quantitative approaches (Laviosa 2002: 10; in Valentini 2009: 33).

Even if the descriptive studies paved the way to the introduction in translation studies of the methods developed by computational or corpus

linguistics, the two approaches are slightly different. Indeed, even if both approaches include empirical research based on the collection of authentic material and the application of probabilistic measurements; descriptive studies also use extralinguistic information to support real data, while corpus linguistics considers each research as a theoretical formulation that can be limited to the single study (Valentini 2009: 34). As a consequence of this approach, in corpus linguistics, data, description, theory and methodology are interdependent factors; such interrelation is considered as a continuous process linking the corpus building to the formulation of hypotheses, their verification and the consequent proposal of theory (Laviosa 2002: 8; in Valentini 2009: 34). In second moment, this theory can be considered as a new hypothesis to be tested through the development of the initial corpus (Valentini 2009: 34).

Mona Baker (1993, 1995) proposed to apply the means and the methods of the corpus linguistics to the translation studies, in order to study the relationship between linguistic structures and their use by translators. In this way, translation studies shifted the focus of research from product to process (in Valentini 2009: 35). Valentini observes that recent developments of translation studies are marked by the discovery of the context of the translation process (Baker 1996, 1999; Tymoczko 1998; Mason 2001); indeed, mere statistical descriptions of how language is used are no longer enough to formulate effective theories. Thus, the traditional quantitative approach is integrated with qualitative methods taking into account the researcher's subjectivity on the basis of extralinguistic factors as the receiving public and the professional translation process (Baker 1999: 25; in Valentini 2009: 36). Corpora-based researches constitutes the most diffused methodology in current translation studies (Valentini 2009: 36). In the last years, with some delay, this methodology has been more and more adopted also in audiovisual translation studies and interpreting studies. The reasons of this delay are to be found in the time needed both to collect the material, and to build multimedia corpora, considering the technology needed for the identification of audiovisual basic units and the alignment of multimodal data (ibidem).

In general, a research study where data extracted from a corpus are used is defined corpus-based study. However, in a more specific use of the term, in a corpus-based study the corpus is used to test or explain theories preexisting the corpus, and not a theory that stemmed from the corpus. A corpus-based study, in general, reveals the characteristics of the structure and

the use of language, from general and abstract patterns, to specific and idiosyncratic schemes; in addition, the basic units of analysis are not questioned. This approach, where both qualitative and quantitative methods are adopted, can be used to formulate research hypotheses, related to theories. In addition, both top-down and bottom-up methods can be adopted (Tognini-Bonelli 2000: 155-156). A different approach to corpus data led to the creation of the term corpus-driven study. In this case, the researcher considers the data as a whole, for general descriptions of the topic analysed; moreover, the representativeness of the sampling, its size, the criteria adopted for the collection of the material, etc. are relevant to the type of research.

## **2.1 The corpus of reference: CorIT (Corpus Italiano di Interpretazione Televisiva)**

In order to better study the gestaltic perception of TV broadcast simultaneous interpretation, a corpus was built up. Thus, the present research is a corpus-based interpreting study, where the main reference corpus is the CorIT (Television Interpreting Corpus), created and developed by the School of Modern Languages for Interpreters and Translators (SSLMIT) at the University of Trieste:

CorIT presently consists of more than 2,700 interpretations (Interpreted Texts – IT) both consecutive and simultaneous mode, from various foreign languages into Italian, delivered on television during various types of broadcasts. Owing to its structure, it may be defined as an open, multimedia, “partiellement parallèle” [partially parallel] corpus (Falbo 2012: 156).

Nonetheless, CorIT “appears much more as a multimedia archive” or a “collection of multimedia texts” than a real “corpus”, because it has not been completely transcribed; only after transcription it will “constitute a digital corpus, fit for digital interrogation” (Falbo 2012: 156). However, the term “corpus” is used by its creators and developers to refer to “the aim of the project” (ibidem). CorIT is “constantly updated with new recordings” (ibidem), therefore it has the characteristics of “open corpus” and “monitor corpora” (cf. Barbera et al. 2007: 51-52 in Falbo 2012: 156), because it is made of different subcorpora. One of these is the presidential debates subcorpus (A), which is replaced by a new subcorpus (A1) when a new US or French presidential debate is added, hence the idea of a series of “monitor corpora”. The author concludes suggesting the use of “open corpus” as “the most efficient definition in operational terms”, while it can be simply considered as

a “corpus *in fieri*” or “corpus in progress”. Another subcorpus of CorIT is made up by “original texts (OT) delivered in a language other than Italian”, but only a part of them is available, for this reason the “CorIT can be defined as a partially parallel corpus” (Falbo 2012:157). CorIT is a “multimodal corpus” but also a “speech corpus” because the software used to transcribe it, WinPitch<sup>1</sup>, allows to automatically align transcripts to audio and video tracks. WinPitch also allows to automatically save transcripts in .rtf format, therefore “CorIT is also a spoken corpus and may be interrogated as such” (ibidem: 34).

The CorIT project stemmed from the “need to study interpreting from a comprehensive perspective for the sake of ecological validity” (Falbo 2012: 155). Moreover,

CorIT was created with the aim of tackling an issue highlighted by many (Shlesinger 1998), namely overcoming the limitation of single case studies and focusing on more comprehensive analyses. CorIT may not represent an exhaustive picture of the interpreting world (Falbo 2001), nevertheless, it certainly provides an opportunity to investigate television interpreting, an interpreting sector that is increasingly gaining momentum (Falbo 2012: 178)

CorIT has engendered many MA theses (cf. Falbo 2012: 182-185) and many studies by its creators and main developers (e.g. Straniero Sergio 2007; Falbo 2001, 2005, 2012; and Dal Fovo 2011, 2014).

### **2.1.1 Transcription conventions**

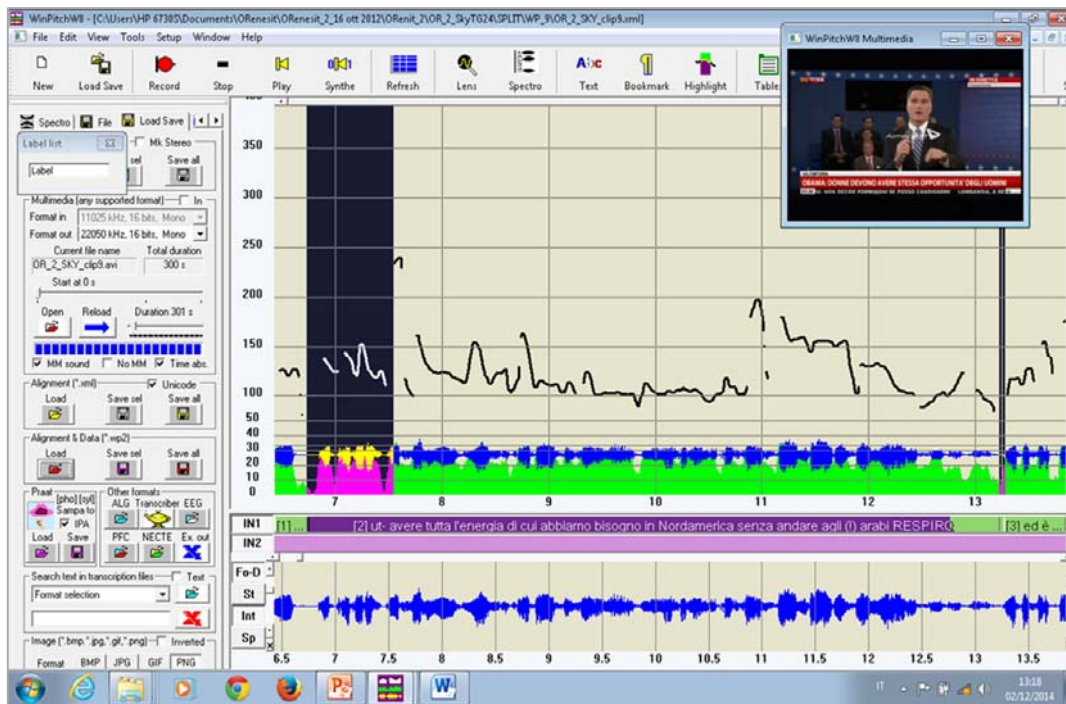
Transcription is fundamental for the analysis of spoken discourse; it represents the first analysis of a spoken discourse, which can be the object of further analyses (Falbo 2012: 174-175). However, as a first analysis, the transcription always reflects the purposes of further analyses, from which depends what is considered necessary to transcribe (ibidem).

The transcription of an audiovisual recording is a “secondary representation” of an “original event”, of which the recording itself is a “first representation” (Falbo 2012: 174-175). However, the second representation of the transcription of a spoken discourse is limited by the “categories created for written texts” (ibidem: 175). With the idea of overcoming this limitation, it was decided to transcribe the interpreted texts through WinPitch, since this software permit to have audio and video tracks always accessible for analysis (activity) of transcription (ibidem). In addition, audio and video tracks are

---

<sup>1</sup> WinPitch Pro, by Pitch Instruments Inc., is a Windows® based speech analysis tool, created by Philippe Martin.

accessible also after the transcription activity, since the transcript is linked to the tracks, or it can be taken separately in .rtf format. The software function for transcription also allows to represent overlapping in turn-taking, through the correspondence layer-speaker, where the layer is the digital space containing the transcript.



**Figure 2.0.** Screenshot from WinPitch. The central largest window is the “analysis window”, where the pitch contour (melodic curve – in black), the intensity (volume – in blue) and the waveform (in green) are showed. On the top right of the image there is the video track window. Below the analysis window there are the layers containing the transcript.

The analysis window (see figure 2.0) of the software allows the researcher to select a segment of the acoustic signal, zoom in it to better analyse the pitch contour, the intensity and the waveform. A special function of the software permits to slow down the speed of speech in order to better analyse the acoustic signal. This function resulted of great importance for the purpose of the present study, which considers the aspects of the continuity and variability of the acoustic signal, and the relationship between sound material and form (see below – chapter 3). A text window allows the researcher to read the transcript in .rtf format while the transcription is in progress. Text, sound and image are aligned.

For the transcription of the CorIT, a series of conventions were adopted, on the bases of a literature review of transcription in interpreting studies (cf. Falbo 2005) and the norms illustrated by Blanche-Benveniste and Jeanjean (1987; in Falbo 2005, 2012: 174-176). The result was summarized in the following passage (1987; in Falbo 2012: 176):

The CorIT project will therefore be based on an orthographic – hence more legible – transcription method, with only a few exceptions of phonetic transcripts (e.g. interlanguage transcription, cf. Falbo 2007); punctuation should be relinquished completely, to avoid the risk of considering the transcript as a written text; capital letters should be reserved to proper names exclusively; speakers should be distinguished and identified in the transcript; any doubt expressed by the transcriber should be indicated, so as not to consider an individual decision as a universally acknowledged truth. A reliable software and respect for established conventions should guarantee a satisfactory level of objectivity, or at least a declared and transparent subjectivity (Falbo 2012: 176).

The transcription conventions identified by Falbo are reported in Table 2.1 (below).

At the basis of the transcription conventions identified by Falbo there is the concept of transcription as a ‘paradoxical task’, as it was proposed by Blanche-Benveniste and Jeanjean (1987: 115; in Falbo 2005):

Transcrire de la langue parlée tient un peu du paradoxe: garder dans une représentation écrite certaines caractéristiques de l’“oralité” ; faire le “rendu” de la chose orale tout en restant dans des habitudes de lecture établies depuis longtemps pour la chose écrite ... On va se trouver tiraillé entre deux exigences: la fidélité à la chose parlée et la lisibilité de son rendu par écrit. (Blanche-Benveniste e Jeanjean 1987: 115; in Falbo 2005: 31)

The reason why it is impossible to transcribe orality is mainly due to the fact that speech is not made only of words, but also of melody, emphasis, changes of volume, lengthened syllables, etc. (Falbo 2005: 28). All the aspects of the phonic signal, which at the same time conveys the representative signs of a language, transmits both the expression of the speaker and her/his appeal to the hearer will be dealt with in chapter 3. Notwithstanding the limitations represented by the writing, the transcription conventions identified for the CorIT (see figure 2.1 - above) in part allows the representation of expression in speech, in this case, mainly related to the cognitive activity of simultaneous interpretation. Nonetheless, writing is not the only limitation to transcription, there is also the transcriber’s knowledge of what is being transcribed, her/his

systems of expectations (Blanche-Benveniste and Jeanjean 1987: 102-103; in Falbo 2005: 29-30) and his/her encyclopedic knowledge.

Maiuscola iniziale	Nomi propri, di istituzioni, toponimi ecc.
(.)	Pausa. Se è necessario quantificare la pausa, al posto del punto si inserirà la durata della stessa.
par-	Parola troncata
X	Sillaba incomprensibile
XXX	Parola incomprensibile
(pa)rola	Fonemi non comprensibili in modo chiaro, ma che non inficiano la comprensione dell'intera parola
Parola.parola.parola	Ritmo sincopato
MAIUSCOLO	Schiarimenti di voce, deglutizioni, risate, sospiri, colpi di tosse, applausi, commenti dell'interprete, microfono...
Eh, ah, ehm	Esitazioni vocalizzate
Cifre e sigle	In lettere
a: parola: parola: parola:::	Allungamento (vocalico o consonantico)
/variante 1, variante2/	Incertezza e difficoltà di decodifica
((gesto))	Eventuali elementi prossemici
→parola parola←	Ritmo d'eloquio accelerato
←parola parola→	Ritmo d'eloquio decelerato
Ex: palora (!)	Utile sia per segnalare tutti i casi di <i>metatesi</i> , <i>anticipazione</i> , <i>perseverazione</i> e <i>trasposizione</i> (Magno Caldognetto & Tonelli 1993: 178), sia per fugare dubbi su un possibile errore di battitura.
A: dgjioegj <dghdjk> B: <jkgkfg>	Sovrapposizione di due o più oratori.
Nome proprio in ortografia corretta !nome proprio effettivamente pronunciato!	Nel caso di nomi propri pronunciati in modo scorretto o con accento straniero, verrà indicata l'ortografia standard e tra punti esclamativi l'effettiva pronuncia.

**Table 2.1.** Transcription conventions for the CorIT project (Falbo).

All these factors influence perception during the execution of a transcription, and an aspect of their cognitive impact is synthesized by the following statement by Blanche-Benveniste and Jeanjean (1987: 102; in Falbo

2005: 29): “l’oreille est un traître; on écoute ce qu’on s’attend à écouter”. Thus, transcription is subjective and it is only a trace of the result of a communication act (cf. Falbo 2005: 27, 31). With reference to the transcription of the CorIT, it was observed that the use of WinPitch function to slow down the speed of speech during the transcription by students led the transcribers to omit some passages; after a revision by the tutor, students admitted to have listened and comprehended those passages; however, they “forgot” to transcribe them (Falbo 2005: 34). According to the examples reported by the tutor, the segments unconsciously omitted from transcription were mainly related to: i) speaker’s reactions to the situation of spoken interaction (e.g. comments about the turn-taking; comments on what had been just said; other types of deictic information); ii) interpreter’s expression of her/his problems with discourse planning (hesitations); iii) complex words or proper names in the source language (ibidem):

- des tours de parole: “grazie una domanda”, “sono d’accordo”;
  - des segments plus ou moins longs: “ecco ne abbiamo uno qui”; “no no non c’è bisogno di fare polemiche”;
  - des tâtonnements de l’interprète (les parties qui n’apparaissent pas sont en majuscule): “un UN indebitamento”; “ogni A- OGNI giorno”; “colpevole CHI chi è stato”; “a ventidue ANNI ventiquattro anni”; “eh eh EH L- cosa accade”; “il DAL- il dollaro”; “e quindi non è COLPA ancora una volta il colpevole”; “questa sera con UN CON degli aspetti”.
- Nous avons pu également repérer des parties qui avaient été notées comme incompréhensibles :
- “in piccionaia”; “New Hampshire”.

The listening of the acoustic signal of interpreted texts at a low speed rate did not allow the transcribers to solve problems of perception of syllables that could be interpreted as either false starts (i.e. truncated words) or vocalized hesitations. For these problems of annotation, Falbo (ibidem: 35-36) proposed the adoption of the “multi-transcription”, as suggested by Blanche-Benveniste e Jeanjean (1987: 143; in Falbo 2005: 35). Here are some example of multi-transcription (the options are indicated within two forward slashes characters) reported by the author (Falbo 2005: 35-36):

- contro /la, l’a-/ /la, l’a-/ l’avvicinamento”;
- “che è /al, all-/ all’interno”;
- “/del, dell-/ dell’umanità”;
- “/nel, nell-/ nell’aria”;
- “gli iracheni /ah, ha, ha-/ sembrano hanno chiesto”; la présence d’un sujet pluriel renforce à notre avis l’interprétation “ha-”; cette hypothèse est soutenue



par la présence de “sembrano” et le changement successif de planification opéré par “hanno”; mais le choix le plus difficile est entre “ah” hésitation et “ha” verbe avoir;

- “aggiunta di una scadenza a questa risoluzione e:/eh non vorrei suggerirle”;
- “si /eh, e-:/ ebbe”; un [e] avec allongement vocalique se confond parfaitement avec une hésitation.

## **2.2 The thematic corpus: OReNESIT (Obama vs. Romney English español italiano) – the 2012 US Presidential Debate**

A thematic corpus for the present research study was built up, in order to observe interpretations produced in real situation, i.e. authentic material.

### **2.2.1 Digital archive**

Firstly, an archive of digital video recordings of simultaneous interpretations from English into Italian (EN-IT) and from English into Spanish (EN-ES) of 2012 United States Presidential Debates was created. Then interpreted texts were transcribed. Videos were converted into the formats suitable to be transcribed with WinPitch. Videos of interpretations from English into Italian were recorded from Italian state-owned television channel *Rainews24* and private TV channel *SkyTG24*, using a DVD recorder connected to a satellite television receiver; due to some technical problems, only in two out of six Italian television broadcastings<sup>st</sup>(1<sup>st</sup> and 3<sup>rd</sup> *Rainews24* debates) it was possible to effectively record the full version of the debates (90 min). Interpretations from English into Spanish were downloaded from the website of Spanish state-owned television channel *RTVE-Canal24horas*, and the websites of US private television channels *Univisión*, *Telemundo* and *CNN en Español*. The thematic corpus was named *OReNESIT* (Obama vs Romney English español italiano), and can be divided into three subcorpora: (1) original texts (OT), (2) *OReNIT* (Obama vs Romney English italiano) and (3) *OReNES* (Obama vs Romney English español). The archive is made of 17 videos for a total 1,147 minutes, i.e. 19.4 hours (cf. Table 2.2).

### **2.2.2 Transcription of the corpus OReNESIT**

The transcription of the OReNESIT corpus, which is a sub-corpus of the CorIT sub-corpus of presidential debates, was carried out by the researcher in collaboration with Babetto (2013), Penso (2014) and Tomasetig (2014): each of them analysed the 22% of the corpus, for a total of 66% of the corpus, the remaining 33% was transcribed by the researcher. Babetto (2013), Penso (2014) and Tomasetig (2014) also analysed different aspects of simultaneous

interpretations (see below – section 2.2.4) within their respective MA theses, which were co-supervised by the researcher. The transcription of ORe nesit is based on the transcription conventions established for the CorIT (see above – section 2.1.1).

CorIT – ORe nesit digital audio-visual archive (Television Interpreting Corpus - Obama vs Romney english español italiano) 2012 United States Presidential Debates							
subcorpus	Original	ORenit		ORenes			
language/language pair	EN	EN-IT		EN-ES			
broadcaster	<i>nytimes.com</i>	<i>RAI</i>	<i>Sky Italia</i>	<i>RTVE</i>	<i>Univisión</i>	<i>Telemundo</i>	<i>CNN</i>
Broadcaster's state	USA	Italia	Italia	España	USA	USA	USA
Broadcaster's status	private	state-owned	private	state-owned	private	private	private
TV channel	-	<i>Rainews 24</i>	<i>SkyTG24</i>	<i>Canal24horas</i>	<i>Univisión</i>	<i>Telemundo</i>	<i>Cnn en Español</i>
1 <sup>st</sup> debate	90 min	90 min	77 min	90 min	90 min	-	20 min
2 <sup>nd</sup> debate	90 min	74 min	70 min	90 min	90 min	9 min	8 min
3 <sup>rd</sup> debate	90 min	90 min	37 min	90 min	90 min	90 min	-

**Table 2.2.** ORe nesit audiovisual recordings.

The 33% of the corpus (the researcher's share) also includes the transcription of audible breaths ("RESPIRO"). This decision was taken with the aim of analyzing the distribution of audible breaths in simultaneous interpreting speech to check the relationship between such distribution and quality of speech. In fact, Goldman-Eisler (1968) studied the function of breathing in speech production, and modern methods of teaching of voice and speech control (e.g. Coblenzer and Muhar 1999/2004; Veneziano 2000) consider breathing fundamental. On the basis of the transcription executed, the researcher observed that in poor SI speech quality audible breaths occurred more frequently, nearby silent pauses, hesitations and syllable lengthenings; while in good quality SI speech, breaths were either not audible or not perceivable. However, this hypothesis should be tested through a dedicated corpus-based analysis which was not carried out within the present study.

With particular reference to the present study, the transcription was carried out as *praxis*, i.e. natural practice of the theoretical paradigm (see below – chapter 3) for the analysis of simultaneous interpreting speech described in chapter 3.

All the transcribers of ORe nesit, as other transcribers of the CorIT, faced the problems of perception and annotation described by Falbo (2005) and reported above (section 2.1.1). In addition, the transcribers could

experience the difficulty, sometimes the impossibility, of recognizing the phonemes as basic units of the phonic signal; in fact, despite the (sometimes reiterated) use of the WinPitch device to reduce the speed of speech, it was impossible to define the phonemes, especially in cases of distorted or weak signal, mainly in correspondence of hesitations, pronunciation of difficult words and unknown proper names. Consequently, the transcription activity has allowed the transcribers to experience that the syllable is the natural basic unit of the acoustic signal (cf. Bühler 1934/1983: 314-328; Albano Leoni 2006: 179-182; see below – section 3.4). As a consequence of a researcher's mistake in distributing the archive material for transcription (all transcribers worked on all the four versions of the three 2012 presidential debates), a section (clip 9: five min) of the first presidential debate was transcribed twice by two subjects. The two versions reported below (table 2.3) offer the opportunity to observe the limitations of the orthographic transcription and all the aspects discussed so far (section 2.1.1 and this section); they allow the reader to compare and contrast the subjectivity (and the objectivity) of transcription.

<b>CorIT – Orenesit &gt; Orenit_1_SkyTG24_(1) Pag. 8/24</b>	
De Gregoris:46-77min/clips [6, 9-15]	Penso (2014):31-45min/clips [7-9]
Clip 9 – transcript – WinPitch	Clip 9 – transcript – WinPitch
<p>*IN1: Centoventidue piccole banche che invece sono- hanno dovuto eh chiudere (in)vece (.) vuole eh (.) quindi lei parla di ridurre l- eh di ab- di abolire questo tipo- que- questa normativa sul settore eh: eh bancario ma ci sono alcuni aspetti di questo d- normativa che eh portano a una maggiore trasparenza (.) al leverage ma (.) parliamo di quelle più gr(an)di (.) RESPIRO no no facciamo rispondere al presidente prima (.) su Dodd-Frank e quello che lei ha appena detto</p> <p>*IN2: È un grande esempio sì eh RESPIRO (.) i- (.) il motivo per cui ci troviamo in una crisi economica davvero così sorprendente RESPIRO è rif- è: ascrivibile a: eh a quello che è stato fatto non soltanto a Wall Street eh (.) c'erano XXmente RESPIRO prestiti concessi a dismisura (.) eh anche se le persone non si qualificavano per poter avere questi prestiti eh agenzie di credito ehm che parlava di grandi investimenti dando il grado A ma non erano grandi</p>	<p>*IN1: centoventidue piccole banche che invece sono hanno dovuto eh chiudere invece (.) vuole (.) quindi lei parla di ridurre di: ab- di abolire questo tipo que- questa normativa sul settore e:h b- bancario (.) ma ci sono alcune aspetti di questo (del)la normativa che (.) hm portano a una maggiore trasparenza (.) al leverage ma (.) parliamo di quelle più grande (.) no.no facciamo rispondere al presidente prima (.) su Dodd-Frank e quello che lei ha appena detto</p> <p>*IN2: è un grande esempio sì e:h (.) i (.) il motivo per cui ci troviamo in una crisi economica davvero così sorprendente (.) è ref- /e:h, è:/ ascrivibile a e:h a quello che è stato fatto non soltanto a Wall Street (.) eh (.) c'erano veramente (.) prestiti concessi a dismisura (.) anche se le persone non si qualificavano per poter avere questi prestiti (.) e:h a:agenzie di credito e:hm (.) che parlava di grandi investimenti dando il grado a ma non erano grandi investimenti e</p>

<p>investimenti e anche eh eh: (.) prodotti bancari che le stesse banche non capivano RESPIRO per poter realizzare profitti enormi nella consapevolezza che il sistema era vulnerabile che cosa abbiamo fatto siamo intervenuti RESPIRO (.) con la maggiore riforma di Wall Street dagli anni trenta RESPIRO eh abbiamo ehm detto eh appunto di eh:m min- eliminare questi eh comportamenti di rischio ho (.) in manie(r)a tale eh RESPIRO da evitare gli- i debiti sbagliati nel frattempo abbiamo assicurato che RESPIRO eh tutto l'aiuto fornito alla: hm alle banche era ripagato con interessi Romney ha detto che vuole RESPIRO eh revocare Dodd-Frank !Dotfrank! eh (.) eh (.) son d'accordo che eh eh- per eh eh funzionare il mercato deve avere delle regole RESPIRO m-hm lui vuole eh revocare questo questo progetto RESPIRO ehm (.) ci chiediamo se gli amerihani pensano che il problema che gra- (a)bbiamo avuto e che (.) c'è stata una enorme regolamentazione di Wall Street beh se questo è il pensiero (.) Romney RESPIRO è il candidato di questi elettori (.) ma non è quello che io credo</p> <p>*IN1: no non è giusto non è giusto (.) dobbiamo avere una: una: regolamentazione di Wall Street del mercato azionario RESPIRO RESPIRO ma io non direi che cinque banche sono: troppo: ho: grandi per essere messe in liquidazione e dargli carta bianca RESPIRO la normativa sulla protezione dei consumatori la Dodd-Frank RESPIRO distrugge le piccole banche un'altro aspetto di Dodd-Frank della normativa dice RESPIRO hm dare i mutui a que(II)i che non aveva diritto esatto ecco perché noi ci trovi(am)o n(ell)a situazione attuale la- la normativa Dodd-Frank è giusto cioè bisogna dare i mutui RESPIRO che (.) d- sono devono essere coperti (.) però non è mai stato definito che cos'è un buon mutuo e un cattivo mutuo ma sono due anni che non sappiamo che cos'è un- un- un mutuo accettabile (.) le banche non prese- prestano soldi (.) perché questo ha- ha- hm: ha inciso sul mercato edilizio perché la gente non ha i-ri- i soldi i mutui per poter cre- RESPIRO costruire una casa</p>	<p>anche eh e::h (.) prodotti bancari che le stesse banche non capivano (.) per poter realizzare profitti enormi nella consapevolezza che il sistema era vulnerabile (.) che cosa abbiamo fatto siamo intervenuti (.) con la maggiore riforma di Wall Street dagli anni trenta (.) eh (a)bbiamo e:h detto eh appunto di ehm: hm elimina:re questi eh comportamenti di rischio oh (.) (i)n maniera tale eh (.) (GUTTURALE) da evitare gli i debiti sbagliati nel frattempo abbiamo assicurato che (.) e:h tutto l'aiuto fornito alla e:h alle banche era ripagato con interessi (.) Romney ha detto che vuole (.) eh revocare Dodd-Frank (.) eh (.) (GUTTURALE) sono d'accordo che eh (.) per: e:h funzionare il mercato deve avere delle regole (.) hm: (.) lui vuole eh revocare questo questo progetto (.) eh ci chiediamo se gli americani pensano che il problema che gra- abbiamo avuto è che (.) c'è stata una enorme regolamentazione di Wall Street beh se questo è il pensiero (.) Romney (.) è il candidato di questi elettori (.) ma non è quello che io credo</p> <p>*IN1: no non è giusto non è giusto (.) dobbiamo avere una: una: regolamentazione di Wall Street del mercato azionario (.) ma io non direi che cinque banche sono: troppo e:h grandi per essere messe in liquidazione e dargli carta bianca (.) la normativa sulla protezione dei consumatori la Dodd-Frank (.) XXX distruggere piccole banche (.) un /altr-, altr'/ aspetto di Dodd-Frank della normativa dice (.) eh dare dei mutui a quei che non aveva(no) diritto esatto (.) ecco perché noi ci trova- nella situazione attuale (.) la hm: la normativa Dodd-Frank è giusto cioè bisogna dare i mutui (.) che (.) d- sono devono essere coperti (.) però non è mai stato definito che cos'è un buon mutuo e un cattivo mutuo ma sono due anni che non sappiamo che cos'è una un: un mutuo accettabile (.) le banche non prese- (.) prestano soldi (.) perché questo ha ha hm: ha inciso sul mercato edilizio perché la gente non ha i ri- soldi i mutui per poter cre- (.) costruire una casa (.) non è la</p>
---	---

<p>non è la normativa Dodd-Frank anzi (.) RESPIRO a volte però non voi non avete avuto una normativ- un appioccio (!) un approccio chiaro RESPIRO non voglio che la normativa (.) incida in modo negativo sull'industria edilizia (.)</p> <p>*IN2: C'è un'altra:: chiara distinzione tra voi due signori ad- hm (X X / XXX) adesso parliamo sempre l'assistenza sanitaria dove so benissimo che c'è una notevole differenze RESPIRO ci(o)è tra il programma suo presidente Obamacare (.) avete due minuti (.) lei eh: governatore Romney parlerà per prima lei vuole abrogare (.) il: il programma perché (.) sempre in base alla mia esperienza eh:: nel New Hampshire una donna mi è: venuta vicino ed ha detto che una don- io non po- una donna mi ha detto che io mi posso permettere l'assicurazione per me e per mio figlio RESPIRO eh un'altra coppia nel Wisconsin mi ha detto lo stesso vogliamo voglio ridurre RESPIRO anche altra gente: piccole: imprese dicono (ch)e il costo dell'assistenza sanitaria è proibitivo (.) ma ma (.) sfortunatamente (.) con il programma del presidente l'Obamacare il cosi- famigerato Obamacare (.) ci costerà (.) i- a livello individuale dumilaecinqucento dollari a testa in più cioè quando il presidente (.) fu eletto quattr anni fa lui ha detto che lui avrebbe ridotto il costo dell'assicurazione sanitaria eh di duemilaecinqucento mentre invece è stata ha XXX il contr- è aumentata</p>	<p>normativa Dodd-Frank anzi (.) ma a volte però non voi non avete avuto una normativa un appioccio (!) un approccio chiaro (.) non voglio che la normativa (.) incida in modo negativo /sull', su l'/ industria edilizia (.) [...] [Interpreter 2 takes over – omitted transcription]</p> <p>c'è un'altra:: chiara distinzione tra voi due signori a: XXX adesso parliamo sempre l'assistenza sanitaria dove so benissimo che c'è una notevole differenze (.) t- tra il: (.) programma suo presidente Obamacare (.) avete due minuti (.) lei (.) e:h governatore Romney parlerà per /prima, primo/ lei vuole abrogare (.) il: il programma perché (.) (.) sempre in base alla mia esperienza (.) e::h nel New Hampshire una donna mi è: venuta vicino e ha detto che una don- io non po- una donna mi ha detto che (.) io non mi posso permettere l'assicurazione per me e per mio figlio (.) e:h un'altra coppia del Wisconsin mi ha detto lo stesso vogliamo voglio ridurre (.) anche altra gente (.) piccole: imprese che dicono (.) che il costo dell'assistenza sanitaria è proibitivo (.) ma (.)sfortunatamente (.) con il programma del presidente (.) l'Obama.care il cosi- il famigerato Obamacare !Obamaché! (.) ci costerà (.) i- a livello individuale (.) du(e)mila e cinquecento dollari a testa in più (.) cioè quando il presidente (.) fu eletto quattro anni fa lui ha detto che lui avrebbe ridotto (.) il costo dell'assicurazione sanitaria (.) di duemilacinquecento mentre invece ha (e)sattamente il contrario ha aumentato</p>
---	--

**Table 2.3.** Excerpt of transcript from the thematic corpus ORenesit – first 2012 presidential debate (English-Italian from SkyTG24).

As it is evident from this excerpt of the transcript, the convention ‘accelerated speech rhythm (€ parola parola ≈)’ and ‘decelerated speech rhythm (≈ parola parola €)’ at the end was not adopted for the CorIT.

### 2.2.3 The corpus Orenesit

Not all the recordings of the archive (see Table 2.2) were considered in the corpus building; in fact, the English-Spanish interpretations from Telemundo and Cnn en español were not included in the corpus, as can be seen in Table 2.4.

2012 US Presidential Debate	Broadcaster – Channel	Number of words
1	RAI – Rainews24	14,473
	Sky Italia– SkyTG24	12,789
	RTVE – Canal24horas	15,337
	Univisión	14,125
2	RAI – Rainews24	13,447
	Sky Italia – SkyTG24	11,788
	RTVE – Canal24horas	15,775
	Univisión	14,960
3	RAI – Rainews24	16,478
	Sky Italia – SkyTG24	5,767
	RTVE – Canal24horas	15,946
	Univisión	16,785
<b>Total words</b>		<b>167,670</b>
<p><b>Note:</b> the words counted (through Word application) include disfluencies, transcription of noises, guttural stop, audible breaths (in the 33%), etc.</p>		

Table 2.4. Orenesit corpus.

### 2.2.4 Corpus-based analyses

Babetto (2013), Penso (2014) and Tomasetig (2014), within their respective MA theses, co-supervised by the researcher, carried out three analysis on the 77% of the Orenesit corpus, since the transcription had not been completed yet. With reference to the present study, the most relevant results from each of them are reported in the following sections.

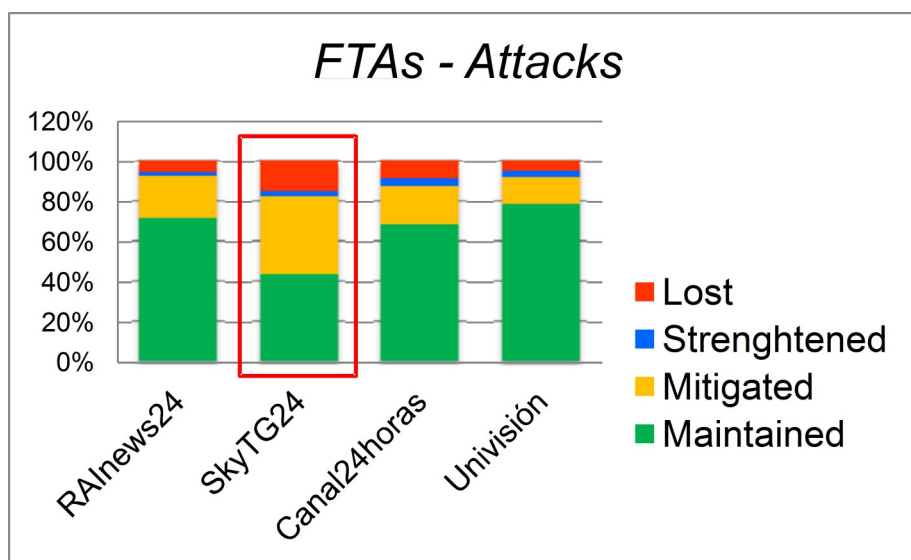
#### 2.2.4.1 Analysis of rendition of face-threatening speech acts (Babetto 2013)

Babetto (2013) analysed the renditions of face-threatening speech acts, identified on the bases of the line of research of the face-work, initiated by

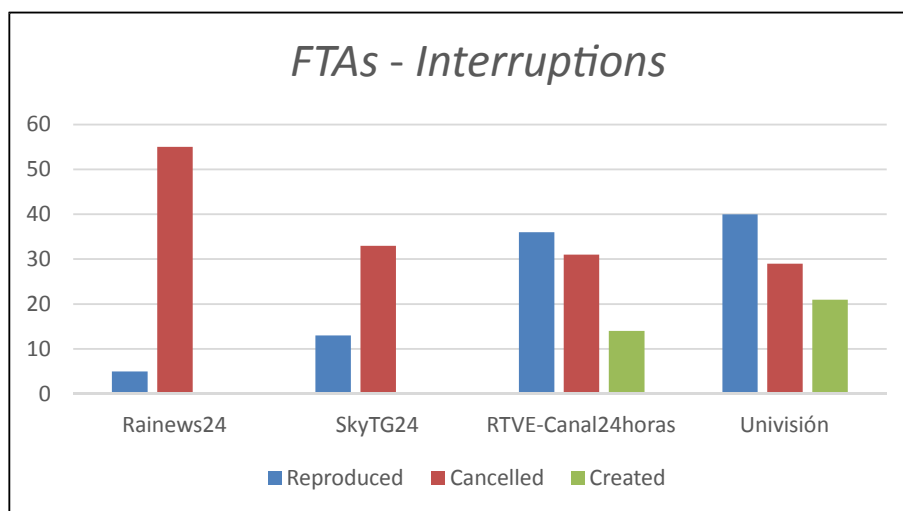
Goffman (1967), and successively developed by Brown and Levinson (1987) and Kerbrat-Orecchioni (1992; 2011) (in Babetto 2013). After an adaptation of the categories defined by Brown and Levinson and Kerbrat-Orecchioni, the author defined the following categories for the corpus-based analysis (Babetto 2013: 51):

- 1) Atti linguistici che minacciano la *face* positiva dell'interlocutore, o **attacchi**: critica, canzonatura, derisione, disaccordo (obiezione, contestazione, confutazione), biasimo e rimprovero, provocazione e sfida; rientra in questa categoria anche la *polirudesse* di Kerbrat-Orecchioni;
- 2) Atti linguistici che minacciano la *face* negativa dell'interlocutore: suggerimenti, consigli, **interruzioni**;
- 3) Atti linguistici che possono danneggiare la propria *face* negativa: promesse;
- 4) Atti linguistici che possono danneggiare la propria *face* positiva: ringraziamenti, scuse, confessioni, **espressioni di accordo**.

From the summarizing tables reporting the number of face-threatening “attacks” and “interruptions” (cf. Babetto 2013: 96-97), two histograms to show and compare the rendition of “attacks” and “interruptions were plotted:



**Figure 2.1.** Rendition of face-threatening attacks in the four interpretations of the three 2012 US presidential debates (OReunit – 60%) – based on Babetto (2013: 96-97).



**Figure 2.2.** Rendition of face-threatening interruptions in ORe nesit (60%) – based on Babetto (2013: 96-97).

Results show that the Italian interpretation broadcast by SkyTG24 presents the lowest number of the attacks maintained by the interpreters, and the highest number of both mitigated and lost attacks. As far as interruptions are concerned, results show that Rainews24 broadcast the highest number of cancellations and the lowest number of reproductions. These results do not depend only on the interpreters' skillfulness, but mainly on the broadcasters' choices. For example, Rainews24 and SkyTG24 had only two interpreters in the booth taking over, while Univisión and Canal24horas had three interpreters, one for each actor of the debate (Obama, Romney and the moderator or the selected citizens, according to the format of the second debate). In addition, in the booth of Rainews24 there was only one microphone (see below – section 2.2.5), which represented an obstacle for the reproduction of interruptions. When interpreting these graphs, as well as the following ones (next figures), the reader has to confront the results with the data related to the total time of recordings which are not homogeneous (see above – Figure 2.2).

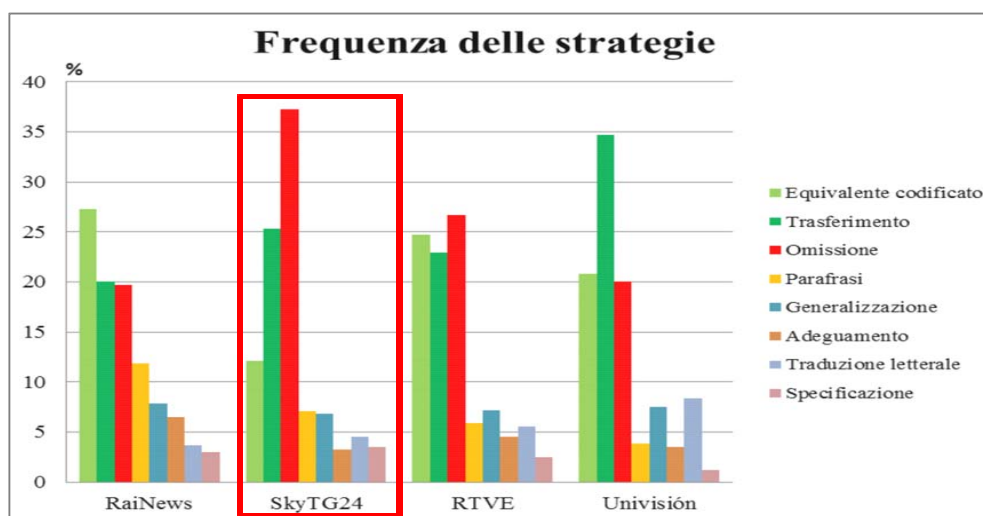
#### **2.2.4.2 Analysis of strategies for the translation of culture-bound elements (Tomasetig 2014)**

Tomasetig (2014), mainly referring to the translation studies on culture-bound elements (CBE) by Florin (1993) and Viezzi (2004), defined a specific series of CBE, dividing them into: international references; national (USA)



references (institutions, proper names, toponyms, non-institutional associations and companies); and allocutive forms (Tomasetig 2014: 31-32). On the bases of Pedersen (2007), the author classified the renditions of CBE in the following translation strategies: cultural adaptations, literal translation, official equivalent, transfer, generalization, specification, paraphrase and omission (Tomasetig 2014: 79-83).

The frequency of the above mentioned strategies adopted in the four versions of the televised interpretations of 2012 US presidential debates are reported in the following figure (2.3; in Tomasetig 2014: 79):

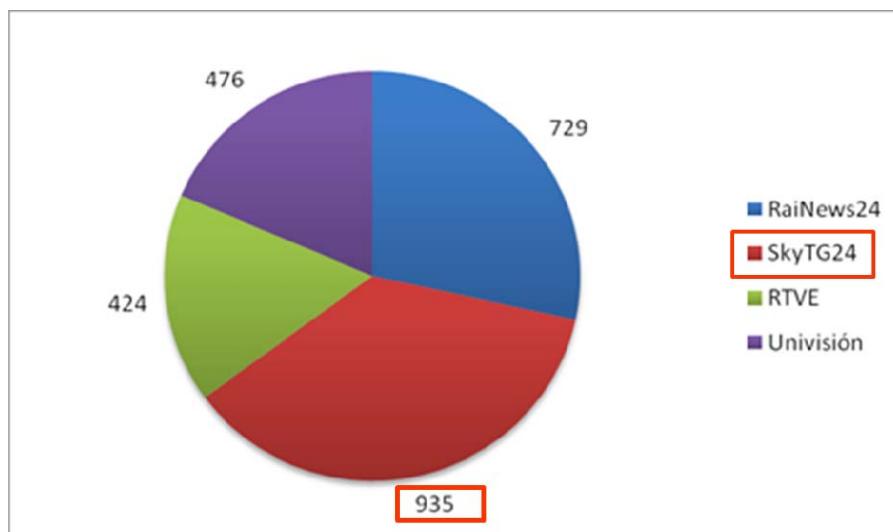


**Figure 2.3.** Frequencies of rendition strategies for the translation of culture-bound elements in ORenesit (77%) (Tomasetig 2014: 79).

Results show that the interpretation broadcast by SkyTG24 presents the highest number of omissions and the lowest number of official equivalents; it presents an inhomogeneous use of the rendition strategies of CBE.

#### 2.2.4.3 Analysis of self-repairs (Penso 2014)

Penso (2014) based her analysis of self-repairs (substitution, recycling, substitution + recycling, other repairs) mainly on the MA thesis by Papa (2011), the results of which were in part confirmed. The categorization of self-repairs was mainly based on Fox et al. (2009).



**Figure 2.4.** Occurrences of self-repairs in Orenesit (77%) (Penso 2014: 92).

From the analysis of self-repairs, the highest number of occurrences was found in the interpretation by SkyTG24 (figure 2.4).

### **2.2.5 TV interpreters, situation and working conditions (RTVE and RAInews24)**

When the Orenesit corpus was being built, the search for digital recordings on the web led to discover that one of the interpreters of the 2012 presidential debates who worked for the RTVE, Daniel Sánchez Reinaldo had posted on his personal web site (<sanchezreinaldo.com>) information on the interpretations of the 2012 presidential debates, as well as links to the videos of the debates. He also mentioned others interpreters who worked with him in the interpretation sessions. From an interview published on the website “manzanares DIGITAL” (20 October 2012) it was possible to know that he has been working as an interpreter for RTVE (in radio and television programs) since 2007; thus, he also interpreted the 2008 presidential debates. He is a freelancer interpreter also working for private radio stations (e.g. Cadena Ser). We report two excerpts from the interview where he gave some information about his job:

**¿Cómo han sido estos dos debates entre Obama y el candidato Romney desde el punto de vista de la interpretación?**

Han sido difíciles porque en un debate nunca sabes lo que te puede salir. Te puedes informar antes, leer, documentarte, pero llegada la hora estás cara a cara con la persona a la que interpretas, sin más armas a tu alcance. Además, aunque

conozcas bastante bien la política estadounidense, nunca lo puedes saber todo. Por ejemplo, preparando el debate, vi que un punto de diferencia entre Obama y Romney son los "doughnut holes". Literalmente significa "agujero de donuts". Ya investigando vi que así es cómo se llama al periodo sin cobertura sanitaria que pueden tener los jubilados en EE.UU. Si te sale eso en el debate sin habértelo mirado te quedas pensando en una décima de segundo: "he oído agujero de donuts, ¡pero cómo voy a decir eso!". Lo mismo ocurrió con la Gallina Caponata. Lo oyes claramente, pero piensas: "¿cómo se van a poner a hablar de Caponata en un debate? No puede ser, pero lo he oído...voy a decirlo". Y lo traduces pensando que al día siguiente vas a salir en todos los zappings.

Aparte de la terminología, que siempre es difícil, estos dos primeros debates han tenido otro aspecto complicado: ambos candidatos hablaban a gran velocidad, dando mucha información en cada frase, dando muchas cifras (de hecho corre el rumor de que Romney en el primer debate metió en su atril una hoja con cifras cuando no estaba permitido). Ahí intentas pegarte lo máximo posible a tu orador e ir rápido, pero a veces tienes que hacer un resumen porque es imposible ir a la misma velocidad que él, escuchando lo que dice, pensando e interpretando.

[...]

### **He leído que sueles documentarte previamente. ¿Cómo se prepara un debate electoral como el de Francia o ahora los de Estados Unidos?**

El trabajo previo es la única ayuda que tenemos para preparar una interpretación, porque nunca tenemos delante de nosotros el discurso que se va a leer. Por eso, una buena preparación es indispensable. Para los debates me he tenido que aprender el programa electoral de Obama, las diferencias con el de Romney (igual que para el debate de Francia), leer artículos de opinión, las últimas noticias, estar al tanto de las declaraciones de los dos candidatos, de las polémicas (por ejemplo, cuando Romney dijo a puerta cerrada que el 47% de los estadounidenses no le votarían porque son unos "mantenidos por el gobierno").

Todo eso te ayuda a que se te encienda la bombilla cuando lo oyes en el debate. Es como un hilo del que tirar, ya sabes de qué va. Por ejemplo, Obama piensa contratar 100.000 profesores de matemáticas y ciencias en 4 años para mejorar la calidad educativa. Al interpretar, en cuanto oigo "100.000 profesores" ya sé de qué va el tema. Sé que son profesores de matemáticas, de ciencias, sé en cuánto tiempo los quiere contratar...y eso me permite tener un momento de relajación mientras sigo hablando.

(<<http://manzanaresdigital.blogspot.it/2012/10/obama-hace-unos-discursos-muy-sensatos.html>> last access 02.2016)

From the same web page, among other photos, there was that of the RTVE interpreting boot equipment (see Figure 2.5)



**Figure 2.5.** RTVE interpreting booth equipment.  
(<http://manzanaresdigital.blogspot.it/2012/10/obama-hace-unos-discursos-muy-sensatos.html>).

Thanks to an interview given to the researcher by Letizia Tesorini (2013: personal communication), one of the interpreters of 2012 presidential debates who worked for Rainews24, it was possible to discover that the RTVE booth equipment is similar to that at Rainews24. Here, the interpreters work with two monitors: the one on the left (teleport screen) transmits the images taken from the international circuit (transmitted by Reuters or Eurovision), while the one on the right transmits the image of the Italian broadcaster (with its peculiar superimposed graphics). Sometimes the signal from the international circuit is not clear, due to reception problems. Other problems are related to the difference in volume between the audio signal of the international circuit and the headphones volume; to adjust this volume, the interpreters have to phone the transmission control room, which is far from the booth; this is good for acoustic isolation, but not for communication between the two facilities. Other problems are related to the slight recorded broadcast of the Italian channel with respect to the transmission by the international circuit; thus, the images on the two monitors are not synchronized; consequently, the interpreters have to handle the lip sync lag. Moreover, the superimposed graphics of the Italian channel also may create problems to the interpreters. Sometimes, from the transmission control room the Italian newscaster's

microphone is not switched off; therefore, the interpreters are disturbed by the noises coming from the studio and the newscaster's voice (which is not broadcasted, because only the interpretation is); among the noises there is the interpreter's echo, because also the newscasters listen to the interpretation.

In the Rainews24 booth there is only one microphone; this prevents the interpreters from reproducing the speakers' overlaps in turn-taking or interruptions, which in the case of presidential debates can be moments of confrontation (see above – section 2.3.3.1).

Letizia Tesorini said that fidelity to the original text, accuracy and voice control are the main aspects of interpretation according to which she is judged; she said she is appreciated mainly for her voice.

When asked how she handled tension during the first years of work as a television interpreter, she answered “just working” (*ibidem*), then admitted that the stress induced by television interpreting and voice control caused her physical problems.

Answering to the question on the preparation for the presidential debates, the interpreter said that she studies the television format of debates (see below – section 4.5.2.1) and the political programs of the candidates; get informed about recent events, especially those related to foreign politics; studies the national policy, which is considered the most difficult and complex topic, presenting many unknown elements; and, finally, she studies the character, her/his personality and elocution.

The interpreter has been working for Rainews24 for 14 years; her two colleagues also have been working for many years. All of them are contractual project-based workers and turn over to guarantee a continuous service in case of breaking news or important events.

### **2.3 Conclusion**

The interpreters' words reported above (section 2.2.5) may help to interpret the results of the analyses based on Orenesit, some of them are reported above (section 2.2.4). The testimony by the Italian interpreter about the importance of voice in TV interpreting (section 2.2.5) seems to confirm the rationale of the present study.

As far as the transcription is concerned, we can agree with Meyer's statement (1998) according to which transcription allows us to see what the ear cannot understand (in Falbo 2005: 36, 26). Indeed, the orthographic transcription adopted for the CorIT permits to see the damaged acoustic

signal that may create problems to the hearer when detecting languages signs; nevertheless, part of the rhythm cannot be detected through transcription. For example, when the duration of silent pauses is not indicated, a string with a high distribution of silent pauses and syllable lengthenings may appear as a poor quality simultaneous interpreting speech segment; however, when listened to, this judgment may be overturned, because the pauses are not long, the syllable lengthenings are not due to hesitations but to a good control of intonation. Conversely, it can happen that, by listening to the interpretation, the many hesitations (false starts, repetitions, corrections, filled pauses, long silent pauses, etc.) impede the comprehension, and the listener is subject to a high cognitive effort; then, after days or months, by reading the transcript of a 5-minute excerpt, he can appreciate a good terminology and good sentence constructions if hesitations are excluded. The perception of reading is different from that of listening; therefore, transcript can be useful especially when the reader is the transcriber, who reads the transcript with the memory of the listening, or if the reader has listened attentively the interpreter.

Results from the three analyses showed that the same Italian interpretation of the three 2012 Presidential Debates, namely the one broadcast by the private channel SkyTG24, presented: i) the highest number of lost or mitigated renditions of face-threatening attacks; ii) the highest number of omitted translations of culture-bound elements (together with the lowest number of codified equivalents and, in general, a highly dishomogeneous use of the strategies analysed); and iii) the highest number of self-repairs. If one considers Bühler's proposal of instrumental model of language (1931: 76-85; see below – chapter 3), according to which the concrete acoustic phenomenon is made of three moments, each one can convey three different signs: i) symptom (index), referred to deictic function and based on the sender; ii) signal, referred to the appeal function and based on the receiver; and iii) symbol, referred to the representational function and based on objects or states of affairs; then in self-repairs the expressive function prevails, in face-threatening acts it is the appeal function that prevails, while in culture-bound elements prevails the representational function. The term “prevails” was used to mean that the three kind of signs corresponding to the three linguistic functions coexist at the same time, referred to the same acoustic phenomenon; however, given the nature of the three objects of analysis, the function (type of sign) indicated may be considered predominant. The fact that the main deficiencies in self-repairs,

face-threatening acts and culture-bound elements converged to (were found in) the same speech (interpretation) could represent a support to Bühler's proposal (see below – chapter 3).





## Chapter 3

# THEORETICAL FRAMEWORK FOR A NEW QUESTIONNAIRE FOR A GESTALTIC ASSESSMENT OF TV BROADCAST SIMULTANEOUS INTERPRETATION

### 3.0. Introduction: prosody, voice and language

#### 3.0.1 Prosody

The word prosody (gr. *prosōidia* «accent, modulation of voice», compound of *prós* «next to» and *ōidē* «singing») is a term of classic metrics, where it designs the study of the verse. Initially it indicated the choice of words in the song; later the term was used by Latins to refer to aspects of word accentuation and syllable quantity. Currently, it is used for the rules of versification concerning phonetic aspects, as accent and rhyme. In linguistics, prosody covers more than a single phoneme, since it mainly concern speech and refers to the modulation of some parameters that were significant also in the traditional use of the term. Among these parameters, the most important ones, from the acoustic point of view, are the fundamental frequency of voice, duration and intensity. Their modulation permits the execution of the *prosodic traits* as accent, tone, juncture, intonation and rhythm. These traits are also called *suprasegmental* ones, because they concern more than one phonetic segment, exercising their influence on the prosodic fields, i.e. units superior to the phoneme representing a structure of their own. (Gili Fivela, 2011, my translation).

Both the results of above mentioned Interpreting Studies focused on prosody (see above 1.4.2) and the study by Goldman-Eisler (1968 – see above 1.4.1 and 1.4.3) show the limits of the phonologization of prosody, starting

from the physic variables that define it: tempo and intensity for rhythm, frequency for intonation. As a matter of fact, all these variables are displayed along a continuous flow, for which there is not any given discrete representation as writing could be for phonology, morphology, etc (Albano Leoni 2009:41). Prosody cannot be reduced to the model of double articulation, because its units can only be defined on the bases of the relationships among its components (ibidem: 41-42). These components are extremely variable, and this variability is always expression of the speaker's communicative intention, attitudes and moods, with effects on the semantics and the pragmatics of utterance (ibidem: 42-43). Prosodic and attitudinal variations constitute of two continuous variables, which hardly can be discretized and correlated (ibidem: 44). It is also hard to find a sense in the covariations, the same happens for voice (ibidem). Nevertheless, speakers know very well the mechanism of prosody and grasp the senses it can generate, intertwining with syntax and semantics (ibidem: 44-45).

Prosody is realized by a complex but 'extremely economic' and 'natural' mechanism (Albano Leoni 2009: 45-46):

Questa componente di straordinaria importanza nella comunicazione orale si realizza attraverso un meccanismo articolatorio estremamente economico, basato su: a) una variazione di poche decine di Hz rispetto alla frequenza fondamentale propria di ciascun parlante, b) una variazione che può anche essere di pochi millisecondi nella durata delle sillabe, e c) piccole variazioni nella pressione dell'aria espiratoria a ridosso dei diversi ostacoli glottidali e supraglottidali.

The 'basic principle' that regulates this continuous phonic flow is 'rhythm':

Il principio di base del continuum fonico, che ne garantisce producibilità e percepibilità, dalla sillaba all'unità tonale e oltre, è infatti quello del ritmo: alternanza tra periodico e aperiodico, tra arsi e tesi, tra lungo e breve, tra picchi e avvallamenti di sonorità intrinseca, di intensità, di durata, di altezza (Albano Leoni 2009: 46).

Albano Leoni maintains that prosody generates meanings, not only because it conveys grammatical meanings, but also because it 'conveys the manifestations of emotions, moods, attitudes, intentions, judgments, evaluations' (Albano Leoni 2009: 50). In a footnote, the author invites the reader to observe that these aspects of prosody are also aspects of voice (cf. ibidem).

For example, as for voice, for prosody there have been different attempts to create groups of categories to summarize the infinite number of meanings conveyed by the continuous flow. This is, according to the author, the best way to approach a ‘grammar of prosody’ (cf. Albano Leoni 2009: 50).

One of the most successful examples of it, according to the author, is the attempt by O’Connor and Arnold (1961: 30-71; in Albano Leoni 2009: 50-51): they identified 10 basic prosodic patterns; for each of them realizations in the categories of ‘statements’, ‘questions’, ‘commands’, ‘interjections’ were identified; it was also found, for example, that utterances of group 1, of the category statements, convey a series of attitudes (“cool, calm, phlegmatic, detached, reserved, dispassionate, dull, possibly grim or surly”).

Halliday (1967: 9; in Albano Leoni 2009: 51) proposed a series of 5 tones that in the continuous flow of speech are selected by the speaker and can be analysed in 5 correspondent tonal groups. He considered “intonation in English as meaningful” (Halliday 1967: 10; in Albano Leoni 2009: 52), and “intonation systems” as part of grammar, and “not as if they merely carried a set of emotional nuances superimposed on the grammatical and lexical items and categories systems” (Halliday 1967: 18; in Albano Leoni 2009: 52).

More clearly, according to Albano Leoni, this approach to prosody as a significative aspect of language was shared by Crystal (1969: vii, 282-308; in Albano Leoni 2009: 54) and Bolinger (1986, 1989; in Albano Leoni 2009: 55). Crystal took into consideration ‘non-segmental models of phonology’ and the ‘interrelations’ among ‘intonation’, ‘gestures’ and ‘context’; while Bolinger paid a special attention to the ‘constituents of intonation/pitch contours’ and to the description of the execution of these contours in different communicative situations.

In addition, there are other reasons that lead to consider prosody as an ‘active component of signification and constituent of meaning’; for instance the ‘role of the receiver’ of the message, who activates his/her ‘systems of expectations’ (Albano Leoni 2009: 55). In addition, lexicon should also be considered as an aspect involved in the perception of prosody, even if research has not yet paid much attention to it (*ibidem*: 56). There are evidences, instead, that seem to prove that prosody plays a fundamental role in the ontogeny of language, since the first experiences of segmentation of the continuous phonic flow by children, as well as their first constructions of meanings, are based on prosodic information (Mehler and Dupoux 1990: 216-

219; Hawkins 1999: 195; Ducard 2002: 134-135; in Albano Leoni 2009: 56-58).

### 3.0.2 Voice

As it happens for prosody, voice is difficult to reduce to ‘discrete and segmentable traits’ (Albano Leoni 2009: 63). For the same reason, voice traits can be approached only through a ‘gestaltic, physiognomic, holistic’ perception (ibidem: 64). This is also due to the fact that voice, besides being the mean of human communication, is a powerful mean of expression: it immediately refers to the ‘individuality of the subject’ and to his/her own language (‘execution’ or *parole*). The voice conveys information on the person of the speaker, her/his physical, psychological and social status; hence the physiognomic and gestaltic traits. According to Laver (1980: 2; in Albano Leoni 2009: 64), “voice quality, as a major vehicle of information about physical, psychological and social characteristics of the speaker, has a vital semiotic role to play in spoken interaction”. Maybe for the kind of information it conveys, voice ‘holds a symbolic power (common to cultures distant in time and space) in virtue of being the manifestation of an interiority otherwise unattainable’ (ibidem). Being emitted from the body, the voice is projection of the body, it carries the signs of the body; therefore, its essence is corporeal, tactile, sensual (cf. Zumthor 1992: VIII; Bologna 1992: 65; Barthes 1973/1975: 65-66).

For its physical essence as voice-of-the-body, the voice reveals the speaker’s uncounscious, her/his feelings, emotions, moods, attitudes and feelings; in this respect, the voice configures itself as “gesture”, and “vocal gestures” may configurate a “vocal style” (cf. Fónagy 1983: 160-169). It is not by chance that Fónagy, to study the vocal expression, integrated acoustic and physiologic analysis, that helped him to identify the ‘phonetic modifications’, generated by ‘basic emotions’ and ‘defined attitudes’; while the method of ‘semantic tests’ helped him to trace the ‘psychic content’, through the listener’s perception of speech (cf. Fónagy 1993: 8-16). As projection of body, the essence of voice, besides being physical, is also metaphysical (cf. Bologna 1992: 23ff); hence, it acquires a strong symbolic power that reaches a divine status. Therefore, voice, besides revealing the body, it also reveals the soul. Studies in anthropology revealed that for primitive cultures, sound, and consequently the human voice, was the direct manifestation of spirit. The metaphysical essence of human voice comes from the sonic substance that,

according to primitive cultures, is present in all things, and is the trace of their original creation by the creator divinity (cf. Schneider 1970: 150ff). From this same trace, unique for any creature, comes the sonic strength of words, which is, for primitive cultures the real symbol of voice, and not what the voice articulates in language. The importance of words resides in its acoustic nature, and not in its linguistic or semiotic function; to this respect, the onomatopoeia is already a transposition of sound on a linguistic plan, and not the mere sound, it is a sonic symbol, therefore without force (cf. Schneider 1970: 163-165).

Magic formulas (which comprehend invocations, singing and cry) are rendered effective by the force of words, i.e. their sound; for instance, words are not important for their meaning, but for their ability to act, to have an effect through their sound (cf. Combarieu 1982: 147-148). According to primitive cultures, ‘the vocal gesture is not a simple way to excite the air, it is a mean of execution’; for this reason, ‘the use of voice is a sign of election’ (Combarieu 1982: 147). For instance, the voice of the shaman invoking or singing coincides with the voice of our ancestors or god, having the power of acting on spirits, things, human beings and events (cf. Schneider 1970: 159-160; Combarieu 1982: 24). The first vocal action undertaken was the act of creation, since god created the world through a sound emanation; therefore, as a reproduction of the act of creation, the voice-in-the-word creates all that it indicates, designs or connotes: anything the listener can give it a sense (cf. Schneider, 1970: 20-24).

Voice was identified with singing because the first forms of songs were full of groups of sounds and syllables whose signification was not verbal but ritual, unintelligible to the same singers; syllables and sounds were functional to melody, rhythm, repetition (cfr. Combarieu, 1982: 17-18). Thus, in songs, the sound (rhythm, melody) prevails on words that are dissolved in it; the vocal force makes words magic; this explains why *chants* are *enchanting* (cf. Combarieu 1982: 17-18, 34-37). The ancient greek word *ōidē* “ode” designated both “magic formula” and “singing”; at the time of Homer, the “oral epic poets and musicians” *aoidos* were also “thaumaturgists” or “miracle-worker” (Combarieu 1982: 36).

The above-mentioned contributions from anthropology may help to explain why so many documents prove the importance of voice in religion, law, magics, poetry, pedagogy, linguistics, oratory, acting and, lastly, in philosophical and linguistic reflection on the origins the nature of language

from ancient times [...] to the XVIII century (Albano Leoni 2009: 65). Voice is important also in everyday life of human activity, for the infinite number of senses, meanings, references, uses of the term, mainly for being an instrument of communication, action, seduction, persuasion, social activity (rites, prayer, rites, etc.), poetry, etc., it is the manifestation of inner states and identity (ibidem: 65-66). For all these reasons, the voice has preserved its power of generating meanings *per se*, i.e. its representative, symbolic and evocative ability, not necessarily is linked to language (ibidem: 66-67). This power of voice, that was strong in primitive cultures, mainly oral cultures, has remained present in populations that developed writing, and possibly will be even more relevant in the new media (ibidem: 66).

According to Albano Leoni (2009: 69-70), the significative aspects of voice can be referred to two main themes: phonosymbolism and manifestation of emotions and moods. Phonosymbolism is based on the power of voice to generate meanings sometimes additional other times autonomous in a holistic or iconic dimension where signifier and signified coincide (ibidem: 70). It comprehends all those phenomena that try to explain a ‘natural (thus universal) relationship between sounds and others sensorial perceptions or knowledge categories (sizes, movement, inside, outside, high, low, etc.)’ (ibidem). Therefore, the phonosymbolism is generated by an ‘integration of interrelated sensorial experiences’ (ibidem: 71).

At the basis of the interrelation would be the similarity of forms of stimuli, similar forms from stimuli of different mode; this is a way to explain *Gestalttheorie*, according Péterfalvi (1978: 45-46; in Albano Leoni 2009: 71), who considered the sonic symbolism as a particular case of synaesthesia. According to the same author (Peterflavi 1978: 47-55; in Albano Leoni 2009: 71), and beside the *Gestalttheorie*, there are other proposals explaining the nature of the association in synaesthesia, i.e. the cultural nature (e.g. traditions), or the physical nature (e.g. small objects produce acute sounds). Researches on phonosymbolism have been conducted mainly in the fields of poetry and philosophy, only marginally in linguistics; nevertheless, testimonies of such associations have been collected also by linguists, for example Fónagy (1983: 57-106; in Albano Leoni 2009: 71).

Albano Leoni (2009: 71) maintains that ‘the manifestation of phonosymbolism represents a primordial, natural component of communication that precedes history and is radicated in biology’. However, it is not possible to clearly define to what extent such component is still present

in the processes of signification, especially because experiments conducted so far on this field are based on tests with closed questions; therefore, their results hardly can be generalized (*ibidem*). To this respect, and with reference to the present study, it was decided to follow the Interpreting Studies research line of questionnaire-based surveys on quality assessment of interpretation (see chapter 1). A questionnaire was built up after defining a theoretical paradigm, of which the questionnaire is an application. The questionnaire for the pilot survey is “a test with close questions”; however, the questions were defined on the bases of semiotic, linguistic and psychological theory; moreover, the closed questions ended with a “semantic differential” (cf. Osgood 1952; in Marradi 2007: 158ff), an instrument still used in psychology of perception (Marradi 2007: 158; see section 4.5.1). The questionnaire included 19 questions related to 3 different simultaneous interpreting speeches (one with an experimental variable). Data from the survey were statistically treated, so that, in the end, it was possible to reach a synthesis of the aspects elicited by questions, through the analysis of similar responses for the three interpretations.

As to the voice as manifestation of emotions, feelings, moods and attitudes; the main problems for research, according to Albano Leoni (2009: 74) are: i) how to segment and label emotions and feelings; ii) how to identify which articulatory, prosodic and consequent acoustic indices correspond to a specific emotion expressed by the speaker; iii) how to identify – among the indices that constitute the acoustic event – those generating a particular impression. In addition, it is difficult to find definitions of emotions and attitudes that are shared by different subjects. For instance, emotions are continuous phenomena that are not defined by ‘categorical limits’; therefore, how to be sure that the same phonic manifestation can be defined by different subjects as worry, fear, scare, fright, panic? (*ibidem*).

However, all this problems have been tackled in research on voice (e.g. in Magno Cadognetto, Cavicchio, Cosi 2008; Magno Caldognetto, Zmarich, Ferrero 1998; in Albano Leoni 2009: 75), and one of the best examples, according to Albano Leoni (2009: 75) was that by Fónagy (1983). This researcher executed a perceptive verification and assessment of both sent and received emotions; this approach allowed to create subjective but shared scales of labels, whose definition rest on physiological, psychological and sociocultural bases (cf. Albano Leoni 2009:75) (see below – section 3.6).

### **3.0.3 Linguistic vs. paralinguistic (segmental vs. suprasegmental)**

Voice and prosody share their nature of continuous signals that are difficult to be represented and segmented. For both voice and prosody, there have been attempts to represent their aspects through lists of labels, defined after experiments of perception (e.g. O'Connor and Arnold 1961:30-71 and Halliday 1967: 9; in Albano Leoni 2009: 50-51, for prosody; and Fónagy 1983, for voice – see above).

The labels have been proposed because both voice and prosody convey meanings related to the speakers' attitudes, emotions and feelings. These meanings are 'interwoven' with language, since they are 'executed simultaneously with language structures' (Albano Leoni 2009: 74). According to Halliday (1967:10; in Albano Leoni 2009: 52), the "emotional nuances" are not the only kind of information "carried" by "intonation systems"; and the "emotional nuances" cannot be considered as if they were "superimposed on the grammatical and lexical items and categories". He proposed "the choice of a tone [...] is an independent grammatical selection of its own right" (ibidem: 18; in ibidem). In fact, subjects' reaction to melodic contours, through the assessment of scales, represent an evidence of "the native speaker's awareness of intonation as meaningful in the language" (Halliday 1967: 47-48; in Albano Leoni 2009: 54). If prosody or voice labels can be lexicalized, or if "intonational patterns can be systematized into a formal grammatical statement", then it means that they convey meanings "link[ed] with other grammatical choice" (Halliday 1967: 47-48; in Albano Leoni 2009: 53-54). Albano Leoni (2009: 68, 44) maintains that it is true that the attempts to represent meanings conveyed by voice and prosody have a limitation, i.e. the non-universality of their signification, thus their non-objectivity; nevertheless, it is clear that 'speakers share perceptual experience' and 'knowledge of the mechanisms of signification'.

Prosody is interwoven with syntax, mainly through the complex processes of focalization, highlighting, information dynamics, structuring in units (Albano Leoni 2009: 44-45). Prosody is also connected to lexicon, since sometimes prosodic decoding can depend on lexicon (cf. Crystal 1969: 284; O'Connor and Arnold 1971: 61-62; in Albano Leoni 2009: 56).

Given the physical interrelations of voice, prosody and language, especially considered the concept of voice as gesture, and the correlation between prosody and proxemics, the distinction between linguistic and paralinguistic should be overcome. According to Albano Leoni (2009: 48) the



distinction between *linguistic* and *paralinguistic*, that overlaps to that between *segmental* and *suprasegmental*, concerning descriptive categories, is supported by two arguments. The first argument is that the segmental sequence (ideally in its written form) as a fixed structure, primary structure on which prosody is applied (ibidem). The second argument is the theory of the two parallel channels, one for the encoding of the segmental sequence, the other for the prosodic encoding. In both cases, the prosody level or channel (suprasegmental) is subordinate to the segmental one (ibidem).

The author also proposes counterarguments to these models based on a hierarchy of two channels. The first one is taken from the neurology, the second one from linguistics, the third one from semiotics.

1. Cognition and emotion are not divided from the point of view of their respective single meanings, neither they are from the perspective of nervous structures supporting them, which are strongly interconnected, functionally and structurally (Oliverio, 1999: 53-54; 101-105; in Albano Leoni 2009: 49).

2. From the linguistic point of view, an utterance necessarily reflects an attitude (Bolinger 1989: 67-68; in Albano Leoni 2009: 49) and it cannot exist without a prosodic pattern (Bally 1950: 77-78; in Albano Leoni 2009: 49). The *phoné* can exist without *dialektós*, but *dialektós* cannot exist without *phoné* (Lapsia 1997: 59-69; in Albano Leoni 2009: 49).

3. Thinking of a general semiotic model, suitable to comprehend the linguistic manifestation of voice, prosody and language, Albano Leoni (2009: 76) refers to Bühler's *Organonmodell* (Bühler 1934/1983: 77), where the three functions of language (expression, appeal and representation) converge in a unique sign. According to Albano Leoni, the researcher that tried to find a unitary interpretation of the significant and the signified was Fónagy (1983; in Albano Leoni 2009: 76).

To this respect, it is worth to remember that auditory perception activates analysis and decoding process, but the linguistic signal almost simultaneously activates superior cognitive activities of analysis of the context, cotext, speakers' expectations; these cognitive activities overlap to the mere auditive perception and guide it in decoding the message according to what could be called idiolinguistic schemas (Albano Leoni 2009: 56; based on Bally, Aubergé e Morlec 1997: 22-24 and Bregman 1990) .

On the more strictly linguistic plan, another proposal that eliminates the dichotomy between linguistic and paralinguistic is the one based on the two-

fold sign by Rossi (1999: 9; in Albano Leoni 2009: 76-77), where intonation is strictly linked to morphology, pragmatics and syntax:

Notre définition de l'intonation, et ce que nous venons d'en dire, laissent entrevoir que les primitives de l'intonation sont des *signs à deux faces*, des morphèmes intonatifs, dont le contenu est donné par les dispositifs supérieurs que sont les intonèmes définis par des traits construits sur l'espace 3 dimensions (fréquence fondamentale, intensité, temps) [...]. Notre approche de l'intonation dérive par conséquent d'une conception hiérarchique de la prosodie où les catégories sont le résultat d'une association étroite entre le modul intonatif et les modules linguistiques supérieurs. Elle se range dans la classe des théories morphologiques de l'intonation ; elle s'oppose par là aux théories dites phonologiques dont les partisans affirment l'autonomie de l'intonation conçue comme une représentation de la mélodie définie en termes de phonèmes tonales.

Another proposal considered by Albano Leoni is that of the semantic mode by Benveniste (1974; in Albano Leoni 2009: 77). The semantic mode is based on the “signifiante”, i.e. ‘process of generation of meanings, which is specular to a process of interpretation, both processes resulting from the inseparable connection between symbolic and indexical fields (Albano Leoni 2009: 77-78). The semantic mode belongs to the theory of enunciations and works in the discourse (Benveniste 1974: 81; in Albano Leoni 2009: 77).

The semantic mode was further developed by Meschonnic (1982: 69-115, 215-218), who proposed a concept of language totally formal, where not even the dichotomy of form and substance works, because language identifies with rhythm, and rhythm with discourse; therefore, rhythm is language-as-form; and rhythm is the form of a discourse, and it includes all levels and aspects developed historically by linguistics. Not only grammar, lexicon and prosody, but also voice, proxemics, gestures, history of human being, history of each subjects, her/his biology, physiology and psychology, his/her personal rhythm (language-as-form). In this perspective, the concept of language-as-form (rhythm) is anthropological, historical, and democratic, since it tolerates any form of writing or speaking of each subject (pluralism), being it the result of the history of human kind, and the personal history of each subject (*ibidem* 15-33).

### 3.1. The gestaltic perception and production of speech: Bühler's Theory of language

After considering the aspects of prosody, voice and the criticism to the dichotomy *segmental/suprasegmental*, perception of speech from a gestaltic point of view was studied in order to find how reception of phonic material is processed in communication.

The entire title of Bühler's work is "Theory of language. The representative function of language" (1934/1983 and 1934/1990), this means that the dominant function of language is the representative one, and it is dominant with respect to the other two functions, i.e. expression (*Ausdruck*) and appeal (*Appell*). The relevance of this work to the present research is mainly due to the highly comprehensive approach to the study of language, ranging from semiotics, psychology, biology, medicine, philosophy and linguistics. In this approach, vocal and prosodic aspects are considered significative features of language, and essential in communication, ascribed to the expressive and appeal functions; however, the main thesis of Bühler's work is that all aspects of human language end up to be representational, since the power of this function is so strong that attracts to his domain the other two functions.

#### 3.1.1 The axiom system

The aim of the author was to propose a well-structured work to understand the nature of human language; for this reason he proposed a Theory<sup>1</sup>.

To this aim, Bühler, first identified the two main requirements that, according to his proposal, any theory of language had to fulfill: i) the definition of specific linguistic observations; and ii) the description of the heuristic ideas that lead the linguistic inductions (cf. Bühler 1934/1983: 65).

The author proceeded axiomatically to build up his proposal. According to Bühler, linguistics must be based on observation, and its scientific status depends on the reliability and the precision of its methods of determination of facts. The linguist observer must comprehend what is perceived with senses of the concrete speech event; such comprehension must

---

<sup>1</sup> A theory is "a (scientific) a systematic ideational structure of broad scope, conceived by the imagination of man, that encompasses a family of empirical (experiential) laws regarding regularities existing in objects and events, both observed and posited – a structure suggested by these laws and devised to explain them in a scientifically rational manner" (*The New Encyclopædia Britannica – Micropædia*, 1981, vol. IX, pag. 941).

be subjected to a methodological verification. The concrete speech event is the starting point for the linguistic analysis (ibidem: 69). The linguist, just as any other natural scientist, must carefully select what is to be observed. Therefore, just as in other natural sciences, also in linguistics, huge amounts of data can be mastered and calculated through conceptualization. Such conceptualization can be successful only through a research on the modalities and motivations of the linguist's conceptual world (ibidem: 69). This is the general task of a theory of language as a part of theory of science, according to Bühler; and this task had never been conceived or realized before as a thorough critical analysis of concepts or in comparison with conceptual systems of other disciplines (ibidem: 69-70). This meant, in the view of the author, starting from the basic principles. Therefore, he decided to use axioms<sup>2</sup> in such a way that these be a purely phenomenological – and consequently, gnoseologically and ontologically neutral – identification of basic principles. The identification process can be seen as a reduction process, drawing from the research in linguistics (Bühler 1934/1983: 73). Bühler maintained that the axiomatic method was similar to that applied by Socrates in the so called socratic induction; for instance, Socrates asked skilled handicraftsmen, talked with them, and in this way discovered the principles at the basis of their ability (ibidem: 73). According to the author, Russel and Hilbert interpreted the investigation on the principles of empirical sciences in such a way that, once the results and the available theories were acquired, these were subjected to a process of logic reduction; this was considered the first step of the axiomatic thought. Bühler added that it was also necessary to provide a justification (ibidem: 73-75). The instrumental model of language proposed by the author, in his opinion, presented all the essential traits that could be identified in the concrete speech event. This was the first axiom identified by Bühler in his proposal for a theory of language. The series of all axioms identified is the following (Bühler 1934/1990):

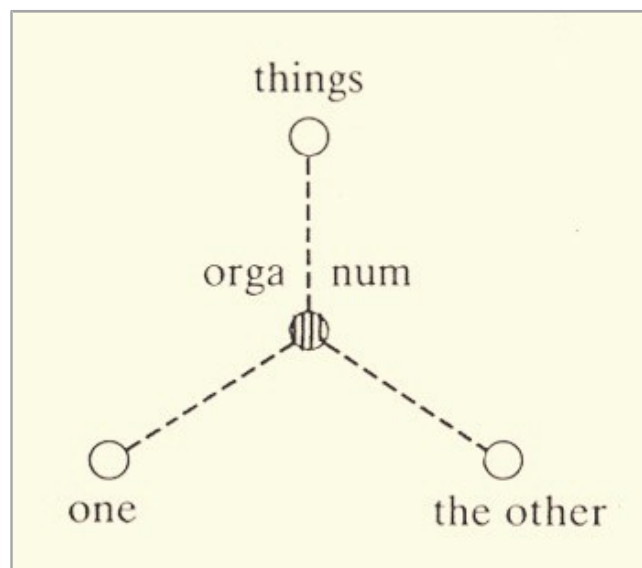
- The model of language as Organon (A)
- The significative nature of language (B)
- Speech action and language work; speech act and language structure (C)
- Word and sentence. The S-F-System of the Type Language (D)

---

<sup>2</sup> An axiom, “according to Aristotle, [is] an indemonstrable first principle from which all demonstrative sciences must start. In addition, each particular science has its own particular first principles on which it builds [...]” (*The New Encyclopaedia Britannica – Micropædia*, 1981, vol. I, pag. 689).

### 3.1.2 Axiom A: The model of language as Organon

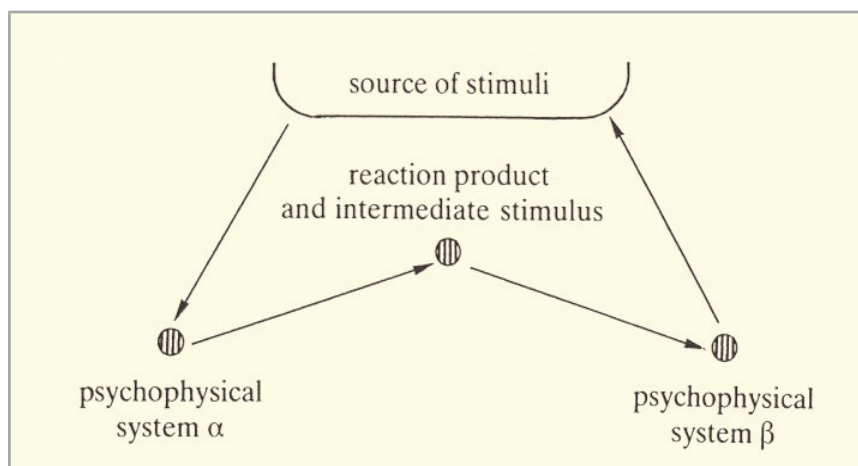
To build this first axiom, Bühler started from Plato's definition of language as an instrument (*organon*), through which two interlocutors communicate on things. All other cases may be derived from this event through a 'reduction' process, since language communication is the richest form of speech event. Two interlocutors and the things about which they communicate make three basic relational elements (Bühler 1934/1983: 77).



**Figure 3.1.** Complex casual connection of elements related to speech, through the mediation of interm-ediante elements (Bühler 1934/1990: 31). One of the interlocutors (*efficiens*) produces an acoustic phenomenon and acts as a stimulus on the other (*effectus*). The central element is the acoustic phenomenon perceptible by senses (cf. Bühler 1934/1983: 77).

Suppose that the speaker's production of the acoustic phenomenon be stimulated (through a preliminary acoustic stimulus) from an object in the perceptual field, and that the listening of such linguistic phenomenon stimulates the hearer, who turns her/his look at the same object – this constitute the causal chain of primary phonic communication. In this model, there are two psychophysical systems, each of them is both 'receiver' and 'sender'. As receivers, they are 'selectors', working on the basis of the principle of the 'abstractive relevance'; as sender, they work as 'producers of signals; in both cases, they take part to the mechanism of interchange of signals, where

the reception of the stimulus constitutes a notice, while the production of signals, an activity (Bühler 1934/1983: 77-79).



**Figure 3.2.** Causal chain of primary phonic communication (Bühler 1934/1990: 32).

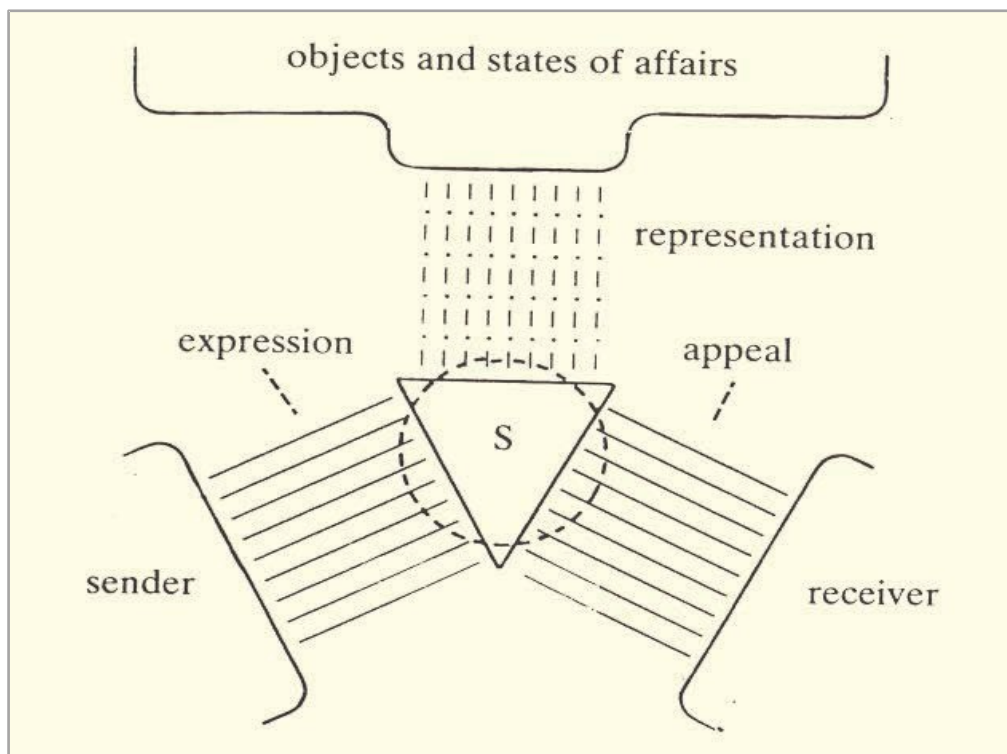
These two steps led Bühler to define his model ‘instrumental’, where the concrete acoustic phenomenon manifests itself in three moments, corresponding to three kinds of sign: ‘symptom’, ‘signal’ and ‘symbol’. Each kind of sign develops a specific function. Thus, the concrete acoustic phenomenon is ‘symptom’ when it develops a ‘notification’ (indexical) function and is focused on the sender, because it is expression of her/his inner states; it is ‘signal’ when develops an ‘appeal’ function and is addressed to the receiver; it is ‘symbol’ when develops a ‘representation’ function and is related to objects and facts. The perception of the concrete acoustic phenomenon is processed through the principle of the ‘abstractive relevance’, i.e. only those aspects that are diacritically relevant are grasped for the comprehension of the sign. In addition, the sensitive datum is always subject to a non-perceptual integration by the receiver (ibidem: 81-82).

According to Albano Leoni (2009: 76), Bühler’s *Organonmodell* represents the ‘most complex and potentially fruitful model’ among the ‘general semiotic model[s]’ created in linguistics to give account of all aspects of language, going beyond the distinction between linguistic and paralinguistic features. This model was considered by Trubeckoj, Jakobson (1960; in Albano Leoni 2009: 76) and Scherer (1995, 2002; in ibidem).

The relevance of Bühler’s *Organonmodell* to the present study is that in it the symbolic system (*darstellt*), the world, the speakers (the sender and the

receiver) are indissolubly woven together, and the speakers are not users or spectators of the representation of the world through language (*Darstellung*), but are its creators and developers (Albano Leoni 2011: 128).

This aspect would allow a redefinition, or even a demolition, of the dichotomy paralinguistic/linguistic, and, at the same time, the inclusion of intersubjectivity in a linguistic theory; where the intersubjectivity is the shared world reaching the heart of language through the speaker (*ibidem*).



**Figure 3.3.** Organon model of language (Bühler 1934/1990: 35). The central circle represents the concrete acoustic phenomenon, the triangle the three moments of its perception, corresponding to three kinds of sign: symbol for representation, symptom for expression, and index for appeal. The triangle includes something less than the circle in virtue of the principle of abstractive relevance, and something more in virtue of the principle of non-perceptual integration. The three groups of lines symbolize the semantically defined relations of the linguistic sign. The broken lines symbolize the indefinite similarity between sounds and things.

### 3.1.3 Axiom B: The significative nature of language

In this part of the “Theory of language” (Bühler 1934/1983: 86-100; 1934/1990: 40-56), the author explains why “language phenomena are significative” and consequently have to be studied through sings. The

“acoustic image of a word is constructed as a sign”, it is made of “elementary characteristics”, the phonemes, which allow to distinguish the acoustic image of a word from similar acoustic images. The phonemes “function as *notae*, features”, and are “differentiating signs” of the acoustic image of a word (1934/1990: 40). At the same time, a word may also function as a “sign for an object”, indicating something or someone in the context (“deictic sign”) and/or referring to an object (“conceptual sign”) (ibidem: 41-43).

As to the deictic sign, the author recalls that animals, and therefore also human beings, have a “psychophysical system” that naturally functions as “receiver and user of signals” (ibidem: 45-46). This biological production of signs by animals can be identified, ‘in their highest level of social life, when a cooperation among individuals demands an expansion of the horizon of common perceptions’ (Bühler 1934/1983: 91):

Productive and creative man, working in cooperation with his fellows, often remains silent as long as each fully understands what the other is doing and acts appropriately. But then a constellation comes up to which our descriptive formula applies, and one partner opens his mouth. Sometimes one word, a random linguistic sign such as ‘(to the) right’, ‘straight on’, or ‘this one’ or ‘stalls, sixth to ninth row’ is enough to give the supplementary guidance needed for the receiver to act appropriately. These are human utterances that will later be described as being empirically integrated. To put it metaphorically, they occur much as do properly placed signposts on human paths; as long as there is only one unambiguously recognizable way, there is no need of signposts (Bühler 1934/1990: 46-47).

As to the conceptual sign, the author recalls the Scholastics’ formula *aliquid stat pro aliquo* (something stands for something else); however he prefers to adopt the general model of proxy (“representation in the sense of standing for” – *Stellvertretung*), because this concept allows him to examine the phenomenon “in terms of relationship”:

Whenever something is a sign, it is only abstract factors by virtue of which the concrete thing functions “as” sign. I have called this fact, which is fundamental for the theory of language, the *principle of abstractive relevance* (Bühler 1934/1990: 48).

Bühler adds that “the proxy element of the complex (*id stat pro aliquo*) always belongs to the realm of what is perceptible”, while this is not true for the other element (ibidem).

In conclusion, the argument of the signs as “intersubjective intermediaries” used as an “implement for orientation in community life” is



consistent with that of human language as an “implement”, or, in Platonic terms “organon”, and with the “paradigm of *homo faber*, a maker and user of implements” (ibidem: 48, 56).

### 3.1.4 Axiom C: Speech action and language work; speech act and language structure

In this part, the author maintains that “speech action” and “speech act” concern the “subjective-related” dimension of language, while “language work” and “language structure” (“form”) concern an objective dimension of language, or “phenomena that have an intersubjective fixation” (Bühler 1934/1990: 58). Speech actions and language works have a “lower level of formalization”, while speech acts and language structures (“form”) have a “higher level of formalization” (ibidem).

	I	II
1.	An	W
2.	Ac	S

**Figure 3.4.** “Intersecting dichotomies” of axiom C (Bühler 1934/1990: 58).

The speech action is related to other attitudes of human being, including non-linguistic expressions, and any speech action has to achieve its purpose, i.e. the communicative intention of the speaker achieved through the speech act. Conversely, the language work is no longer linked to the speaker, to his/her activity and aim, because it is the “product” of the “person’s work” and becomes “independent” from it (ibidem: 60-63). The speech act is related to the speaking subject and is an act conveying a sense (ibidem). The speech act is internal to formal conventions of language (“structure”/ ‘form’); the context is fundamental to define its meaning (sense) (ibidem: 67-80).

### **3.1.5 Axiom D: Word and sentence. The S-F-System of the Type Language**

Strictly related to the axiom A is the axiom D, according to which languages are symbol-field (*Symbol-Feld*) systems, where the two phenomena stand in a relationship of mutual presupposition, or interdependence; and where the word is the symbol and the sentence the field in which the words place themselves (Albano Leoni 2011: 128). Indeed, both words and sentences are linguistic forms, which cannot be considered as categories, since they are mutually implied, and can only be defined in terms of co-presence (Bühler 1934/1983: 122; in Albano Leoni 2011: 128). In addition, the ‘semiotic power of languages is given by the synergy of fields’ (symbolic and deictic; see below – next section), ‘and the principle of the two-class system (S/F): symbol-word and field-sentence’ (Albano Leoni 2011: 129):

A code of symbols, whether written or unwritten, must be limited, as is the written or unwritten lexicon of a spoken language, and for the same reasons, by reason of the limited capacity of human memory. Especially when the problem is that a number of people must make a mental note of the individual coordinations so that they can be used in intersubjective communication with sufficient diacritical precision and fluently, the capacity limits are not all very generous. [...] It is not because we and our fellows are acrobats of mnemonics that there are practically no limits to the novelties that we can represent in language such that everything is intersubjectively understandable, but because such acrobatics are not required by a field system of the type language (Bühler 1934/1990: 88-89).

### **3.1.6 The deictic field and the symbolic field of language**

According to Bühler, human language is a system of signs operating in connection with a field. The author identifies two main fields: deictic and symbolic. The deictic field concerns the situational moments of a concrete speech act; in human language it is based on three coordinates: here, now, I. The deictic field is related to deictic words, used to follow the voice, or the gestures; for example, words or phrases like in “front of/behind”, “on the right/left”, “on/under”, etc. By using these words, the speaker refers to her/his person, his body, involving – in the indication – his/her “(conscious, experienced) *tactile body image*” (Bühler 1934/1990: 145) related to the visual space, even though orientation is not limited to the sense of view:

When the same person uses words like *in front-behind*, *right-left*, *above-below*, another fact becomes apparent, namely the fact that he senses his body, too, in *relation* to his optical orientation, and employs it to point. His (conscious,

experienced) *tactile body image* has a position in relation to visual space. Spatial orientation in animal or man can never be an affair of the visual sense conceived in isolation. Otherwise we should not be able to understand quite a few well-known facts. Of the human being we know that data from the visual, the tactile and the aural sense are registered and evaluated *together* by the recording device just mentioned, and that further endogenous contributions are also registered there, contributions of so-called kinæsthesia, made by movements of our head and body (Bühler 1934/1990: 145).

The author identifies three types of indication: the *demonstratio ad oculos*, the anaphora and the *Deixis am phantasma* (“imagination-oriented deixis” – Bühler 1934/1990: 140). The *demonstratio ad oculos* (“ocular demonstration” – *ibidem*) is the visible indication, immediately reached in the physical situation of the subjects involved in the communication act through the use of deictic words and/or deictic gestures; it is the immediate (non-mediated) result of the analysis of the situation by the actors of communication (Bühler 1934/1983: 131-132, 154-156). This is an example proposed by Bühler:

If I am standing as the commander nose to nose in front of gymnasts lined up in a row, convention requires that I choose the command ‘forwards, backwards, right turn, left turn’ not according to my own orientation system, but to that of the others, and the translation is psychologically so simple that every group leader learns to master it (Bühler 1934/1990: 118).

In the *demonstratio ad oculos*, the position of the speaker, the identification function developed by his/her voice, the orientation conferred by the voice to the receiver assumes a value in the deictic field, conveying the information related to the coordinates of I-here-now. The meaning (sense) of the communicative act is strictly related to the appeal, the expressive functions of language, and the position of the subjects in the situation (1934/1983: 154-156).

The deictic field is present also in discourse through deictic words or phrases like ‘as mentioned before’, ‘this’, ‘therefore’ etc., used by the speaker/writer to indicate parts in the global context of discourse (Bühler 1934/1983: 132, 173-176). If the *demonstratio ad oculos* activates the orientation in the intersubjective-spatial-temporal situation of the communication act; then, the anaphoric deixis activates the orientation in the linguistic flow of discourse, through deictic terms (*ibidem*). In this case, the deictic field is strictly connected (overlapped) to the symbolic field (*ibidem*).

Seen from a psychological perspective every anaphoric use of deictic words presupposes one thing: that the sender and the receiver *have the flow of speech in*

*front of them* and can reach ahead and back to its parts. It must hence be given as a whole to both the sender and the receiver to such an extent that wandering is possible, comparable to the wandering of the gaze on an optically present object. All of this does not surprise the psychologist; for he knows that not only the flow of speech but also other formed acoustic sequences demand and permit such wandering along the sequence, resumption of what is past and preconstruction of what is about to come. The adequate production and reception of every piece of music, for example, requires something similar, though it is not exactly the same. (Bühler 1934/1990: 138-139).

When the *demonstratio ad oculos* is reported through a narration, hence, with the use of also the anaphoric deixis, then we have a *Deixis am phantasma* (literally: demonstration from what is shown, or from the mental image). In this case, the physical situation of a communication act is transferred to a fictional level, and the author of such transference (the narrator), in order to orientate the receiver towards a phantasmatic object, uses deictic terms just as in the *demonstratio ad oculos* (Bühler 1934/1983: 176-181). The only difference is that the receiver uses his internal or inner eyes and ears, and not the real senses, to follow the indication with his/her eyes, or hear the voice to recognize the speaker and his/her position in space (*ibidem*); thus, deictic terms are used to evoke the scene :

It is not at all the case that imagination-oriented deixis *completely* lacks the natural deictic clues upon which ocular demonstration is based. Rather, the speaker and hearer of a visual description of something absent possess the same talent and resources that permit the *actor* on the stage to make something that is absent present and which permit the *audience* to interpret what is presented on the stage as a mimesis of something absent. The language used for ‘perceptual’ things is completely adapted to this fictional play, and language should only be called perceptual to the extent that it uses the resources (Bühler 1934/1990: 142-143).

According to the author, the deictic field is like the scene of any act of speech, because it works as a subjective orientation system, through the deictic elements “here, I, now”, and is related to the communicative situation, not necessarily linguistic; this means that it is not necessarily words or sounds that are significative, or convey a meaning, but also silent gestures or even silence.

If the deictic field refers to the situation, based on the indication of the subject, the time and the space of the communication act, the ‘symbolic field provides the accomplished discourse with a class of operational and explicative clues that could be covered by the name context’ (Bühler 1934/1983: 201). The symbolic field is the space, or ‘surface’ where language

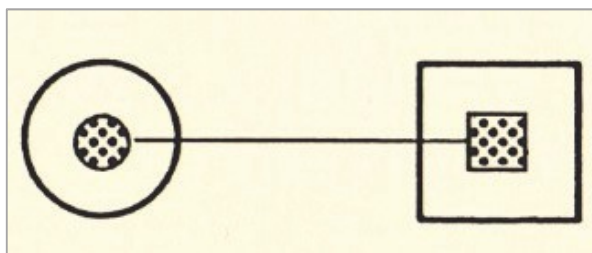
symbols are ‘arranged’; its main purpose is ‘to permit a more general and insightful interpretation of the relationship between lexical and syntactical aspects of language’ (ibidem: 203). The symbolic field is fundamental for a language sign to be symbolic or not, since ‘the field values of a language sign stem from the symbolic field’ (ibidem: 235-236). To give an example taken from the ‘symbolic fields in non-linguistic representative systems’ (cf. ibidem: 231-245):

[...] the isolated notes as they are listed in the lexicon contains no marks of the pitch. There is only one sign in the lexicon for all semibreves that are encountered in the piece of music, only one for all minims, and so on, regardless of how high or low the tones symbolized by them are. On the music-paper, it is purely a concern of the field to specify the pitch, whereas on the other hand the field has no part in the specification of the (relative) length of the tone: for the relative length of the tone is symbolized by the form of the notes alone (Bühler 1934/1990: 207).

Similarly, the ‘significative’ value, or meaning, that a (material) word is assigned to in a dictionary appears as an ‘alien’ in the field of the grammatical (syntactic) determinations (Bühler 1934/1983: 236). With respect to the theory of the symbol, the author refuses the logicians’ proposal of ‘arbitrariness’ being at the basis of “symbolization” because this is a ‘negative determination’; therefore, he proposes a “correlative definition”: “every symbol needs a field and every field needs symbols if serviceable representations are to be possible” (Bühler 1934/1983: 238-239; 1934/1990: 210-211). This means that “those signs that are extraneous to the field must also be *open* for field values with which they are vested: they must be *fieldable* [capable of taking place in a field and assuming a field value]” (ibidem). For example:

Of course I cannot transplant the lexical units of the musical notation to the map nor the geographical symbols to my music-paper, then to bestow field values upon them. The note symbol is no [capable of assuming a field value in the map field], it is not ‘fieldable’ there because it does not symbolize a geographical entity that could receive a local value (Bühler 1934/1990: 211).

Let us see how the symbolization works, according to Bühler’s proposal (1934/1990: 249-250), in the ‘linguistic-conceptual signs’, by considering the following figure:



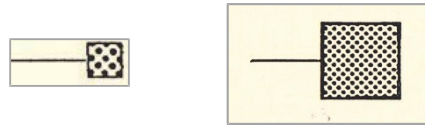
**Figure 3.5.** Gestaltic symbolization (Bühler 1934/1990: 250).

The inner circle (black) represents the *flatus vocis*, i.e. the “concrete sound material” of a word (e.g. “horse”), which is the “sensually perceptible aspect of language sign”; the outer circle (white) represents all those aspects of the sound material that are not considered relevant to the naming function of the sign, according to the principle of abstractive relevance. These aspects are related, for example, to the peculiar speaking voice of the word “horse” by different speakers; in fact, “the acoustic differences in speech are pathognomically and physiognomically significant, but they are irrelevant for the naming function of the [...] word horse”. It is the same phenomenon that happens when we use a traffic light, where the “valid conventions” are more or less as follows: red → blocked way and green → the way is free. Every signalling entity (e.g. lamp) is a *concretum*, with many determinations such as form and size, “but it is only the factor red or green as contained in the convention that is relevant for traffic and for the partners of traffic”. Getting back to the figure, the small square drawn within the larger one represents the “*abstractum*” ‘horse’ in an utterance such as “the horse is not a ruminant”, “because it has the same determinations as any concrete horse, but fewer of them”. The author maintains that “there is nothing very unusual about the fact that phonetic phenomena behave much the same in their capacity as names” (Bühler 1934/1990: 249-250); this argument will help to understand the expressive aspect of voice (see below – section 3.1.1.8).

The two squares of the figure are used by Bühler (1934/1990: 252-261) also to illustrate the concepts of naming words and connotation.

Some Scholastics explained that adjectives such as *albus*, do not note only the property, the colour factor ‘white’ [small full square], but that in one stroke a bearer of this property is also noted, con-noted, no specific thing (of course), but something or other to which the factor is to be attributed [larger empty square] Bühler (1934/1990: 252-253).

John Stuart Mill proposed that “abstract names such as the redness” and “proper names” are “naming words” that “provide no connotations” (ibidem: 252); in this case the figure should have only the small full square or one single big full square (ibidem):



**Figure 3.6.** Gestaltic symbolization of proper names according to John Stuart Mill proposal (Bühler 1934/1990: 253).

Bühler criticizes Mill’s proposal for proper names as “diacritic” (ibidem: 130) or names that “signify a subject only” (Mill 1974:31; 1877:35; in Bühler 1934/1990: 253) and lack a “symbolic function”; conversely, he proposes that proper names “have and etymon” (Bühler 1934/1990: 260), and “are attributed deictically; it is not exactly the symphysical surrounding field but something analogous that becomes relevant here” (ibidem: 263). In conclusion, for these names he proposes the form of two squares (figure 3.5), where the empty square represents a surrounding field.

### 3.1.7 Gestalt and speech

The notion of *Gestalt* (= form) perception, together with the theory of colours, is used by Bühler (1934/1983: 206-207) to explain the theory of ‘inner field’ and ‘surrounding field’. In detail, the contrast of colours is determined on the basis of the perception of a patch of colour on a surface that is influenced by the surrounding field of the same mark: the influence between inner field and surrounding field is mutual (ibidem) – each element of the figure 3.5 (above), either the two circles or squares, serve to illustrate the notion. The gestaltic principle implicit in this perception is the impossibility to reduce the whole to the sum of its parts:

[...] sense data usually do not occur in isolation, but are embedded or integrated into various mental processes as encompassing wholes in which they are correspondingly subject to various modifications. The term ‘surrounding field’ seemed appropriate and has gained currency.

It is almost a matter of course that the special group of sensible entities or sensually perceptible processes that we call the signs of language is no exception to this. [...] There is no need to prove to an expert that the most important and most interesting surrounding field of a language sign is its context; the individual sign appears in company with its fellows, and the

company proves to have effect as a surrounding field (Bühler 1934/1990: 175-176).

Nevertheless, there are indications and denominations that, apparently, have either no context or a limited context (Bühler 1934/1983: 207-210); for example, the utterance “straight on” could be enough to eliminate the ambiguity in a communicative situation where one has to decide which way to follow in front of a crossroad. In these cases, the surrounding field that on the linguistic plan is activated by the co-occurring signs in the field (co-text), is replaced by the praxis, i.e. by a system of social and cultural conventions shared by the actors of the communication. This system constitutes the “surrounding empratic context” (Bühler 1934/1990: 176), where the linguistic sign is used as a diacritic element, when it is necessary to make a choice, where the sign is just a praxis:

Before me is a set of examples collected from everyday life containing sentences in various degrees and nuances of incompleteness, broken off or with gaps, and also words completely lacking context or with only scanty context. On unbiased examination it seems to be quite unimportant whether such words are deictic particles or whether they have a naming function (Bühler 1934/1990: 176-177).

Beside the sympratic (or empratic) surrounding field, the author identifies other types of surrounding fields: the symphysical and the synsemantic (Bühler 1934/1990: 176). The surrounding symphysical field is characterized by “isolated names, that is, names lacking context” that “can be encountered *physically attached* to what they name” (ibidem: 179). This is the case of brand names stamped on products, place-names written on signposts, proper names of the owner or maker of objects on the same objects, titles of books and of chapters – all names “physically connected and affixed to what is named” (ibidem: 179-180).

The surrounding synsemantic field is constituted by gestures, expressions, co-occurring with phonic signs (ibidem 186-187). This field intervenes itself in the determination of other types of symbolic values. This intervention is activated by a second and complementary perceptive action: the first action concerns elements co-occurring in the field; the second concerns these elements and the cognitive ability of the subject in determining a semantic value at a local level. To explain the synsemantic field, Bühler draws on the semiotics of painting:



If a painter mixes the same grey on his palette three times and three times places physically the same patch of grey in a developing picture, this patch can three times (or more often) receive a different pictorial value in the context of the painting (Bühler 1934/1990: 186).

*Gestalt* psychology is mentioned by Bühler (1934/1990: 171) also to explain the perception of the relationship between form and matter in the acoustic aspect of words; in this perception, the phoneme is the diacritic element of the word, used by the hearer in cases of distorted signal or weakened communication to recognize the word; the phoneme is then an element of the surrounding field (Bühler 1934/1983: 342-345).

It is not phonology, but grammar, or better, lexicology which characterizes certain parts of the sound stream of an utterance as words and word elements. That is also one of the presuppositions of our list. Further, it is modern psychology that empirically points out that in addition to the phonetic features (=phonemes) certain Gestalt qualities are also part of the phonetic character of these formations. Just as there are larger forms, the so-called sentence melody, so too do the same formations exist on a smaller scale in the word itself. There are word accents and word melodies [...]. They have a place among the musical modulations that either can be of direct syntactic relevance (such as the so-called sentence melody) or are of relevance by way of modulation of the sound of the individual word. [...] every word has a phonetic shape, which is not entirely determined by the expressive function alone, but in part also contributes to making the symbolic value and the syntactic valence of the word (Bühler 1934/1990: 199-200).

Not only words, but also prosody has gestaltic features, and also prosodic features, in the gestaltic perception of the acoustic aspect of words, facilitates the recognition of a word:

This fact [damaged signal or distorted or weakened communication] becomes theoretically productive because we can state more or less exactly which factors and constituents of the phonetic impress are first and foremost subject to weakening, fading and distortion under the circumstances mentioned. In the terminology of acoustics it is the noises, in the terminology of phonetics the plosives that are altered before everything else. The limit of their range is quickly crossed with increasing distance from the speaker, and the telephone also weakens and distorts them. The vowel sounds are more resistant in both cases, and along with them or bound to them certain well-characterized complex, characteristics (or Gestalt qualities) also resistant: for example melody, or pitch contour of the sound system, further the rhythmical impress (strong – weak, short – long), and finally the waves of gravity and saturation of the vowels. In fact, these complex characters are often sufficient to fulfill the reduced diacritical requirements. The word images are then recognized primarily by their acoustic shape and certainly not only by their itemized description (Bühler 1934/1990: 322-323).

The author states that the notion of acoustic face of words, where the phonemes are just phonic “marks, features, criteria, *notae* [...] and correspond to the features of things” (Bühler 1934/1990: 316), is consistent with the visual perception of mime phenomena, which belong to the field of expression (ibidem: 324-326):

The evidence presented both in older and newer studies [...] indicates that within the flux of continuous changes in the face and in the larger bodily movements involved in communication with bodily expression certain pregnant factors are extracted or appear to be particularly salient. This procedure is quite familiar to sculptors and painters, who fix human expression in stone and paint; the fact that it is also quite familiar to the partners of trivial, everyday contact was proved in my book about expression. From a psychological perspective, exactly the same thing is done with the acoustic image of the word. In receiving, the hearer extracts from acoustic continuum certain pregnant factors which he needs for diacrisis (Bühler 1934/1990: 3325).

Even if expression changes, according to the emotional states, and the acoustic shape of words is modified, the “diacritical itemized description remains intact”. Similar “laws of constancy” are observed in the “constancy of the size of visible thing through change of distance”, or that “the colour of visible thing through the change of lighting”, or that of the “loudness of auditory entities through change of distance” (ibidem: 327).

It is clear that Bühler’s phonology stems from communication and the natural tendency of human beings to abstraction, and not only from the phonic signal.

### **3.1.8 Voice and expression**

It is relevant to the present study to point out the role of voice in communication, considering the role of voice in television interpreting (see above – sections 1.1.1.3, 2.2.5). The phonic material also conveys physiognomic and pathognomic<sup>3</sup> information: the former is related to possibility to recognize the speaker; the latter is related to the possibility to infer the emotional status of the speaker, or the impact of what is being said/heard on his person. In this respect, voice is expression:

The phenomenon of the sounding word is a continuum, and is susceptible of continuous nuancing in innumerable many dimensions. Let us begin with the

---

<sup>3</sup> ‘Semiotics of emotions’ (Bühler 1968/1978: 43). The word “pathos” means ‘lived experiences’, the ‘actual subject’s movements of the souls’, but also the ‘durable passions and states of the soul’; movement (voice included) is the most important factor of pathognomic phenomena (ibidem: 49).

most obvious by mentioning the gender differences between human voices: the voices of men, women and children are different, and every word sounds different from the mouth of a man or a child. That holds to such an extent that the speaking voices of a few dozen people around me form part of what I recognize the individuals by. That is, we pay attention to the *physiognomic* features in the acoustic image of a word and use them in verbal contact. Moreover, the speaking voice is a seismographically finely tuned organ of expression; we notice by it, sometimes by the acoustic image of a single word, what is going on in the sender. The phonetic image is thus capable of *pathognomically* informative modulations (Bühler 1934/1990: 324).

However, the author maintains that these aspects “must not disrupt the set of constant factors” that allow the interlocutors of a given language community to “recognize the units that can be linguistically registered (as they are entered in a dictionary)”, they “must not deform them all, or at least to such an extent that their diacrisis would become impossible” (*ibidem*). The physiognomic and pathognomic features of expression are related to the psychophysics of the subject, in particular, to vital organs, muscles and nerves that serves the eating and breathing functions (cf. Bühler 1933/1978: 76-81, 151-152), being communication a vital function for human beings (cf. *ibidem*: 81) – this argument is consistent with the principle of the instrumental function of language. The idea of the ‘direction of relationship’ observed in movements determined by ‘attraction’ or ‘rejection’ drive, (cf. *ibidem*: 82, 98, 173), once again confirms this principle. Moreover, expression is linked to the ideo-motor and cardio-motor systems (cf. *ibidem*: 85-88); visible bodily movements are symptoms of psychic states (*ibidem*: 150-153) and they have to be defined in relation to action (*ibidem*: 168, 172).

Fónagy (1983) studied the physiognomic and pathognomic information related to the voice, and proposed, with evidences, that the movements of the melodic curve (pitch contour) reflect the movements of the soul (see below – section 3.6).

Physiognomic and pathognomic features coexist with the linguistic units in the phonic material; and even when they do not “disrupt” the “linguistic units”, are perceived by the hearer all the same. Considering that expressive movements and voice are actions that reflect psychic states, and that they can be observed also in other animals (cf. Bühler 1933-1968/1978: 79, 113-114), it may be supposed that pathognomic and physiognomic information are perceived as instinctively as they are produced, that is, more directly than linguistic units.

With reference to the present research study, it may be supposed that, considered the importance assigned to the voice in television and film interpreting (see above – sections 1.1.1.3, 1.1.2.8), both the expressive and the representative features of the phonic material result amplified in perception, due to the medium. It has to be considered that the pathognomic and physiognomic features of a simultaneous interpretation are mainly constituted by hesitations, silent pauses, filled pauses, syllable lengthenings, etc., that reveal the mental states of the interpreter with respect s/he has said, is saying, translating, or not. These expressive features of course disrupt the representative features, especially when amplified by the medium. With reference to the gestaltic symbolization proposed by Bühler, the interpreter operates a double or meta symbolization: that of the source language and that of the target language. Considering the organon model, this operation is time and energy consuming, therefore the interpreter's mental process is in part expressed through hesitations, that at the same time serve the appeal function, because the listener is there, waiting for the translation.<sup>4</sup>

### 3.2 Rhythm as form (Benveniste 1966)

A long time before psychology, a concept similar to the *Gestalt* perception existed in the ancient Greek language, precisely in the ancestor of the word 'rhythm'. According to the philological proposal by Benveniste (1966: 327-335), the original meaning of the word "rhythm" is 'form' (rhythm < gr. ῥυθμός < ῥεω 'to flow'); in detail, the ancient Greek ῥυθμός:

d'après les contextes où il est donné, désigne la forme dans l'instant qu'elle est assumée par ce qui est mouvant, mobile, fluide, la forme de ce qui n'a pas consistance organique: il convient au *pattern* d'un élément fluide, à une lettre arbitrairement modelée, à un péplos qu'on arrange à son gré, à la disposition particulière du caractère ou de l'humeur. C'est la forme improvisée, momentanée, modifiable. Or, ῥεω est le prédicat essentiel de la nature et des choses dans la philosophie ionienne depuis Héraclite, et Démocrite pensait que, tout étant produit par les atomes, seul leur arrangement différent produit la différence des formes et des objets. On peut alors comprendre que ῥυθμός, signifiant littéralement «manière particulière de fluere», ait été le terme le plus propre à décrire des «dispositions» ou des «configurations» sans fixité ni nécessité naturelle et résultant d'un arrangement toujours sujet à changer. Le choix d'un dérivé de ῥεω pour exprimer cette modalité spécifique de la «forme» des choses est caractéristique de la philosophie qui l'inspire; c'est une

---

<sup>4</sup> In line with this assumption, modern methods suggest to start from the control of breathing to reach body consciousness in order to achieve a better control of (bodily and vocal) expression in speech (cf. e.g. Coblenzer and Muhar 1999/2004).

représentation de l'univers où les configurations particulières du mouvant se définissent comme des «fluements» (Benveniste 1966: 333).

As it can be inferred from the quotation, the author reaches this conclusion after a historical analysis of the use of the word, especially by the ancient Greek authors. Thus, the term is not present in Homeric poems; however, it was found in the texts by Ionic authors, the tragic and lyric poetry, and in the Attic prose, especially among philosophers (ibidem: 328). In detail, as a way of example, *ῥυθμός* was used to refer to: the form of atoms, or institutions (Democritus); the moral or physical formation (Democritus); the form of letters of the alphabet, meaning the display of the signs of writing (Leucippus, Herodotus); the individual and distinctive form of the human character (Archilochus); the personal forms of human mood or character (Anacreon); the perceptive form of pain (Sophocles); the proportioned display of richness and poverty (Plato); the form of personality (Plato) (ibidem: 328-332).

In the same Ancient Greek texts studied by the author, another similar term to refer to 'form' was *σχῆμα* ('scheme'), used differently with respect to *ῥυθμός*, in fact, while *σχῆμα* designated an accomplished, fixed form, considered as an object (ibidem: 333), *ῥυθμός* referred to a form not fixed yet, but as it was perceived/produced by the subject in a moment of its development (ibidem: 333).

How did the meaning of *ῥυθμός* change from 'form' to the actual meaning of durations and the intervals regulating the succession or the order of movements of human beings, their activities, their attitudes, or even things and events? (ibidem: 333, 325). According to Benveniste's findings (1966: 327-335), the actual meaning of the word rhythm already existed in ancient Greek, and was referred to the result of an *a priori* secondary specialization of the meaning 'form' (which certainly existed until the V century B.C.) operated by Plato (ibidem: 333). For example: in the *Philebus* (17 d), Socrates insisted on the importance of intervals in order to study music, he added that rhythms and measures were aspects inherent to the body movements, and they were subjected to numbers; in the *Symposium* (187 b), he says that rhythm is the result of fast and slow, first opposed, then harmonized, as harmony is a consonance and consonance a chord; in the *Laws* (665a), he states that the order in the human movements is called rhythm, just as harmony is the order of voice, where high and low merges, and the union of two orders is the choral art (ibidem: 334).

The conclusion proposed by Benveniste (1966: 334-335) is the following:

On voit comment cette définition procède du sens traditionnel, comment aussi elle le modifie. Platon emploie encore *ῥυθμός* au sens de «forme distinctive, disposition, proportion». Il innove en l'appliquant à la *forme du mouvement* que le corps humain accomplit dans la danse, et à la disposition des figures en lesquelles ce mouvement se résout. La circonstance décisive est là, dans la notion d'un *ῥυθμός* corporel associé au *μῆτρον* et soumis à la loi des nombres: cette «forme» est désormais déterminée par une «mesure» et assujettie à un ordre. Voilà le sens nouveau de *ῥυθμός*: la «disposition» (sens propre du mot) est chez Platon constituée par une séquence ordonnée de mouvements lents et rapides, de même que l'«harmonie» résulte de l'alternance de l'aigu et du grave. Et c'est l'ordre dans le mouvement, le procès entier de l'arrangement harmonieux des attitudes corporelles combiné avec un mètre qui s'appelle désormais *ῥυθμός*. On pourra alors parler du «rythme» d'une danse, d'une démarche, d'un chant, d'une diction, d'un travail, de tout ce qui suppose une activité continue décomposée par le mètre en temps alternés. La notion de rythme est fixée. A partir du *ῥυθμός*, configuration spatiale définie par l'arrangement et la proportion distinctifs des éléments, on atteint le «rythme», configuration des mouvements ordonnés dans la durée.

### 3.3 Rhythm as sense of discourse (Meschonnic 1982)

Following Benveniste's philological proposal of rhythm as form (1966), Meschonnic (1982) extended the original meaning of the rhythm to discourse. The key elements of Meschonnic's theory of rhythm are the relationship among rhythm, sense, subject and voice. The author criticizes the structuralist approach to rhythm, where it is considered a formal element, juxtaposed to the sense, just like the signifier is juxtaposed to the signified; similarly, grammar, lexicon, syntax and morphology are conceived as separate levels. Meschonnic recognizes that his theoretical approach to rhythm was possible thanks to the philological proposal by Benveniste, who first linked rhythm to sense:

A partir de Benveniste, le rythme peut ne plus être une sous-catégorie de la forme. C'est une organisation (disposition, configuration) d'un ensemble. Si le rythme est dans le langage, dans un discours, il est une organisation (disposition, configuration) du discours. Et comme le discours n'est pas séparable de son sens, le rythme est inséparable du sens de ce discours. Le rythme est organisation du sens dans le discours. S'il est une organisation du sens, il n'est plus un niveau distinct, juxtaposé. Le sens se fait dans et par tous les éléments du discours. La hiérarchie du signifié n'en est plus qu'une variable, selon les discours, les situations. Le rythme dans un discours peut avoir plus de sens que le sens des mots, ou un autre sens. Le «suprasegmental» de l'intonation, jadis exclu du sens par des linguistes, peut avoir tout le sens, plus que les mots. Ce n'est pas seulement la hiérarchie du signifié qui est ébranlée,

mais les «subdivisions traditionnelles», comme disait Saussure: syntaxe, lexicale... Le sens n'est plus le signifié. Il n'y a plus de signifié. Il n'y a que des signifiants, participes présents du verbe signifier (Meschonnic 1982: 70).

Thus, the rhythm organizes the sense of a discourse; and in a discourse it can have more sense than words; intonation can be more significative than words.

The rhythm is the main conveyor of sense, mainly because it represents the subject, and precedes the subject, since it is considered our prehistory:

Priorité sur la pensée, antériorité chronologique du mètre, le rythme est aussi, d'abord une antériorité anthropologique, une préhistoire en nous. L'archaïque comme une mémoire de l'oubli, non un passé mais une permanence, une fois de plus l'origine comme un fonctionnement. Leroi-Gourhan [1965: 220] note que «Les marques rythmiques sont antérieures aux figures explicites», et «L'art primitif débute par conséquent dans l'abstrait et même dans le préfiguratif». [...] Les rythmes sont la part la plus archaïque dans le langage. Ils sont dans le discours un mode linguistique pré-individuel, inconscient comme tout le fonctionnement du langage. Ils sont dans le discours un élément de l'histoire individuelle (Meschonnic 1982: 100).

Before words, before the comprehension of meaning (sense), rhythm represents unintentionality, it is related to the unconscious; therefore it can show in the language the attitudes of the subject. In Meschonnic's theory of rhythm, the subject becomes the protagonist of his/her discourse:

Si le sens est une activité du sujet, si le rythme est une organisation du sens dans le discours, le rythme est nécessairement une organisation ou configuration du sujet dans son discours. Une théorie du rythme dans le discours est donc une théorie du sujet dans le langage. Il ne peut pas y avoir de théorie du rythme sans théorie du sujet, pas de théorie du sujet sans théorie du rythme. Le langage est un élément du sujet, l'élément le plus subjectif, dont le plus subjectif à son tour est le rythme (Meschonnic 1982: 71).

The author distances oneself from the theory of sign, according to which language preexists the discourse, and discourse is but a special use of signs, a choice in the preexisting system of signs. Meschonnic reduces the dualism between language and discourse to discourse only, and maintains that the work includes the language. Consequently, the speaking subject, according to the theory of sign, operates the choice of signs, giving rise to discourse. Conversely, in the theory of rhythm, the discourse is not conceived as the use of signs by the subject, but as the activity of the subject in history; the historical and cultural subject informs the discourse with his/her rhythm and,

through the rhythm, with the sense, which becomes inseparable from the subject:

Dans la théorie du rythme que Benveniste a rendue possible, le discours n'est pas l'emploi des signes, mais l'activité des sujets dans et contre une histoire, une culture, une langue, — qui n'est jamais que discours, où la définition de la langue apparaît essentiellement grammaticale, un certain rapport du syntagmatique au paradigmatique, qui reprend, *redécoupe* les catégories anciennes. Le rythme comme organisation du discours, donc du sens, remet au premier plan l'évidence empirique qu'il n'y a de sens que par et pour des sujets. Que le sens est dans le discours, non dans la langue (Meschonnic 1982: 71).

The author extends to politics the relevance of the subject in the discourse; in fact, he proposes that semiotics tends to the totalitarianism, because it is inclined to unify and comprehend everything in signs: “Si tout est signe et systèmes de signes, tout est sémiotisable” (ibidem: 74). Conversely, the theory of rhythm is opened to the subject, and to any corporal aspects that can concern language; it is opened to the indefinite, to a fragmentary reality.

In the theory of rhythm, the subject is both historical and ahistorical at the same time: ahistorical because any language presupposes the presence of a subject, who is not historically determined as the individual; historical because s/he is subject of a discourse, thus socially and individually linked to time and history (94-86-87-88).

The realization of the presence of subject in discourse through rhythm is not only limited to the spoken language, but also in written language.

Once defined the relevance of sense and subject in the theory of rhythm, Meschonnic (1982: 216-217) can give his definition of rhythm:

Je définis le rythme dans le langage comme l'organisation des marques par lesquelles les signifiants, linguistiques et extralinguistiques (dans le cas de la communication orale surtout) produisent une sémantique spécifique, distincte du sens lexical, et que j'appelle la signifiante: c'est-à-dire les valeurs, propres à un discours et à un seul. Ces marques peuvent se situer à tous les «niveaux» du langage : accentuelles, prosodiques, lexicales, syntaxiques. Elles constituent ensemble une paradigmatique et une syntagmatique qui neutralisent précisément la notion de niveau. Contre la réduction courante du «sens» au lexical, la signifiante est de tout le discours, elle est dans chaque consonne, dans chaque voyelle qui, en tant que paradigme et que syntagmatique, dégage des séries. Ainsi les signifiants sont autant syntaxiques que prosodiques. Le «sens» n'est plus dans les mots, lexicalement. Dans son acception restreinte, le rythme est l'accentuel, distinct de la prosodie — organisation vocalique, consonantique. Dans son acception large, celle que j'implique ici le plus souvent, le rythme englobe la prosodie. Et, oralement, l'intonation. Organisant ensemble la signifiante et la signification du discours, le rythme est



l'organisation même du sens dans le discours. Et le sens étant l'activité du sujet de renonciation, le rythme est l'organisation du sujet comme discours dans et par son discours.

The author says that he took the term “significance” from Benveniste (1974: 51; in Meschonnic 1982: 72) but assigned a different value to the term with respect to the original meaning of “propriété de signifier” used by Benveniste in “Sémiologie de la langue” (ibidem):

Pris dans la paradigmatic et la syntagmatic d'un discours, le rythme sens et sujet fait une sémantique généralisée, fonction de l'ensemble des signifiants, qui est la *signifiance*. [...] Le rythme dans le sens, dans le sujet, et le sujet, le sens, dans le rythme font du rythme une configuration de l'énonciation autant que de l'énoncé. C'est pourquoi le rythme est le signifiant majeur. Il englobe, avec l'énoncé, l'infra-notionnel, l'infra-linguistique. *Le rythme n'est pas un signe*. Il montre que le discours n'est pas fait seulement de signes. Que la théorie du langage déborde d'autant la théorie de la communication. Parce que le langage inclut la communication, les signes, mais aussi les actions, les créations, les relations entre les corps, le montré-caché de l'inconscient, tout ce qui n'arrive pas au signe et qui fait que nous allons d'ébauche en ébauche. Il ne peut pas y avoir de sémiotique du rythme. Le rythme fait une antisémiotique. Il montre que le poème n'est pas fait de signes, bien que linguistiquement il ne soit composé que de signes. Le poème passe à travers les signes. C'est pourquoi la critique du rythme est une anti-sémiotique (Meschonnic 1982:72).

### 3.4 The acoustic face of words (Albano Leoni 2009)

Albano Leoni (2009) criticizes the segmental paradigm that has dominated the theory of spoken language from its first traces in western philosophy to the birth of phonology and until today. According to this paradigm, the signifier is a linear and continuous signal that ordinary hearers-speakers or expert analysts segment in order to assign a meaning to parts of it. In this way, the signal becomes discontinuous and not necessary linear; these parts are called by linguists ‘units of first articulation’, or ‘monemes’ or ‘morphemes’. According to linguists, not only the signifier, but also the signified is naturally constituted by a ‘linear succession’ of basic units, segments called ‘phonemes, or ‘units of second articulation’. On this level, though, the criterion of identification of units is not the assignation of meaning (Albano Leoni 2009: 79-80), but each phoneme is analysed as a bunch of traits or double-way matrix of acoustic/articulatory traits (ibidem: 81).

Since the segmental paradigm has dominated linguistics for 2,500 years, the author outlines a history of phonology to find out how linguistics arrived to define the phoneme, and how it developed after its definition.

### **3.4.1 Brief history of the phoneme**

The historical outline is divided into four stages: i) prehistoric stage; ii) psycholinguistic stage; iii) structuralism; iv) generativism, postgenerativism and cognitive stage.

#### *3.4.1.1 Pre-history*

The notion of basic unit of phonic structure of a language traces back to the ancient Greeks' reflection on *grámma* (letter) and *stoicheíon* (element), later, to that of their Latin equivalents *littera* and *elementum* (Laschia 2001; Desbordes 1998; in Albano Leoni 2009: 83).

In Chapter 20 of Aristotle's *Poetics* one can read that 'Stoicheíon is indivisible voice, not an ordinary indivisible one, but a voice that originates a composed voice' (Albano Leoni 2009: 83)

The Latin interpretation of this passage was: "*littera est pars minima vocis articulatae*" ('letter is the basic unit of articulated voice') (ibidem: 84). Therefore, for the Greeks and the Latins the only way to represent the basic unit of language was graphics, and was the graph already provided by the alphabet: letter = sound (ibidem: 84-85).

#### *3.4.1.2 The psycholinguistic stage*

At the end of the XIX century, studies on articulatory and acoustic phonetics revealed that the phonic signal is continuous and variable (Albano Leoni 2009: 85). Moreover, studies in psychology, especially in Germany, focused on the physiology of perception and developed the notion of "representation" of percept, and consequently on the association of representations (ibidem: 86).

Baudouin de Courtenay (1985: 9; in Albano Leoni 2009: 87) operated a synthesis of the findings of these studies by defining the phoneme as the psychological equivalent of the linguistic sound, i.e. a unitary representation taking shape in one's mind through a fusion of impressions originated by the pronunciation of a sound. Nevertheless, this unitary representation of the phoneme is associated to a certain number of single representations, which can be articulatory or acoustic.

However, while de Courtenay only considered the individual language, Paul (1920; in Albano Leoni 2009: 86), also took into account the intersubjective dimension defined as the average of individual usages.

In the psycholinguistic stage, the basic unit (*pars minima*), was no longer illusorily represented by the letter (*littera*), but it was projected in the immaterial psychic representation of percept. Nevertheless, the materiality of sound was still present, and the problem of its limits, the need for segmentation, required the existence of a *pars minima*.

### 3.4.1.3 Structuralism

Structuralism is divided by the author into European and American (Albano Leoni 2009: 89-109). In Europe, the beginning of this stage was marked by the publication in 1929 of the Theses of Prague (*Travaux du Cercle Linguistique de Prague*), while its ending coincided with the publication of *Grundzüge der Phonologie* by Trubeckoj (1939/1971). During the ten years between the publication of these two works, there were numerous conferences and publications related to phonology and phonetics, among them the four international conferences titled *Travaux du Cercle Linguistique de Prague I-IV*; in this debate among linguists, the notions of ‘structure’, ‘opposition’, ‘relevance’ and ‘function’ emerged (cf. Albano Leoni 2009: 90).

In the *Theses*, the phoneme was defined as ‘the most simple and significative acoustic-motorial images in a given language’, these ‘images’ were ‘subjective’ and corresponded to the ‘objective acoustic-motorial factors’ that had only an ‘indirect relationship with linguistics’, therefore they ‘could not be identified with linguistic values’; the subjective acoustic-motorial images were considered ‘elements of a linguistic system only to the extent they developed a differentiating function of meanings (*significations*) in the system’; the ‘sensorial content of such phonological elements’ was considered ‘less essential of their mutual relationships in the system (structural principle of the phonological system)’ (Garroni e Pautasso 1979: 25; in Albano Leoni 2009: 91).

According to Albano Leoni’s proposal (2009: 93-94), Bühler took part to the debate, perhaps at the beginning he was attracted by the phonological distinctive feature (Merkmal) that allowed him to extend his notion of distinctive relevance to the single sounds. However, Bühler’s phoneme, differently from Trubeckoj’s, is a material form (the sound matter had a phonematic impress, mainly in words) with little diacritic autonomy, because

the diacritic role of phoneme (or its features) was to be used only in case of non-recognition of words through the *Gestalt* perception (ibidem: 94).

In the next stages of the debate on phoneme, Bühler's suggestion was not followed by phonology, which stressed the importance of the distinctive sign (*Signalement*) and left the physiognomic idea behind, as well as the representation of the signifier (ibidem). At the end of the decade, the *Grundzüge der Phonologie* (Trubeckoj 1939/1971; in Albano Leoni 2009: 100) hardly dealt with the nature of the phoneme, since the phoneme was no longer considered as the psychic representation of a physical event or an phonatory intention, then individual, but an indefinite element with a distinctive function (cf. Albano Leoni 2009: 100-101):

Chiamiamo fonemi le unità fonologiche che, dal punto di vista di una data lingua, non si possono divider in unità fonologiche minori susseguentesi. Quindi il fonema è la più piccola unità fonologica di una lingua. Il lato significativo di ogni parola nella lingua si può dividere in fonemi, si può rappresentare con una determinata serie di fonemi (Trubeckoj 1939/1971; in Albano Leoni 2009: 100).

A common ground between the European and the American structuralism was the ‘*a priori* assumption that a *pars minima* called phoneme existed’ (Albano Leoni 2009: 103). The beginning of the American structuralism coincided with the publication of the work *Language* by Sapir (1925; in Albano Leoni 2009: 103). Sapir considered the phonic units as variable and working in a system; but he also evoked the ‘mental dimension’ through: i) the notion of ‘psychological distinction’ of each unit (Sapir 1925: 39; in Albano Leoni 2009: 103); ii) and the argument according to which ‘the data perceived by senses acquire significance only in a form intuitively sensed’ (ibidem: 51; in ibidem).

Bloomfield recognized the infinite variety of phonic phenomena and that speech is continuous (1933: 77; in Albano Leoni 2009: 104); according to him, in the speech it is possible to recognize the similarity of two utterances thanks to the “gross acoustic features”, in any utterance:

“certain ones are distinctive, recurring in recognizable and relatively constant shape in successive utterances. These distinctive features occur in lumps or bundles, each one of which we call a phoneme” (Bloomfield 1933: 79; in Albano Leoni 2009: 104).

Therefore, the most important aspect of phonemes is their distinctiveness as a result of their distribution in utterances; they “are not

sounds, but features of sounds which the speakers have been trained to produce and recognize” in speech (Bloomfield 1933: 79; in Albano Leoni 2009: 104). According to Bloomfield, phonology cannot deal with the study of meaning, but ‘only assume that in a language some utterances are similar in form and meaning’ (Albano Leoni 2009: 104).

Following Bloomsfield, Bloch in his postulates (1948: 143-144; in Albano Leoni 2009: 105-106) also appeared distrustful of the relevance of psychology, philosophy and semantics to phonology. Bloch also maintained that, despite instrumental measurements cannot demonstrate the segmentation, this takes place in the mind of a people trained in phonetics (ibidem: 12; in ibidem), forgetting that perception is subjective and instrumental data are objective (Albano Leoni 2009: 106).

Hockett (1955: 25; in Albano Leoni 2009: 106-107) also maintained that it is not the “phoneme” (“or other phonologic unit”) that is “physically present in the speech signal”, but rather “a *representation* of the phoneme”. This appears to be a subversion of the notion of representation, since in Hockett the sound (material) is a representation of the phoneme (immaterial), and not the phoneme being a representation of material (inductively built from the observation of the datum) (Albano Leoni 2009: 107).

Jakobson and Halle, in *Fundamentals of language* (1956; in Albano Leoni 2009: 107-108), unified the American and European structuralism by synthesizing in binary matrices both Bloomsfield’s features and Trubeckoj’s components of the phonological content, creating a model of representation that is still used. Nevertheless, if the features are not phenomena of phonic material that can be perceived or measured, then the phoneme is still immaterial (Albano Leoni 2009: 107-108). This ambiguity, together with the undefined limits (material, mental or ontological) of the segment, were not solved by the structuralism (ibidem 109).

#### 3.4.1.4 *The generativist and post-generativist stage*

The work *The Sound Pattern of English* (Chomsky and Halle 1968; in Albano Leoni 2009: 110) marked the beginning of this stage. In it, the relationship between signifier and signified, still present in the structuralism through the odd couples, completely disappears; the *parole* is not considered the subject matter of linguistics, and language is not an intersubjective phenomenon, since it is only considered as grammar, i.e. the ideal speaker’s competence (Albano Leoni 2009: 110). Nevertheless, the notion of

representation is still present: firstly because the grammar is “the system of rules represented in the mind of the speaker-hearer” (Chomsky and Halle 1968: 4; in Albano Leoni 2009: 110); secondly because ‘the phonetic representation of an utterance in a given language is a matrix where rows are labelled as features of universal phonetics’ (‘the columns indicate the sequence of segments of the utterance’) (Chomsky and Halle 1968: 5; in Albano Leoni 2009: 110-111). The matrix appears to be a ‘metalinguistic representation’, i.e. ‘how a model represents a phenomenon’ (Albano Leoni 2009: 111). This means that the ‘object of the representation’, or the ‘mental process’ is not the “percept”, or the ‘physical datum’ of the psychologist stage; hence, the generative model did not investigate the ‘nature’ of either the *pars minima* or its representation, which remains discrete and segmental – it did not question the structure of speech and its physical aspect (Albano Leoni 2009, 111-112). The same is true for the following generative phonologies, for example the so called natural phonology (Stampe 1979; Donegan and Stampe 1979; Hurch and Rhodes; in Albano Leoni 2009: 113-114), or the so called articulatory phonology (Browman and Goldstein 1990; in Albano Leoni 2009: 114-115), where ‘the discrete, abstract, dynamically defined gestures’ (Browman and Goldstein 1990: 342; in Albano Leoni 2009: 114-115) are but a new rendering of the features, this time from the production perspective, just as Jakobson’s or Chomsky and Halle’s two-dimensional matrixes (Albano Leoni 2009: 115). In the so called auto-segmental and metrical phonology (Goldsmith 1990; in Albano Leoni 2009: 115-116) the ‘axiom of the linear succession of discrete elements (segments) still persists; however, it is integrated with a representation on different levels interrelated’ (Albano Leoni 2009: 116). The optimality theory (Prince and Smolensky 1993; McCarty 2002; in Albano Leoni 2009: 118) appears to be one of the best ‘attempts of simplification of the generative model’ (Albano Leoni 2009: 118). In fact, the theory is based on an “input” of ‘linguistic primitives’ (allegedly innate and universal) and their combinations, a ‘transformational’ mechanism that filters the primitives through two devices, and an “output” (McCarty 2002: 70, 10; in Albano Leoni 2009: 18). Two devices operate the transformation: “Gen” generates a series of possible candidates to the output, “Eval” selects them. The mechanism is regulated by two principles: ‘fidelity’ (output must not modify the input) and ‘restraints’, i.e. natural limitations concerning both the combinations of primitives (‘universal constraints’) and the transformations from input to output (‘special constraints’). The optimality theory’s approach is segmental,

since the phonological units both of input and output are segments, and the theory does not say anything about their nature (Albano Leoni 2009: 18).

#### 3.4.1.5 *Cognitive phonologies*

This stage of the brief history of the phoneme outlined by Albano Leoni (2009: 82-127) is due to the publication of works specially dedicated to cognitive phonologies in the framework of cognitive linguistics. Cognitive phonologies differ from generative phonologies because the subject of study is no longer the ‘mind of the ideal speaker’, or ‘the transformation of a deep structure’; but rather the ‘biology and psychophysics of speakers’. If the focus of cognitive linguistics is the ‘construction and the representation of knowledge through languages’, then ‘studying how phonic stimuli contribute to that construction and representation’ should be the most important thing (Albano Leoni 2009: 119). Is this really the case in cognitive phonologies?

Lakoff (1987: 61-62; in Albano Leoni 2009: 119) considers the phoneme ‘the basic category of phonology and consequently the supposed content of cognition’ (Albano Leoni 2009: 119). Langacker (1987: 388-394; in Albano Leoni 2009: 119) ‘recognizes the primacy of syllable in the orthogenesis of language’, but the ‘basic unit of the representation of the signifier’ is the ‘segment’ (Albano Leoni 2009: 119). In Durand and Laks (2002; in Albano Leoni 2009: 121), the ‘signify is completely absent’, there is no ‘reference to the context and the world’ (Albano Leoni 2009: 121). Pinker and Jackendoff (2005; in Albano Leoni 2009: 121) also make no reference to the contribution of the ‘hypotheses of meaning’ and the ‘context’ to perception; but they describe how human auditory system works, underlining its ‘ability to discrimination’. Durand and Laks (2002: 1; in Albano Leoni 2009: 121) seem to confirm the suspicion that cognitive phonology ‘in some cases’ appears to be ‘a branch of neurosciences that studies neuro-motor controls for the production and the reception’.

There is no further development on the ‘relation between percept and representation’ (Albano Leoni 2009: 122). Myers (2000: 245; in Albano Leoni 2009: 122) maintains that “a phonological pattern is the distribution of a psychological sound category. Phonology is the theory of such patterns”. Nathan (1996: 107, 109; in Albano Leoni 2009: 122) defines the “nature of the phoneme as mental unit”, and “features are [...] the mental images of the physical parameters that constitute speech”. There may be an unconscious comeback to the associative psychology in the connection between sound and

representation, but the above mentioned definitions are not so far from Baudouin de Courtenay's 'mental images of physical parameters' and Trubeckoj's features as 'acoustic-motor elements of a phonic representation' – the object of representation are still the basic units, mainly the phonemes, sometimes the syllables, as in the case by Segui and Ferrand (2002; in Albano Leoni 2009: 123).

The so called laboratory phonology may be included in the category of cognitive phonologies because it considers language as a 'cognitive system' that has to be 'explained' as a natural ('physical') phenomenon of human beings as a result of their 'evolution', and their 'interaction with the environment', the 'social interaction' being 'subsumed within natural phenomena' (Pierrehumbert, Beckam and Ladd 2000: 274-275; in Albano Leoni 2009: 123-124). Laboratory phonology follows the 'United States tradition of mathematization of linguistics, formalization, and measurability of phenomena' (Albano Leoni 2009: 125). It is implicit in this line of research, as in other neurosciences, the possibility to describe human beings 'objectively', not considering the 'subjectivity' (Oliverio 1999: 3-12; in Albano Leoni 2009: 125). This approach guarantees a particular attention to the sound matter and to the continuous nature of the signal (Albano Leoni 2009: 125). The theoretical paradigm is segmental and discrete, but the perspective in this case is overturned: since the phonological module is categorical and non-discrete, also the phonetic module is rendered non-discrete, despite the ascertained non-categorical nature of the signal. In this way also the phonological categories become natural, because the physical (articulatory and acoustic) non-linearity allow to consider phonetics as nearly categorical (Pierrehumbert, Beckam and Ladd 2000: 284-285; in Albano Leoni 2009: 125).

In these approaches the phoneme is mentioned occasionally and is never defined (Albano Leoni 2009: 126).

### **3.4.2 Conclusion**

At the end of his review on phonology, Albano Leoni maintains that the question on the nature of the phoneme, both physical and representational, is still open but no longer considered (2009: 126). The distinctive traits (with the relative binary matrixes) have acquired 'phenomenological dignity'; however, they cannot represent 'real phonological processes' (ibidem: 127). The only thing that has remained constant is the phoneme identification with the basic unit (*pars minima*): a discrete nature but, paradoxically, undefined in its limits.



The author hypothesizes that the basic units of segmental phonologies is a ‘slippery thing, maybe inexistent’ and tests the hypothesis through an analysis of phonic matter and its perception (*ibidem*: 128).

### 3.4.3 Criticism of segmental phonologies

As to the substance, or the phonic matter, the signal is not segmental, because the only discontinuities that can be observed are those of silent pauses and the silence at the end of occlusive sounds (not always present) (Albano Leoni 2009: 128). Therefore, the division of words in single parts is only due to an artificial representational requirement of a phonetic transcription (Paul 1920: 34-36; in Albano Leoni 2009: 129). The reason of the impossibility of segmentation are mainly due to the co-articulation phenomenon, characterized by a ‘parallel cooperation of different articulatory organs (larynx, tongue, soft palate and lips) whose individual activities have different durations’, with non-synchronized movements, and ‘subjected to the inertia and the non-accurate neuromotorial activity that control them’ (Albano Leoni 2009: 129-130). Only salient points can be defined in a real sequence of speech sounds (*ibidem*: 130). Results from studies on coarticulation (for example, Hawkins and Smith 2001; in Albano Leoni 2009: 130) show that the ‘information unit’ can be ‘smaller or bigger than a phoneme, sometimes extended to syllable or beyond’. The lack of a defined number of segmental units in the inventory of a given language supports this thesis (Albano Leoni 2009: 130-132). Diphthongs consist of ‘one articulation’, a ‘dynamic vocalic articulation’ characterized by ‘a more or less marked change in trajectory’ and a spectral modification (Laver 1994:284; Levelt 1989: 432; in Albano Leoni 2009: 133). On the perceptual side, the ‘phonological interpretation’ of the diphthong is debated, since there are both ‘monophonemic’ (Roca and Johnson 1999: 190-203) and ‘biphonemic’ interpretations (in Albano Leoni 2009: 133).

Another counterevidence to the segmental phonology proposed by the author is represented by the variability and the transiency of the signal. To this respect, results from spoken corpora analyses (e.g. Calamai 2002a-b; Landi and Savy 1996; Marotta 2001, 2005; Vallone, Caniparoli and Savy 2002; Sorianello and Calamai 2005), have shown the existence of phones cancelled or realized in different ways from those expected by hearers (in Albano Leoni 2009: 133-134). An analysis on Italian spoken corpus (Albano Leoni and Clemente 2005) revealed that, out of 603 occurrences of the phoneme /p/,

the 38% were considered non-canonical by the tagger ('cancellations', 'anomalous realizations', 'non-recognizable' as [p] but as [b] or [pp]); in the case of vowels, the canonical realizations were 44%; and 48% in the case of diphthongs (in Albano Leoni 2009: 134-135). As to the variability, the 'phonic stability of phonemes' may also depend on prosody, since 'notable phenomena of vowel and consonant reduction' tend to occur in 'zone of prosodic depression of the phonetic string, often in coincidence with accelerations of the speech' (ibidem: 135-136). These cases of non-conforming phones, due to the transiency and imprecision of the signal, seem to be consistent with the language as a system, especially if one considers that some "context effects may be inherent part of the signal itself, reflecting important properties of the speaker and the linguistic content of the message" (Pisoni 1986: 156; in Albano Leoni 2009: 136). In conclusion, it seems that the sound matter, or the phonic substance, does not support the 'primitive nature of the phoneme as a discrete *pars minima*' (Albano Leoni 2009: 137).

The same can be said, from the point of view of perception; in this case the context proved the incidence of the context in the comprehension process. Results of shadowing experiments (repetition of a text by a subject, after the listening) showed that progressive decontextualization of a word caused a progressive reduction of its comprehensibility by the subject; this happened either when the text was unknown by the subject, or it was manipulated by inserting silent pauses or changing the word order (Albano Leoni e Cutugno 1999), or the visual component of a multimodal message was suppressed (Cerrato e Albano Leoni 1998; Cerrato, Falcone e Albano Leoni 1998; in Albano Leoni 2009: 137, 140). In addition, the 'historians and theorists of writing and the psychologists' reviewed by the author (cf. Albano Leoni 2009: 141-147) 'seem not to confirm the hypothesis of a primitive nature of the phoneme', and writing seems to be a 'metalinguistic representation of speech' and not its natural 'projection' (ibidem: 148).

In conclusion, the author proposes that comprehension of speech seems the result of the processing of the acoustic and articulatory conformation of the sound matter, and its context and co-text, where prosody plays a fundamental role; hence, the comprehension appears to be more an 'interpretation' than a 'decoding' process (Albano Leoni 2009: 148-154).

### 3.4.4 The proposal for the acoustic face of words

The starting point is the original meaning of structure in linguistics, i.e. the conception of the whole not as the sum or the juxtaposition of its components, but as the relationship among the parts that determines the whole, in such a way that any single part does not preexist to the whole, nor the whole preexist to the parts (Albano Leoni 2009: 155). When the term *structure* was introduced in linguistics, it was related to *Gestalt*, key word of psychology of form born at the end of the 19<sup>th</sup> century. Even if *Gestalt* and *structure* were different, they shared the concept of the relationship between the whole and its parts described above; moreover, this relationship was ‘always defined *in praesentia*, i.e. in terms of ‘syntagmatic structures or *Gestalten*’, just like the structures of ‘a crystal, a molecule, a sentence or a broader text, a construction, an animal or vegetal body, a mental pattern giving shape to a perception’ (ibidem: 156-157). However, observes the Albano Leoni, in a ‘phonological, morphological and semantic structure’ the parts are not physically compresent, like in an atom, a perceptive pattern, a sign, or in a sentence’, but the relationship in this case is in *absentia*’, i.e. ‘paradigmatic’ (ibidem: 157). Thus, Albano Leoni (ibidem) proposes that ‘the peculiar nature of phonological units may have allowed or encouraged a mechanistic and reductive interpretation of the phonological structure’, more similar to that of mathematic abstract structures than to that of physical structures of psychological and social sciences (Piaget 1968/1994: 165; in Albano Leoni 2009: 157-158). Therefore, the author proposes that the structural relationship between the whole and its parts, in linguistics, takes place in the *langue* (or competence) (Albano Leoni 2009: 158-159). On the contrary, the syntagmatic relationship between the whole and its parts, or the relationship between the speaker/hearer and the structure may take place ‘at a local level’, in such a way that ‘the use or comprehension of a word’ is a continuous process of ‘solution finding’, through the ‘network of systemic correlations among the coexisting forms’ (De Mauro 2007: 27; in Albano Leoni 160-161). In simpler words: ‘the relationship whole-parts’ should be considered as a ‘local relationship within a sign or a group of contiguous signs’ (Albano Leoni 2009: 163). Both this mental process, and the ‘neural state’ of the ‘hearer’, ‘associated to a situation’ (Kravchenko 2008: 185-189; cf. Maturana and Varela 1980/2001; in Albano Leoni 2009: 161, 162-163), ‘manifest themselves through phonic aspects’ (Albano Leoni 2009: 162-163).

The physiognomic perception and representation of speech is then proposed by Albano Leoni (2009: 165-192) as an alternative to the segmental phonology, that he criticizes, among other things, for having focused too much on the *langue*, the signifier and the speaker, in detriment of the *parole*, the signified and the hearer (cf. *ibidem*: 17-23). The physiognomic (and holistic) approach adopted by the author is the same one ‘followed by the Gestalt psychology at the beginning of the XX century (Köhler 1947)’, according to which ‘the meaning is expressed directly in the totality of the perceived corporeal organization’ (Fortuna 2005: 9; in Albano Leoni 2009: 166):

La capacità di distinguere, memorizzare e riconoscere un numero molto grande di volti, di fisionomie, di ambienti, di luoghi è nota. Riconosciamo, e distinguiamo dalle altre, la faccia di una persona conosciuta anche se si fa crescere i baffi o si toglie gli occhiali o ingrassa o dimagrisce o si tinge i capelli o invecchia o ride o piange. Distinguiamo due facce diverse anche se ci sono ignote, e il loro successivo riconoscimento sarà tanto più facile quanti più saranno gli elementi contestuali di cui disporremo (ad esempio se ne udremo anche la voce, o se l’ambiente e la situazione dell’incontro saranno gli stessi). Il riconoscimento di una faccia conosciuta, o la distinzione tra due facce (note o ignote, o delle quali una sia nota e l’altra ignota) partono dalla percezione visiva globale e per concludersi non hanno bisogno di passare attraverso la scomposizione del tutto nei suoi presunti tratti (ad esempio misurando la lunghezza del naso o la distanza tra gli occhi e così via), neanche quando dobbiamo distinguere i volti di due gemelli omozigoti (Albano Leoni 2009: 166).

The author maintains that the notion of acoustic physiognomy allows to consider both the *Gestalt* and the notion of a structure *in praesentia*, or local (see above – this section) (*ibidem*: 167). As to the application of the *Gestalt* to language:

[...] bien souvent nous appréhendons les ensembles *avant* même de discerner leurs parties – si tant est que nous les discernions jamais. Il se trouve en effet que nous lisons très bien, sans pour autant individualiser chaque mot, *a fortiori* toutes les lettres; nous écoutons aussi des mélodies, sans pour autant en détacher chaque note; et nous reconnaissons bien tel ou tel regard, ironique ou engageant, sans pour autant noter la couleur des yeux, que nous serions bien incapables d’évoquer. [...] Nous affirmerons ensuite que les *formes*, c’est-à-dire les unités organisant les champs perceptifs, ne sont pas moins immédiatement données que leurs parties (paraphrase of Wertheimer by Rosenthal and Visetti 2003: 65-66; in Albano Leoni 2009: 166-167).

The ‘hypothesis of a physiognomy of language’ was already proposed by Wittgenstein (in Fortuna 2005: 7; in Albano Leoni 2009: 169). However, Albano Leoni’s idea of a sound face of words was taken from Bühler

(1934/1983: 328-346; in Albano Leoni 2009: 169-175), who proposed a ‘perception and recognition of speech is holistic, according to a gestaltic modality, of physiognomic type: the whole is recognized as a representation of a phonic *silhouette*’ (Albano Leoni 2009: 170). In fact, Bühler’s sound face is characterised by ‘material marks’ (named *notae* or *diakritisches Signalement*), similar to Trubleckoj’s, that ‘allow recognition when communication is difficult, there are too many elements to be differentiated, differences are small, or the circumstances are adverse’ (Bühler 1934/1983: 43-44; in Albano Leoni 2009: 170). Bühler specifies that physiognomic recognition may be completed by the analytic recognition; however, the ‘starting point’ is the word, whose physiognomy may change, and ‘the phoneme is *Signalement* of the word’, which is ‘superordinate to the phoneme’ (Albano Leoni 2009: 171). Bühler’s idea of sound face of words is linked to the theory of the symbolic and deictic fields (see above – section 3.1.6); he states that ‘linguistic phenomena cannot be considered as separated from the reality of physical phenomena’ (Bühler 1934/1983: 88-89; in Albano Leoni 2009: 172). Albano Leoni (2009: 173) maintains that in Bühler the human language developed starting from ‘indication’ and ‘imitation’ (deixis) and reaching ‘representation’ (symbolism); nevertheless, the final stage is still linked to the real situation of the communicative act, preserving an ‘indirect relationship’ with indication. In Albano Leoni’s words (2009: 173):

Il messaggio (fonico) è innestato nella realtà, in quella fisica del mondo e in quella psicofisica dei parlanti, e questa realtà è parte integrante della situazione comunicativa ed espressiva nella quale si colloca il messaggio.

The coexistence or ‘integration’ of symbolic and deictic fields may explain why the signal, in its ‘segmental representation’, is the only vehicle of sense; the ‘semiotic function’ also explains the economy of human language, and allows to overcome the notions of double articulation and the distinction between linguistic and paralinguistic factors (Albano Leoni 2009: 173-175).

Prosody also is processed through in gestaltic way, since the perception of melodic form remains the same when it is transposed in a different range (modification of the constitutive tones) (Persyn-Vialard 2005: 138-139; in Albano Leoni 2009: 177). According to Albano Leoni (2009: 179) ‘the linguistic signal, which is continuous, consists of a sequence of periodicities and aperiodicities’, rising and falling alternate movements, which are the result of the dynamicity of the physical magnitudes that constitute the signal’. These

features render the signal ‘naturally segmentable’ in basic units identifiable in the syllables (*ibidem*). The syllable is therefore conceived as an ‘articulatory and perceptive salience’, where salience or prominence is the ‘gradual compression and strengthening and following rarefaction of acoustic indices’ (*ibidem*). The author proposes syllable as the ‘physical (and not linguistic or semiotic) *pars minima*’; it is different from the phoneme because it presents ‘a natural articulation, perceptibility, accessibility and universal constituting mechanisms’ (*ibidem*: 180-181). Conversely, as the phoneme, it is not discrete; however, this should not be a problem, since the definition of its limits is not relevant to speakers, because the important aspect is the perceptive salience of its central part (nucleus) (*ibidem*: 181). Again, the physiognomic approach is relevant to this proposal, since ‘a face is a group of individual prominent aspects that are not discrete; a sound face is a group of perceptive prominences hold together by a sense’ (*ibidem*: 181).

In conclusion, ‘voice, syllable and prosody’ are the ‘properties sound face’, and are ‘coessential to it’; the ‘sense’, that holds together these properties, permitting the recognition in case of ‘physical ambiguity’ is another property; each component manifests itself with the others, and the whole is more than the sum of its parts (*Gestalt*) (*ibidem*: 184). A single word or syllable (e.g. hesitation, vocalization...) are emitted by a voice and have a prosody, and consequently can be significative (*ibidem*: 183-184). Nevertheless, a complete understanding is impossible without considering the sense, which is related to a context (*ibidem*: 185-186), which includes the relationship among the interlocutors, their (linguistic) knowledge of the world, and their systems of expectations. At the end of his argumentation Albano Leoni (2009: 191) states:

Le argomentazioni che ho presentato sono un *plaidoyer* a favore della linguistica della *parole*, o forse, meglio, di una linguistica del discorso, che sia semantica e non semiotica, che sia interpretativa, che sia una linguistica della significazione.

This conclusion seems similar to Meschonnic’s proposal of rhythm as form of discourse (see above – section 3.3), especially considering that Meschonnic’s rhythm is made by marks that are perceived as a unique form, even if he does not make any reference to the *Gestalt*; however, both Meschonnic and Albano Leoni clearly defend a semantic approach to signification and refer to the indeterminateness of sense.

### 3.5 Voice, audio-vision, television

The methodology used to build a questionnaire to elicit a holistic perception of quality of a television broadcast simultaneous interpretation has to take into consideration the influence that the medium may have on perception. Television interpreting studies (e.g. Mack 1999 and Straniero Sergio 2007: 54) underline the importance of voice and the so-called formal aspects in this mode of interpretation. In addition, results from the first survey on quality expectations in television interpreting (Kurz and Pöchhacker 1995) showed the respondent's sensitiveness to criteria like voice, accent and fluency. Results from the survey by Russo (2005) also showed the importance of voice in simultaneous interpreting of films; in fact, "voice quality" was the criterion mostly appreciated, followed by "overall quality", "style" and "fluency of delivery" (Russo 2005: 12). Hence, the need to investigate the relationship between voice and audio-vision in speech perception.

#### 3.5.1 The perception of sound on screen (Chion 1990/1994)

Audiovision is not the mere addition of two sensorial perceptions, "audio" plus "vision", it is the effect of a unique "transsensorial perception", which can be referred to as "rhythm", since in film vocabulary the term is not specifically related to *audio* or *vision* (Chion 1990/1994: 136). The "transsensorial perception" is mainly due to the fact that everything that is "spatial", concerning both image and sound, and can be referred to the "visual impression", while the "auditory impression" is the result of the processing of everything that is temporal, again in terms of both image and sound (see below) (ibidem). In other words, when a rhythmic phenomenon reaches us via a given sensory path; such path, either eye or ear, is perhaps nothing more than the channel through which *rhythm* reaches us. Once it has entered the ear or eye, the phenomenon strikes us in some region of the brain connected to the motor functions, and it is solely at this level that it is decoded as rhythm (cf. ibidem: 136). Auditory perception (processing) does not occur at the very instant that sound is heard (recognised). Contrary to how it may seem, "we don't hear sounds, in the sense of recognizing them, until shortly after we have perceived them". This "paradox" is due to the "ear's temporal threshold" (ibidem: 12). Hearing does not "occur in continuity":

The ear in fact listens in brief slices, and what it perceives and remembers *already* consists in short syntheses of two or three seconds of the sound as it evolves. However, within these two or three seconds, which are perceived as a

gestalt, the ear, or rather the ear-brain system, has minutely and seriously done its investigation such that its overall report of the event, delivered periodically, is crammed with the precise and specific data that have been gathered (Chion 1990/1994: 12).

In audiovision, sound, voice and language constitute an “added value” for each other, because they structure the image; their contribution to the sense of the image is fundamental (ibidem: 5-7). The clear impression, from a present or remembered experience, one has of an image is created by “the expressive and informative value with which a sound enriches a given image”, to such an extent that the information seems to be generated by the image, and finally belongs to it (cf. ibidem: 5). This is so true in cinema that “sound cinema” is primarily “vococentric” and “verbocentric”, considering the importance it gives to the voice with respect to other sounds, most of the times (ibidem)<sup>5</sup>. The notion of added value also works the other way round, i.e. the image constitutes an added value for sound. The phenomenon of added value is particularly evident in sound-image synchronism, due to the principle of “synchresis” (ibidem: 63). “Synchresis” is a word composed by the words “synchronism” and “synthesis”: it designates “the spontaneous and irresistible weld produced between a particular auditory phenomenon and visual phenomenon when they occur at the same time” (ibidem). Synchresis originates from a “point of synchronization”, i.e. a moment when an auditory phenomenon and a visual one occur at the same time (ibidem: 58). It is also “a function of meaning, and is organised according to gestaltist laws and contextual determinations” (ibidem: 63). Synchresis is the result of “dubbing, post-synchronization and sound-effects mixing” (ibidem).

If sound plays a fundamental role in “framing” vision, then the “added value” influences the perception of time in the image. The sound “temporalizes images” especially when, in television or cinema, the sequence of shots does not follow a temporal succession. In these cases, sound: (i) confers a “temporal animation of the image” by better defining “the perception of time”; (ii) “imposes a sense of succession” by “endowing shots with temporal linearization”; (iii) “*vectorizes* or dramatizes shots” giving the sequence “a goal” and creating “a feeling of imminence and expectation” (ibidem: 13-14).

---

5 However, the introduction of *Dolby* technology, which improves sound definition, gives more importance to noises; therefore, “speech is no longer central to films” – “talking film” turns to “sound film” (Chion 1990: 165-166).



In audio-vision, it is possible to distinguish three modes of listening (ibidem: 25-31): “causal listening”, “semantic listening” and “reduced listening”. In causal listening, the subject naturally tends to look for the source of sound; in semantic listening, s/he uses a code to interpret the message, while in reduced listening, s/he focuses on the single aspects or traits of the sound itself, no matter the source or the code (ibidem: 32). Reduced listening is the kind of listening “encouraged” by the “acousmatic situation” because it focuses attention on “sonic textures, masses and velocities” (ibidem). A sound is “acousmatic” when the listener cannot see its source or “originating cause” (ibidem: 71). Semantic listening, instead, is the kind of listening that prevails in cases of “audiovisual counterpoint”, which occur when an “auditory voice” is “perceived horizontally in tandem with the visual track”, i.e. when the visual track is not linked to the audio track. Audiovisual counterpoint occurs, for example, in television “when the image goes its own way and the commentary goes another” (ibidem: 38-39) (ibidem: 38-39). Nonetheless, in audiovision, mainly in the cinema, the dimension that prevails is not the “horizontal” one, but the “vertical” one. Soundtrack, “the simple end-to-end aggregation of all sounds in a film”, is actually produced separately but is interrelated with the visual track: “each audio element enters into simultaneous vertical relationship with narrative elements contained in the image (characters, actions) and visual elements of texture and setting”. For this reason, “there is no image track and no soundtrack in the cinema, but a *place* of images, plus sounds” (ibidem: 39-40). Just as in film editing (montage), it is possible to “cut” the visual flow into “shots”, but this is not possible with the sound flow (there is no “sound shot” or “unit of sound montage”); so in perception, “the sound segment, especially if it lasts any time at all, does not synthesize into any particular bloc or totality in our perception” (ibidem: 40-43). There is no “auditory shot”. Therefore, “the spectator’s general impression of sonic flow will result not from characteristics of editing and mixing conceived separately but from all the elements combined” (ibidem: 46). However, the “point of reference” of this perception is provided by “visual cuts” (ibidem: 44). As a consequence, if sound “temporalizes the image”, then the “image magnetizes sound in space”. Indeed, in cinema the image corresponds to the “frame”, which is the “container” that “preexists” the images and “can remain after the images disappear”. However, there is “no auditory container for film sounds, nothing analogous to [...] the frame”. For this reason, the “the audiovisual scene” (the “container”) is the “frame”. “Film sound is that which is contained or not

contained *in an image*”, “there is no auditory scene”; therefore “*there is no soundtrack*” (Chion 1994: 66-71). If in the cinema perception passes through the image, in television “sound is always foremost”: “never off-screen, sound is always *there*, in its place, and does not need the image to be identified” – “television is illustrated radio” (Chion 1994: 157)<sup>6</sup>

As to the relationship between sound and image in film, an “offscreen sound” is a sound that is acousmatic (invisible) with respect to what is shown in the shot. Conversely, an “onscreen sound” is a sound whose source is visible, because it is represented in the shot. While a “nondiegetic sound” is a sound whose source not only is invisible, i.e. not shown in the shot, but it does not belong to the story which is being represented, it is an external sound; “voiceover commentary, narration and underscoring” are examples of nondiegetic sound (cf. *ibidem*:73).

Thinking of an “audiologovisual poetics” (*ibidem*: 169-184), it is possible to “distinguish three modes of speech in film”: “theatrical speech”, “textual speech” and “emanation speech” (*ibidem*: 25-31). Theatrical speech is “perceived as dialogue issuing from characters in the action”. Textual speech is that of “voiceover commentaries”. Unlike theatrical speech, this kind of speech dominates the images, because it “has the power to make visible the image that it evokes through sound”; it represents “the pure and original pleasure of transforming the world through language and of ruling over one’s creation by naming it” (*ibidem*).

### **3.6 Voice as gesture: expression and perception (Fónagy 1983)**

Chion’s proposal of “transsensorial” perception of audio-vision (see above – section 3.5.1) is a gestaltic perception of sound and images; in addition, when the sound is “acousmatic”, that is, its source is not showed by the image, then the audio-viewer naturally tends to imagine (figure out) it. With reference to the present study, this means that the physiognomic and pathognomic information (see above – section 3.1.8) conveyed by the simultaneous interpreter’s voice through the audiovisual mode acquire a peculiar significance, since the medium provides images, but not that of the speaker (interpreter). Thus, it may be supposed that this information draw more attention from an audio-viewer than from a hearer. In his experimental studies on psychophonetics, Fónagy analysed the physiognomic and

---

<sup>6</sup> “Sound television is inconceivable, unlike cinema” (Chion 1994: 165).

pathognomic information conveyed by voice, he also proposed how this information is subject to a symbolization.

### 3.6.1 Voice as gesture

What we consider verbal communication today originated from gesture communication, which comprised not only movements of different parts of the body (hands, eyes, etc.), but also vocal sounds; therefore, gesture communication is ‘prelinguistic’. Body movements served to reduce the tension originated by the communicative intention, therefore ‘ex-pression’ means discharge of tension (Fónagy 1983: 148). It is evident that preverbal communication is still present in live speech, especially in prosodic elements of languages, mainly intonation and rhythm (*ibidem*). The evolution of human language from the preverbal to the verbal stage can be observed in children, who before acquiring a language only communicate through intonation and rhythm. These musical elements of language are ‘coded’, integrated in the system of a given language (*langue*); musical elements of human languages still function as expression of emotions; stressing represents expression through sound emission (*ibidem*: 150).

### 3.6.2 Prosody

Intensity or dynamic stress articulates and organises speech. It divides the flow of speech chain into sequences, into rhythmic groups (Fónagy 1983: 107). Stressed syllables are the product of a special effort by phonatory [...] and, especially, of a strong contraction of expiratory muscles, mainly intercostal and abdominal (*ibidem*: 108). On an acoustic level, these movements also result in the lengthening of stressed syllables, a higher melody (tone movement) and a modification of timbre. ‘By perceiving and interpreting the stress, the hearer experiences the articulatory effort produced by the speaker’ (*ibidem*). On the contrary, dynamic stress is a “*mise en relief*” through an expiratory and articulatory effort. As a “linguistic and perceptual category”, dynamic stress does not produce a rise in intensity; therefore, the physiological effort is inversely proportional to the intensity of the sound produced.

Emphatic stress is characterized by a glottal closure before either the stressed syllable or the stressed word. Emphatic speech is marked by an interrelation between overstressing activity (movement of the stress) and the frequency of emphatic pauses, with strong attacks, giving the speech a *staccato*

rhythm (cf. *ibidem*: 110). Strong and frequent emphatic stressing is a feature of angry and hate-filled attitudes (cf. *ibidem*: 111). From a vocal-physiological point of view, hatred is but a retained anger, since in hatred, expiratory effort is counterbalanced by a simultaneous opposed effort of violent contraction of laryngeal muscles resulting in a closure of vocal folds (cf. *ibidem*: 113). Emphatic stress in hatred becomes a physical ex-pression, pure body language that supplements the verbal language (cf. Fónagy 1983: 114).

Voice is at the basis of phonation, it can be interpreted only by considering it for what it really is: a body attitude similar to other activities originated by movement (cf. Fónagy 1983: 116). Vocal folds move as a consequence of the flow of air produced by expiration and the action of sub-glottal pressure (Fónagy 1983); vocal fold movement is a periodic vibration perceived as both a muscular and auditory sensation (cf. Fónagy 1983: 117). A singing voice is perceived as more pleasant because it is characterised by the regularity of the melodic curve (tone level constant) (*ibidem*). Melodiousness of the voice is linked to a high regularity of the melodic curve, which is, in turn, closely linked to the expression of tender emotions (*ibidem*). The perception of melodic voice is similar to that of musical tone, it is more pleasant than the spoken voice since its decoding requires very little effort (cf. *ibidem*: 118).

Intonation can be considered as a space projection of laryngeal mimicry, since melodic curves depend on the closing and vibration of vocal folds. Some glottal muscles can modify the degree of tension of vocal folds or reduce their vibrating mass (cf. Fónagy 1983: 121). These movements are too subtle to be perceived as tactile or sensitive movements; the ear distinguishes changes of tone frequency corresponding to the number of vibrations per second of vocal folds (*ibidem*). When the vibrations of tensed folds accelerate, then the tone rises, when fold relaxes and vibration decelerates, then tone decreases (*ibidem*). Prosodic elements of language, intonation and rhythmic patterns, have a gestural behaviour: they not only signify attitudes, but *are* the same attitudes, they convey *per se* these attitudes (cf. Fónagy 1983: 149).

### **3.6.3 Attitudes in voice**

Fónagy (1983: 122-137) analysed acoustically and physiologically the relationship between emotions and voice; in particular, he observed the vocal behaviour related to ‘tenderness’, ‘anger’, ‘complaint’, ‘coquetry’, ‘irony’, ‘joy’ and ‘aggressiveness’.

Tenderness has a waving melody, recalling slow movements of caresses (cf. Fónagy 1983: 124).

Anger produces a melodic curve irregularly broken in stressed syllables by abrupt swerves of one fourth or one fifth (cf. *ibidem*: 122-123). Nonetheless, this is not the only prosodic configuration of a speech dominated by an aggressive emotion; in fact, in some cases, the melodic curve can have a lower tone, as does argumentative speech in animated debates (see below); and in other cases, when anger is controlled, there are strong and regular stresses, but tone does not change (cf. *ibidem*: 125-127). In these cases, anger is concealed, but it can be perceived all the same through strong contraction of expiratory and glottal muscles and tension of the tongue and facial muscles (*ibidem*). The melodic pattern of complaint (*ibidem*: 130) is characterized by a regular rise in voice of half a tone followed by falls of half a tone. These ‘slides’ of tone occur at the beginning or at the end of a rhythmic group, depending on the language. Physiologically, a groaning speech is marked by regular breaks in inspiration caused by contractions of diaphragm that consequently produce a fresh inspiration.

In coquetry (Fónagy 1983: 131-132), tone rises by a third in the first stressed syllable to immediately drop to the same level as before and stay there until the end of the sentence; here, instead of falling, it rapidly slides upward. Such a rise in voice towards the end of a sentence, with a weak intensity, expresses a tender, loving, caring attitude. The voice is semi-whispered and modifies both timbre and register: chest voice, with velar timbre, disappears in the stressed syllable during the rise in tone, while the register turns into head voice. (*ibidem*: 132).

The vocal expression of an ironic attitude (Fónagy 1983: 135-136) develops in a three-step sequence. In the first stage, chest voice is accompanied by a contraction of laryngeal adductor muscles, a voiced creak, a very low tone, an almost flat melodic curve, a very relaxed labial articulation. The second stage is characterized by a head voice, a reduction of global acoustic intensity, a muffled voice; tone suddenly rises, tongue articulation becomes tense since the tongue moves upwards and onwards, and the pharyngeal muscles are contracted. In the third stage the register returns to the chest voice, the onset is compressed and again weak, tone falls to the starting point to stay there, the tongue moves backwards. Due to these variations in tension, the three stages of ironic vocal expression may represent

the three stages of classic tragedy: *protasis*, *epitasis*, *apodosis* o *catastrophe*. Irony seems to be a *mise-en-scene* of aggressiveness (ibidem).

Expression of joy (ibidem: 124-125) comprises sounds features of both aggressiveness and tenderness: frequent and strong stresses, sudden tone jumps, high speed. Joy is characterized by a head voice (not a chest voice as in anger). Timbre is clear, articulation is more advanced, pharyngeal and laryngeal muscles are relaxed. What distinguishes joy from anger is the irregularity of its stresses. Melodic movement ‘bounces with joy’, as does the intensity of stresses, which is never the same but changes constantly. Joyful movements are irregular, unpredictable – their melody is not organised, but anarchic (ibidem).

Aggressiveness (cf. Fónagy 1983: 151) is marked by the presence of emphatic stress, consonant lengthening, glottal closure; the emergence of a rhythmic structure of the clause that strengthens stresses reduces melodiousness, simplifies melodic patterns, reduces duration of vowels, and introduces frequent and often irregular pauses (ibidem).

Body attitude at the basis of explanation or logical operation is not very different from aggressiveness (cf. Fónagy 1983: 138). Both cases are characterized by strong stresses, reduction of melodic movement, prominence (if not dominance) of metrical structures, the repetition of the same pattern, the brusque tone reflecting muscular tension, frequent and regular pauses (ibidem).

### **3.6.4 Vocal personality**

Features expressing emotional attitudes become extremely frequent in people’s speech, almost permanent; these personal vocal traits are less and less perceived by hearers due to the ‘law of accommodation’ (Fónagy 1983: 155). Therefore, vocal gestures forming a vocal style detach from emotional attitude to attach themselves to the speaker’s personality, becoming personal messages (ibidem).

In an experiment (cf. Fónagy 1983: 160-169; 1993), subjects who listened to two actresses (Simone Signoret and Gaby Morlay) playing the same short fragments of Cocteau’s *La voix humaine* identified two definite characters by means of a questionnaire on social and personal behaviour. It was found that, on the basis of vocal information, the woman played by Simone Signoret had a self-defeating personality, or introverted behaviour; while the woman played by Gaby Morlay had an active personality, or extroverted behaviour – details

<i>QUESTIONNAIRE BIOGRAPHIQUE (Voix Féminines)</i>		
Texte	GM %	SS %
1.		
a) Toujours attractive, trente-quatre ans .....	45 (58,4)	4 (5,3)
b) Elle était belle il y a dix ans, maintenant elle en a quarante, elle est fanée, le visage légèrement ravagé .....	13 (16,9)	43 (57,3)
c) Jeune femme jolie : 26 ans .....	17 (22,1)	4 (5,3)
d) Laide : 50 ans .....	2 (2,6)	24 (32,0)
2.		
a) Elle est toujours délaissée par ses partenaires .....	5 (6,6)	47 (61,8)
b) D'habitude c'est elle qui se lasse de ses amants .....	37 (48,7)	8 (10,5)
c) Première aventure d'une femme bien rangée .....	6 (7,9)	19 (25,0)
d) Elle n'est pas très touchée par cette rupture .....	28 (36,8)	2 (2,6)
3.		
a) Elle appartient à un milieu modeste, elle est sténodactylo dans une usine .....	14 (25,9)	30 (53,6)
b) Fille d'un riche avocat, veuve d'un industriel, ne connaît pas de soucis matériels .....	40 (74,1)	26 (46,4)
4.		
a) Au cours de leur liaison elle dominait, guidait, conseillait son partenaire .....	49 (79,0)	13 (19,4)
b) Elle se comportait plutôt comme une petite fille qu'il fallait mener par la main .....	13 (21,0)	54 (80,6)
5.		
a) A l'école elle est très appliquée, ses résultats cependant sont loin d'être brillants .....	11 (15,7)	46 (68,7)
b) Une des meilleures élèves, sans se donner trop de mal .....	30 (42,9)	10 (14,9)
c) Cancre .....	7 (10,0)	8 (11,9)

Figure 3.6. Questionnaire on the vocal style of Gaby Morlay (G.M) and Simone Signoret (S.S.) (Fónagy 1983: 162-163) – part I.

about the aspects of behaviour elicited, with results of responses, are shown in figure 1 (see below). Gaby Morlay's voice, compared to Simone Signoret's, presented: a higher phonation rhythm; a lower number of pauses (especially those relative to hesitations) with a shorter duration; stresses more defined, with a regular distribution; vocal folds were more relaxed and therefore had a more relaxed vibration (a more intense sound – a full voice); a higher tone level, with a higher range of tone variation (rising, rising-falling, falling), but with a lower number of falling tones; a more definite articulation, without spasmodic constraint and with a higher and more advanced point of articulation (cf. Fónagy 1993: 16-17). S.S.'s voice has a slow rhythm, pauses are longer and more frequent, stresses are weak, melodic movement is monotonous (cf. Fónagy 1993: 19-21). Gaby Morlay's voice may sound more attractive because it is more melodic, closer to singing voice, more enchanting. It has a higher regularity: vocal folds vibrate in a freer and regular way, therefore producing a sound that oscillates around the same tone. Such regularity is reflected by syllabic repetition: the sequence of sounds is regular, therefore more predictable. This recurrence may be at the basis of enchantment (cf. Fónagy 1993: 22).

It was found that an artistic voice could produce a very high degree of semantic condensation (cf. Fónagy 1983: 238).

Fónagy (cf. 1983: 243-255) conducted a series of experiments called “*jeu d'écho*”, used in experimental psychology, where subjects were asked to mimic utterances after few seconds pause; they were recorded, recording were acoustically analysed and subsequently played for other subjects who had to assess them according to modal or semantic categories identified by the researcher. In one of these “*jeu d'écho*” (ibidem: 251-253), the Hungarian expression “*Az én hibám volt*” [“*était-ce (ou: c'était) ma faute*”] – uttered by József Timár in a performance of *Death of a Salesman* (Arthur Miller, 1949) – was considered as an assertion by four out of ten respondents, while the other six subjects judged it as a question. The two echoes (or shadowed utterances) were considered unanimously either assertive or interrogative; in both cases, the intonation had a rising-falling pattern. On a semantic level, the original “ambiguous” utterance made by the actor was assessed (on a seven-point scale) according the following semantic categories: ‘invitation’, ‘complaint’, ‘surprise’, ‘doubt’, ‘sadness’, ‘regret’. Respondents assigned a high rate (5 to 7) to all the attitudes mentioned, which were all considered expressed by the character. The echoes obtained from the actor's utterance were also assessed



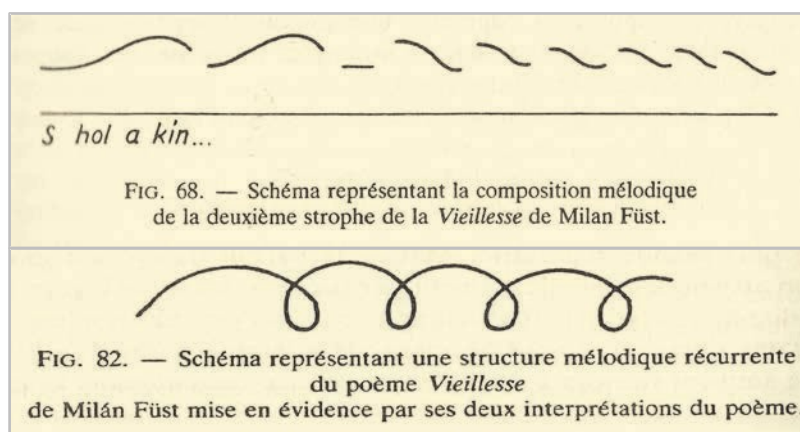
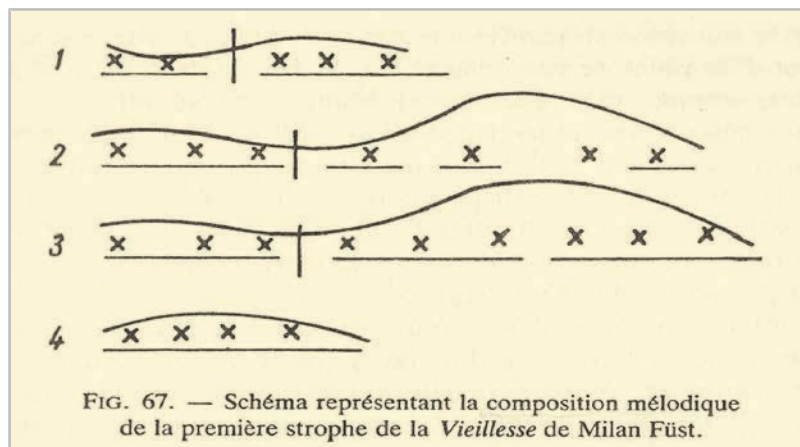
in the same way; from the semantic test three echoes were selected, previously judged as: a ‘question’, an ‘exclamation’ and an ‘exclamative question’. In this case, where one of the attitudes prevailed, the others were assigned a very low rate – ‘contradictory’ attitudes showed a ‘complementary variant’: differing semantic dimensions showed low rates. The average rate of the utterance made by the artist was 5.56, while the average aggregate rate of the echoes was 2.94. Similar findings in similar experiments (cf. Fónagy 1983: 243-251) seemed to confirm the hypothesis according to which particularly expressive utterances made by artists seem to ‘condense’, or overlap with, a variety of several simple utterances, sometimes contradictory, at both prosodic and semantic level (cf. *ibidem*: 254-255).

Texte	GM %	SS %
<b>d) Fait l'école buissonnière en permanence, doit changer deux fois d'école .....</b>	<b>22 (31,4)</b>	<b>3 (4,5)</b>
<b>6.</b>		
<b>a) S'habille avec beaucoup de goût, elle est élégante, même dans des robes peu coûteuses ....</b>	<b>38 (50,0)</b>	<b>10 (11,9)</b>
<b>b) S'habille avec maladresse et peu de goût, elle n'est jamais élégante, même dans une robe relativement chère .....</b>	<b>2 (2,6)</b>	<b>44 (52,4)</b>
<b>c) Comme adolescente, elle portait des robes de jeune fille de vingt ans — à partir de trente elle s'habille un peu plus jeune que son âge .....</b>	<b>32 (44,7)</b>	<b>6 (7,1)</b>
<b>d) Comme adolescente, elle était habillée comme une petite fille de huit ans — depuis qu'elle a trente ans, elle s'habille comme si elle en avait quarante .....</b>	<b>2 (2,6)</b>	<b>24 (28,6)</b>
<b>7.</b>		
<b>a) Elle était une fille à maman .....</b>	<b>11 (16,2)</b>	<b>21 (30,4)</b>
<b>b) Elle était une fille à papa .....</b>	<b>41 (60,3)</b>	<b>12 (17,4)</b>
<b>c) Elle était orpheline .....</b>	<b>1 (1,5)</b>	<b>17 (24,6)</b>
<b>d) Ses parents ne s'occupaient pas d'elle .....</b>	<b>15 (22,1)</b>	<b>19 (27,3)</b>
<b>8.</b>		
<b>a) Elle était la sœur aînée .....</b>	<b>31 (52,5)</b>	<b>34 (54,8)</b>
<b>b) Elle était la cadette .....</b>	<b>28 (47,4)</b>	<b>28 (45,2)</b>

**Figure 3.7.** Questionnaire on the vocal style of Gaby Morlay (G.M) and Simone Signoret (S.S.) (Fónagy 1983: 162-163) – part II.

Further analysis of the poet's reading revealed the ability of “artistic interpretation” to “transpose” the “tangled” lines from the “visual channel” to the “sound channel” of the poem (*ibidem*: 297-307). Transcription in music

notation of the melodic curve of the poet's readings revealed a “recurrent melodic structure” of the poem, confirmed by semantic tests, that goes beyond ordinary (linguistic) prosody.



**Figure 3.8.** Examples of synthesis of recurrent melodic pattern found by Fónagy (1983: 275, 302) in a poem read by the author himself.

In artistic interpretation, intonation seems to follow the rules of music. Thus, melody enriches the text with ‘musical expression’, which adds to the ‘linguistic expression’, serving the ‘representational function’ of language (ibidem: 306). The evocative technique of vocal artist adds a third dimension the melodic movement: ‘melodicity’ - that adds to time (duration) and height (tone) (Fónagy 1983: 310 – for the translation of the term *musicalité* into ‘melodicity’, the source was Fónagy 2001: 102).

On pourrait concevoir la musicalité comme une dimension de profondeur de la mélodie phrastique qui lui [l’artiste vocal] permet tantôt de s’approcher, tantôt de s’éloigner du plein-chant. La musicalité de la voix dépend de la régularité de la distribution des fréquences fondamentales à l’intérieur d’une syllabe (Fónagy 1983: 310).

Melodicity ‘increases with tenderness and sharply decreases in the expression of aggressive emotions’ (Fónagy 1983: 311).

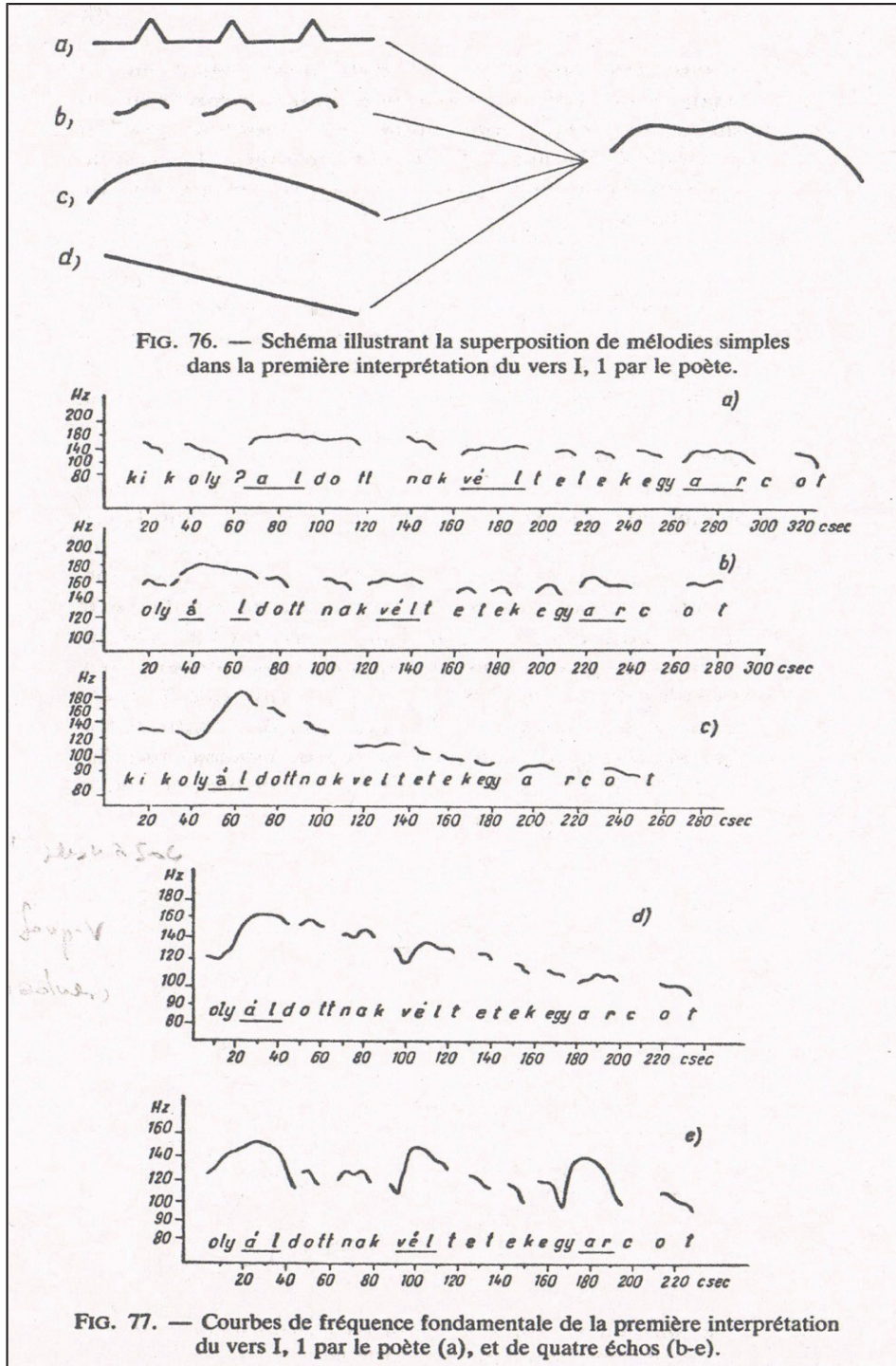


Figure 3.9. Melodic curves compared (Fónagy 1983: 294).

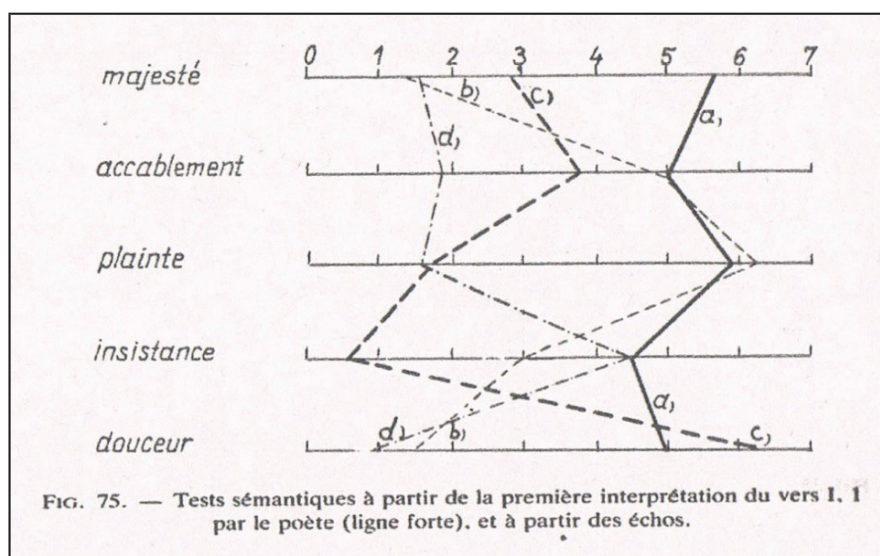


Figure 3.10. Semantic tests compared (Fónagy 1983: 293).

### 3.6.5 Poetry and vocal art

Fónagy's discussion and conclusion (1983: 316-321) about the outcomes of the experiments on artistic voice is reported in this subsection.

Information provided by live speech, compared to that of writing, is higher, since the execution of every single phoneme implies a multiple choice, and every choice is meaningful. Prosody and articulation considered 'expressive' are such because they are perceived as 'varyingly different from everyday experience'. For this reason, the more an interpretation moves away from the 'vocal execution expected according to a text', the more it adds to the same text. However, such deviation has to stay within the limits of comprehensibility, in order to acquire an aesthetic effect and thus be effective. The measure of the aesthetic effect is the 'degree of surprise'. The condition for the 'unexpected' to be 'aesthetic' is that the gap between the expected and the unexpected allows the codification of vocal (stylistic) message: only meaningful gaps can be expressive. Actually, interpretation may depart from the expected performance according to the text, in order to 'better express the content of a poem'. It is the vocal interpretation that realises a poem', 'materialising one of the numerous version a poem entails'. The constraints that a poem imposes on artistic interpretation, or on vocal creation, are provided mainly by the syntactic structure, which determines 'the melodic line of the recitation', but also by the 'frequency of phonemes'. 'Artistic precision' requires the poet not to use ordinary language. Moreover, the poet 'completes'

the text with a ‘sound mimesis’, since s/he makes a double usage of sounds: (i) those in the text (signs in the poem); and (ii) those that are independent from the poem, which are ‘direct expression of emotional contents [that] doubles the signification of sentences’. Hence the opposition between fundamental and occasional sense. A vocal artist also uses both the ‘expressive transposition of melodic forms’ and the ‘expressive distortion’, thus mixing melodic metaphors. In this way, the vocal artist develops a ‘personal condensation technique: the overlapping of simple melodic forms’. Thus, concision, condensation and economy are ‘characteristics of artistic communication’, they ‘provide a lively pleasure’. The poet and the vocal artist use musical means, playing with words and sounds, to express what goes beyond ‘conceptual thinking’; they use metaphors, voice, gesture and mime to ‘say more than they know’, or to say what they still do not know, but what they sense.

Ce langage primitif, le prélangage, mis en relief par la poésie et l’art vocal, survit modestement à l’intérieur de la langue. [...] en raison de son caractère primitif, le prélangage dépasse en efficacité la communication linguistique. C’est en recourant aux moyens du langage primitif que poésie et art vocal se constituent en poésies, en créations et actions (Fónagy 1983: 320).

Poetry and vocal art share a ‘certain equilibrium between increased information and high redundancy’. The lexical, grammatical and phonetic elements in a poem are ‘strictly organised’; even the content follows metrical constraints. These structures are in turn organised in superstructures, or architectures, that ‘create the musical unity of the poem’. The same is true for the ‘more structured prosody’ of a ‘vocal composition’, i.e. ‘more redundant than melody and rhythm in ordinary speech’. Such articulation, or organisation, facilitates perception; this ‘mental economy is calculated in the form of an aesthetic pleasure’.

### **3.7 Constancy of expressive traits and their symbolization**

In the end, Fónagy’s proposal is that also expressive information, or physiognomic and pathognomic information (see above – section 3.1.8), shows constant traits that can be processed by the hearer along the development of someone’s speech. This conclusion seems to be not in line with Bühler’s proposal (Bühler 1934/1990: 324 – see section 3.1.8) of constancy of traits of “factors” of “units that can be linguistically registered”. It may be supposed that, according to Fónagy, expressive factors are not only

deictic signs, but also iconic (the spatial representation of different attitudes in the melodic curve) and symbolic, for their aesthetic effect. With reference to the present study, the simultaneous interpreting speech may be compared to that of a poet reading her/his own poem (see above - previous section), since also the interpreter is the creator of his own interpretation.

### **3.8 Proposal for a questionnaire to elicit a subjective gestaltic assessment of voice quality in television broadcast simultaneous interpreting**

The aim of the questionnaire is to assess the overall quality of television broadcast simultaneous interpreting testing the gestaltic perception of the voice. This approach also takes into account the role that the audiovisual medium may play in perception. For all these reasons, the items of the questionnaire will be different in part from the ones used so far in previous studies of the kind (see below).

The first step to create the questionnaire was the construction of a deep structure, based on the main features defined in the theoretical contributions of *Il volto fonico delle parole* ('the sound face of words') (Albano Leoni 2009), *L'Audio-vision* (Chion 1990/1994) and *La vive voix* (Fónagy, 1983).

#### **3.8.1. Deep structure of the questionnaire: categories related to the sources of the theoretical paradigm**

The traits of *Il volto fonico delle parole* (Albano Leoni 2009) were all taken as properties of reference related to a model of speech perception, namely: 'syllable', 'voice', 'prosody', 'sense', 'context' and '(linguistic) knowledge of the world'. With reference to Chion's *Audio-vision*, 'synchresis' (synchronization + synthesis) was taken as a property, due to the importance of the gestaltic perception of sound-image: the image synchronized to the sound is a unit of meaning, mainly determined by the audio-visual medium. Given the fact that sound (voice) is prominent in television (cf. Vilches 1989: 209-223), the traits of Chion's proposal for an 'audiologovisual poetics' were also taken as properties: 'theatrical speech', 'textual speech' and 'emanation speech'. As to the properties extracted from Fónagy's *La vive voix*, these were: 'expressivity', 'comprehensibility', 'melodicity', 'vocal attitude' and 'vocal personality'. The property of 'comprehensibility' is opposed to that of 'expressivity' in the aesthetic perception of non-ordinary expressive speech (see section 3.5.2.5), just as the simultaneous interpretation speech is. Moreover, from the

expressive point of view, it can be compared to the reading of a poem by the same author, considering that the interpreter creates her/his own interpreted text (see sections 3.5.2.4 and 3.5.2.5).

DEEP STRUCTURE OF QUESTIONNAIRE			SUPERFICIAL STRUCTURE
Properties of:			Questionnaire flow
<i>Il volto fonico delle parole</i>	<i>L'Audio-vision</i>	<i>La vive voix</i>	Quality criteria / Items
<i>Prosody</i>	<i>Synchresis</i>	<i>Expressivity</i>	1. Articulation 2. Hesitations 3. Audible breaths 4. Silent pauses 5. Speed of speech 6. Melodious / monotonous voice
<i>Voice</i>	<i>Textual speech</i>	<i>Comprehensibility</i>	7. Same melody repeated (sung speech) 8. Sweet / aggressive voice 9. Self-confident/insecure voice
<i>Syllable</i>		<i>Melodicity</i>	10. Active / self-defeating personality of the interpreter 11. (In)Expressive voice 12. (Un)Comprehensible voice
<i>Sense</i>	<i>Theatrical speech</i>		13. Interpretation-image synchrony 14. Informativity of interpretation with respect to the image
<i>Context</i>	<i>Emanation speech</i>	<i>Vocal attitude</i>	15. (In)Credible interpreter 16. (Un)Skilled interpreter 17. Simple / complex words
<i>(linguistic) knowledge of the world</i>		<i>Vocal personality</i>	18. Non-natural / natural syntax 19. Simple / complex sentences 20. Comprehensible interpretation (self-assessment of comprehension) 21. Real comprehension of text (multiple-choice test)

**Table 3.1.** Proposal for a questionnaire to elicit a gestaltic assessment of voice quality in TV broadcast simultaneous interpreting.

### **3.8.2. Superficial structure of the questionnaire: items related to the properties of *Il volto fonico delle parole* (Albano Leoni 2009)**

With reference to the property “syllable”, taken from ‘the acoustic face of words’, the related items of the questionnaire were: “articulation”, “hesitations”, “speed of speech” and “same melody repeated”. “Articulation” mainly refers to clear pronunciation, word enunciation, ‘speaking with distinction’, phono-syllabic scanning, which relates to the production and perception of syllables. The item “hesitations” includes self-repairs, repetitions, corrections, vowel and consonant lengthening, filled pauses, vocalizations; all these phenomena may perceptually assume the form of syllables and can be considered as ‘phonological words’ (Albano Leoni 2009: 165). The item “speed of speech” was also related to phono-syllabic scanning. The item “same melody repeated”, meaning “sung speech”, was directly related to the aspect of “melodicity”. It was related to “syllable” via the definition of “melodicity” given by Fónagy (1983: 310), i.e. the ‘musicality of voice depending on regularity and distribution of fundamental frequency in one syllable’ (see also section 3.5.2.4). Actually, the item “same melody repeated” was an attempt to elicit the perception of “melodicity” that Fónagy detected in the artistic use of voice, as we assumed that a voice used in television is expected to have artistic features (Straniero Sergio 1997: 54). The items “audible breaths”, “silent pauses”, “natural/non-natural syntax”, “simple/complex sentences”, “melodious/monotonous voice”, “sweet/aggressive voice” were all related to prosody. Prosody is strictly linked to syntax, since in orality prosody conveys the meaning of noun groups, phrases, clauses and sentences. Silent pauses, be they executed consciously or unconsciously, influence rhythm and intonation; consequently, so do audible breaths, since these are pauses, although not entirely silent. Since we decided not to include “intonation” among the items to simplify the language, then the item “melodious/monotonous voice” has the function of eliciting a pleasant or unpleasant melodic curve. For reasons of simplification of the language of the items, the timbre was not included among the items. The item “sweet/aggressive voice” is also a function of “melodious/monotonous voice” (see section 3.5.2.4), since the impression of the “vocal attitude” (“sweet/aggressive”) depends on the form of the melodic curve, which in this case is highly influenced by “emphatic accents” (see section 3.5.2.3), and then by rhythm.



The items related to “voice” include: “articulation”, “hesitations”, “speed of speech”, “melodious/monotonous voice”; “same melody repeated”; “sweet/aggressive voice”; “active/self-defeating personality of the interpreter”. The last two items are based on the impressions of the speaker’s attitude and the kind of behaviour the voice may convey, as proposed by Bühler (1934/1983 and 1933/1978; see section 3.1.8) and demonstrated by Fónagy (1983; see section 3.5.2.4). The items “comprehensible voice” and “expressive voice” refer to the aesthetic aspects of artistic voice (see section 3.5.2.5). The item “credible voice” is mainly related to the physiognomic information about the interpreter conveyed by his/her voice as perceived through the audiovisual medium (see sections 3.1.8, 4.0.2. and 4.0.3): since television broadcast simultaneous interpretation mainly occurs in TV programs that are completely or partly information programs, TV viewers expect the interpreter’s voice to be as objective as a newscaster’s, in order to be credible (Vilches 1989: 218-219).

If we consider that in TV broadcast simultaneous interpreting, as in film interpreting, voice plays a major role in helping TV viewers interpret the message, i.e. to assign signs to it (Russo 2005; Straniero Sergio 1997: 54; Vilches 1989: 209-223), then the items more strictly related to the “sense” property are “comprehensible” and “expressive voice”, “credible voice” and “comprehensible interpretation”, since they are relevant aspects to measure the subjects’ self-perception of their processing of sense. This is also true for the items “complex/simple words” and “sentences”. The item “active/passive personality of the interpreter” was also considered relevant for the “sense” property, because it provides information on how much the interpreter has anchored the audience, making his/her speech more accessible to sense processing, consistently with the medium and the context.

In order to define the items to be related to the “context”, the main point of reference was not only represented by the communicative situation where a TV broadcast simultaneous interpretation may take place (broadcast journalism of any kind, press conferences, talk shows, media events; a broadcast or recorded conference or convention, etc.), which could be defined as the ‘context of production’. However, we considered more relevant to the sense what could be defined the ‘context of reception’ of a TV broadcast simultaneous interpretation, which takes into consideration the reception of the audiovisual text by subjects, through the audiovisual medium. Therefore, the items more strictly related to the “context” property are:

“melodious/monotonous voice”; “same melody repeated”; “comprehensible interpretation”, “expressive voice”; “credible voice”; “interpretation-image synchrony”; “informativity of interpretation with respect to the image”; “skilled/unskilled interpreter”.

The items related to the “linguistic or non-linguistic knowledge of the world” were: “complex or simple words”; “complex or simple and sentences”; “comprehensible interpretation”; and a test on “real comprehension of text” (interpretation). The first three items were supposed to elicit the subject’s perception of comprehension, or the respondent’s interpretation of text, filtered by her/his paradigms, coloured by his/her personal conditioning, her/his past experience of events; the last item is objective, since it is a test based on the information provided by the interpreted text.

It may be argued that, considering the “deep structure” made of the categories extracted from the theoretical references to which the items are related, the questionnaire is redundant, since all the properties of ‘the sound face of words’ recur in all the items. In other words, it is clear that the gestalt perception considers all the properties of the ‘sound face of words’ all co-existent; nevertheless, the researcher admits that could not avoid a ‘structural approach’ to define the items of the questionnaire. The argument of redundancy is undeniable; however it was inevitable, considering the theoretical approach used, which is based on the total perception of speech production, audiovisual perception and speech perception. The redundancy of the questionnaire is due to: i) the interrelations among intensity, pitch and duration inherent in the word ‘prosody’ (which is the modulation of the three elements); ii) the interrelations among prosody and semantic and syntactic operations in speech, due to the cognitive activity of its production (Goldman-Eisler 1968: 6-10; 90-93); iii) the interrelations among the speech and the person of the speaker, both in production (Goldman-Eisler 1968; see also section 1.4.1) and in perception (Fónagy 1983: 160-169; 1993; see also section 3.5.2); iv) the interrelations among the various formal aspects, and among these and content aspects, detected in the quality assessment of simultaneous interpreting (Collados Aís et al. 2007); v) the interrelations between sound and image in audiovisual perception, which, again, was defined as “rhythmic” (Chion 1990: 136; see section 3.5.1). In addition, there are the proposals of: i) rhythm as form by Benveniste (1966; see section 3.2); ii) rhythm as form of a discourse (Meschonnic 1982; see section 3.3).

## Chapter 4

# PILOT SURVEY ON VOICE QUALITY OF A TV BROADCAST SIMULTANEOUS INTERPRETATION ASSESSMENT THROUGH GESTALT PERCEPTION

### 4.0 Introduction

In this section, the main aspects related to TV and film interpreting that emerged from the Interpreting Studies are reported, before explaining our research. In addition, semiotic aspects related to the perception of voice in television and radio information programs are discussed.

#### 4.0.1 Television Interpreting and Film Interpreting

In an article in which a number of selected studies on simultaneous interpretations of “US Presidential material” – in the section titled “methodological issues” – Franz Pöchhacker (2011:25) maintains that:

Considering the large target population, TV interpreting would also seem to lend itself very well to survey research approaches as used in studies among users of conference interpreting (e.g. Kurz 2002/1993; Moser 1996). Surprisingly little work of this kind has been carried out, however, which may also indicate a lack of interest in such audience studies among TV executives.

The article was mainly dedicated to “empirical analyses, several of them conducted in the framework of recent MA theses completed at the University of Vienna under the author’s supervision”. Among these studies, the one on

“audience preferences” by Schwarthorer (2010; in Pöchhacker 2011:30) was a survey on both quality expectation and quality assessment of television interpreting. The conclusions led the author and the supervisor of the thesis to state that

Subjective judgments as reflected in criteria-based ratings appear to be quite reliable and consistent, in lay audiences as well as those with a background in interpreting. The preference established in Schwarthorer’s comparison between two highly professional German interpretations of President Obama’s inaugural address appears to be rooted in nonverbal (vocal) features, once again confirming previous findings (e.g. Kurz/Pöchhacker 1995) regarding the special importance of prosodic and voice-related qualities in the media setting (Pöchhacker 2011:30).

With respect to TV interpreting, Mack (1999) claimed that:

[...] gli ‘addetti ai lavori’ [gli interpreti] tendono ad ampliare il discorso ad aspetti comunicativi e culturali e all’apporto scientifico dell’interprete al prodotto televisivo, dando quasi per scontate caratteristiche riassumibili come qualità della voce e della presentazione. Le osservazioni dei fruitori si concentrano invece, a vario titolo, su aspetti tecnici dell’ascolto e peculiarità del parlato del (singolo) interprete preso come riferimento. Alla luce di queste osservazioni è comunque evidente che l’aspetto vocale-soprasegmentale è parte inscindibile della percezione del parlato dell’interprete – tanto più nella simultanea televisiva, dove per il telespettatore l’interprete è solo una voce.

In 2007, after many years of research on the Television Interpreting Corpus (*CorIT*), Straniero Sergio published an in-depth study on interpreting in TV talk-shows, outlining this particular genre through conversational, enunciation and performance analyses of the communicative situation at stake. The following conclusion is particularly meaningful for the research subject under consideration:

Nell’IT [interpretazione televisiva], la norma iniziale più importante è quella della *telegenia*, ovvero il grado di fruibilità televisiva. Un’interpretazione può considerarsi telegenica nella misura in cui soddisfa il criterio della forma, indipendentemente dal fatto che il discorso dell’interprete sia una trasposizione fedele del TP. Il primato della forma sul contenuto, l’importanza della voce e, più in generale del *packaging* della prestazione sono ritenute una norma specifica dell’IT da parte del pubblico, dei *broadcasters* (cfr 1.1), degli stessi *media interpreters*, ma anche dei giornalisti che, nel commentare un’IS andata in onda il giorno prima, tendono a soffermarsi quasi esclusivamente sui fattori non linguistici (Straniero Sergio 2007: 541).

Paolo Maria Nosedà has become famous in Italy ever since he became the official interpreter of the TV talk-show *Che tempo che fa*, broadcast by

Italian state-owned television channel *Rai Tre*. With reference to his work he stated that

L'interprete diviene la “voce” della persona che deve tradurre, ma non solo. Per pochi minuti, per ore o giorni, l'interprete diviene il tramite della comunicazione e deve anche saper comunicare il cuore degli ospiti. Questa è appunto una delle caratteristiche salienti e meravigliose del mio lavoro, ma è anche una grande difficoltà: da un lato occorre non tradire la personalità, i contenuti e l'immagine della persona la cui comunicazione devo gestire, dall'altro l'interprete deve rendersi conto dove si collochi il confine fra creatività e realtà. È necessario calarsi nella personalità di colui o colei che si traduce e tuttavia senza esagerare. Scegliere un registro, uno stile, un modo di espressione – spesso in una manciata di secondi – significa anche essere molto umili ed accettare anche tratti caratteriali che possono essere soggettivamente fastidiosi o inopportuni e non filtrarli attraverso un giudizio personale (Nosedà 2008).

However, while the adoption of a ‘subjective’ and ‘personalized’ vocal style by the interpreter could be suitable for a speaker interviewed in the context of a talk-show or a section of a TV information program, it might not be the same for an interpretation of the US Presidential Debate. Even if this is considered a media event, the informative function may prevail over the function of entertaining; therefore, the audience may expect an ‘objective’ voice (see below – section 4.0.2).

As for Film Interpreting, the study by Russo (2005; see above – section 1.1.2.8) is relevant to the present research project, if we consider the following conclusion:

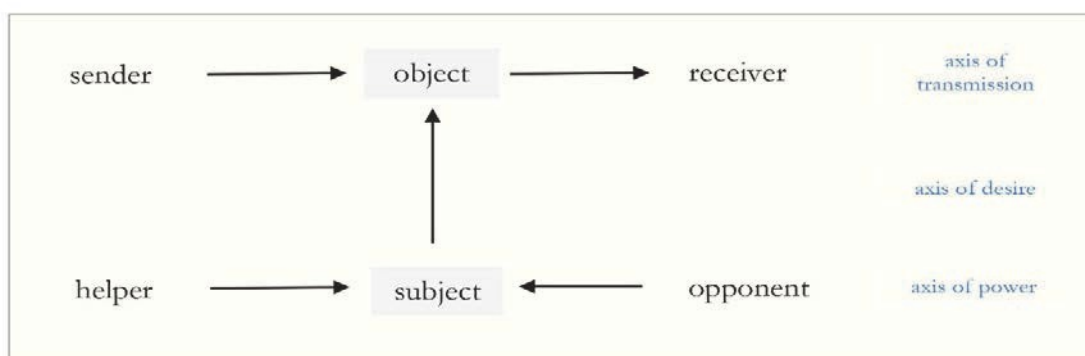
The feature most appreciated in professionals was *pleasant voice* [...] and in students *correct formal and grammatical usage* [...]. The most appreciated features according to user group are: *voice, correct formal and grammatical usage* and *synchronisation* for critics [...] and *voice* for students [...] and "other" [...]. This finding supports the view that presentation skills and synchronisation are particularly relevant to media interpreting, as shown also by Kurz and Pöchhacker (1995) (Russo 2005: 19).

#### **4.0.2 TV information, voice and TV broadcast simultaneous interpretation**

According to Vilches (1989: 209-222) in television narration, the narrator's voice prevails on other sounds and on any image or sequence of images (ibidem: 209). In particular, in television information programs, the voice is ubiquitous, since it can get in and out of the image; therefore, the

voice in television ‘acquires the *status* of individual actant, independently from the body that emits or transform it’ (ibidem).

The term *actant* is taken from the *actantial model* proposed by Greimas (1969: 211-212; in Marchese 1978: 32) to describe the structural roles typically performed in storytelling; in particular, the different roles performed in myths and folktales by an indefinite number of characters can be reduced to six semantic categories, the actants (Marchese 1978:32), that may interact according to the following schema:



**Figure 4.1** The actantial model by Greimas (1966; in Schleifer 1987:103).

An actant only represents functions in narrative: it can be either a subject performing an action or an object receiving it; if one considers the different actions taking place in a story, an actant can be subject and object at the same time, depending on the action considered; an actant may embody itself in a person, a group of people, an animal, a thing, a concept, a moral value, etc.; this objectification (or personification) of one or more actants on a more concrete level gives rise to the *actors*; an *actor* may represent several *actants*, and one *actant* may be distributed in several *actors* (cf. Vilches 1989: 210; Reis and Lopes 1990: 14-15).

With the aim of applying the actantial model to the narrative of TV information programs, Vilches (ibidem: 210-211) assigned the voice the roles of: i) ‘actant of the communication or the enunciation’, with reference to the voice of narrator/narratee, for example in the persons of interviewer/interviewee; and ii) ‘actant of the narration or enunciate’, with reference to the narrating/narrated voice, where a narrated voice can be, for example, the silenced voice of an actor of the information (see below – this section).

The voice in television is an actant that can be present in more than one frame (ibidem: 211). In a television audience’s perception, the narrating voice

takes the place of the speaking character, just as in music videos or song videos, where the sound, with its linear development, creates the unique spatial dimension where the sequence of images are contained (*ibidem*). In television information, the voice is ‘larger than the body and the space’, it is the absorbing frame, the continuous background against which the image is the frame absorbed; the body of the character is fictionalized by the hyperreality of the narrating voice (*ibidem* 211-212). The narrating voice is: i) ‘ubiquitous’, because it occupies any space: in and out of the bodies, in and out of the scene of the news event; ii) as a consequence of this, it is ‘all-seeing’, it is a ‘panoptic voice’, watching over the scene of the news and inside the characters, even if it cannot be seen; iii) ‘all-knowing’, because it knows everything about the event and the people involved; iv) ‘all-powerful’, because it can say what it wants (*ibidem*: 212). This last effect of the narrating voice in television information cancels the essence of the actors of the news event or the opinion, including the personal traits expressed by the voice (origins, psychic status, sex, age, etc.) (*ibidem*). For this reason, the narrating voice in television information has to be ‘uniform’, with a ‘good articulation’, good control of prosody, while ‘personal attitudes and emotions disappear due to the high speed of speech’ (*ibidem*).

La fuerte entrada en escena de la voz en la información diaria, esa tendencia a la globalidad gestáltica de la “situación sonora”, tiene algo de ese carácter autorreferencial que tienen todos los medios de comunicación. Por más que estén hablando de otra cosa, en el fondo, hablan de sí mismos. De ahí que, desde esta “perspectiva auricular”, el sonido de la voz adquiera en la información su *status* simbólico, donde lo que menos importa es el significado, el efecto, el contenido. La voz narradora o dobladora en los informativos se constituye en símbolo de la noticia (Vilches 1989: 213).

The kinds of functions developed by the actant voice (ubiquitous, all-seeing, all-knowing and all-powerful) are also fulfilled by television as a medium (*ibidem*: 213). Thus, it is: i) ubiquitous because it can send its signal from everywhere, and it can be received everywhere; ii) all-knowing because it can remember everything thanks to the audiovisual memory contained in its archive of recordings; iii) all-seeing because it is ubiquitous and all-knowing, for the same reason it is also exhibitionist; iv) all-powerful thanks to its political and economic power, and for all human and material resources spent in its organization and facilities (*ibidem*).

There are two main kinds of voice in newscasts and, ‘broadly speaking, in cinema and television’, i.e. the ‘voice off [screen]’ and the ‘synchronized

voice'; the synchronized voice can be further divided into 'visible voice' and 'hidden voice' (ibidem: 214-216). The voice off has no power on the image, since it is 'lateral' to it, and is 'directed to the audience'; the voice off is the 'object of a communicative pact' between the newscast and the audience, while the 'image is bartered to strengthen the alliance between television and audience' (ibidem: 214). In particular, the voice off initially constitutes a *proposal* that the newscast forwards to the audience; a proposal that the audience can *accept* or *reject*. As a matter of fact, the audience cannot reject this *proposal*, because it is the only one 'permitting the *interpretation* or comprehension of the information'; therefore, the voice off, 'as a consequence of this status of *mandate*, becomes an *evaluation* of the news' (ibidem: 214-215).<sup>1</sup>

The synchronized voice is part of the image, since it embodies things and persons; it also changes according to the changes in focus and perspective (ibidem: 215). This voice is more credible to the audience, not only because of its simultaneity to time and space, but especially because the audience believes it has a broader range (a higher freedom?) of interpretation; as a matter of fact, the synchronized voice is a means of subtle persuasion (ibidem). Within the synchronized voice, the visible voice can be visually referred to the source, or the person; this is the voice of 'historical pronouncements, prominent political figures, and urgent declarations' where the sound quality is subordinate to the image (ibidem). The visible voice has a 'direct visual field', a 'depth of field' where the 'propagation of sound' and its 'transmission' are 'related' to its 'source' and the 'reception' by the 'audience' (ibidem: 215-216). On the contrary, the synchronized hidden voice has a 'recognizable' voice, but it is not visible (ibidem: 216). From the point of view of 'communicative effects', this kind of voice presents an 'enigmatic' aspect which renders it more significant than the visible voice (ibidem). The hidden voice is a case of 'partial suppression of the matter of expression', and as such it represents a case of 'persuasive manipulation' (ibidem). The power of this enigmatic voice, synchronized to the images, but hidden, can be noticed when it is used to comment on or summarize a pronouncement (ibidem: 216-217). In some cases, in a TV information program, this voice can shut the speaker up, when it is the protagonist of the information, for e.g. a politician, because the spectator watches his/her images, but his/her voice is silenced (lowered or

---

<sup>1</sup> An explanation of the notion of *communicative pact* can be found in Casetti (2002: 53ff).



with no volume at all) by the voice off (*ibidem*). The voice off ‘can inhibit any enunciation different from that of the institution [of news]’ (*ibidem*: 217).

[...] la voix *off* représente un pouvoir, celui de disposer de l’image et de ce qu’elle reflète, depuis d’un espace absolument *autre* par rapport à celui qu’inscrit la bande-image. *Absolument autre et absolument indéterminé*. En tant qu’elle surgit au champ de l’autre, la voix *off* est supposée savoir: c’est cela l’essence de son pouvoir.

Car, si la voix *sait*, c’est forcément *pour* quelqu’un, qui ne parlera pas. Pour quelqu’un, c’est-à-dire à la fois *à son adresse* et *à sa place* (Bonitzer 1975:26 – original text in French; in Vilches 1989:217 – Spanish translation).

Bonitzer refers to the voice off in cinema, but Vilches applies his remarks to television; he adds that the power of voice off originates from its anonymous origins’, and from being the collective ‘voice of newscasts’; the voice off can be considered ‘the metaphor of the power of the information’ (*ibidem*: 217). Precisely because the voice off – either in ‘editing’ televised news or in live broadcasts – is used to ‘eliminate the other’ (the actors) by imposing the ‘Voice of the news’; this is an ‘expropriation of the differences represented by the personal accents, the identity of the other’, to create another identity, without personal marks, with an ‘homogeneous meaning’, similar to that of the ‘fiction in the cinema’ (*ibidem*).

Sólo puede existir un acento, una sola inflexión reconocible: la del enunciador de las noticias. Porque un telediario obedece al principio de realidad y la realidad no puede ser subjetiva, no es la voz de tal o cual la que se escucha, sino la voz al servicio de la objetividad. El cuerpo de la voz es como el grano de la foto, afirma CHION (1982) glosando a Barthes. Pero, ¿quién se fija en un acento particular cuando tiene delante el granero de la palabra? Por eso la voz en *off* debe ser una voz sin acento («tener el acento de la televisión», que es como decir: ¡el acento de todo el país!), sin particularidades regionales, sin marcas propias de educación o de clase. Un acento que neutralice la fuga del sentido. La voz en *off* expone, «presenta», no interpreta, se dice en algunas redacciones de telediarios. Pero, en el fondo, la voz en *off* expone más que la imagen del acontecimiento. Las imágenes, que suelen ser breves y pobres, con frecuencia se ven obligadas a ser repetidas, mientras que la palabra no se detiene ni se repite jamás. En el marco del contrato comunicativo, la palabra en el telediario no es un derecho sino un deber: debe de explicar la realidad, de interpretar lo documental, de documentar lo aparente.

El telediario ha separado la voz del cuerpo para congrega la unidad de la voz al sentido. (Vilches 1989: 218)

As to the relationship between voice and the frame in television, Vilches (*ibidem*: 219-222) maintains that the screen influences the reception of voice. Indeed, the voice perceived through television tends to have bigger

dimensions than those of human bodies represented on the screen. In the cinema, the volume is used to measure the distance with the audio-viewer in relation to the dimensions of the image, and it represents the measure of the distance; therefore, the volume is related to the depth of field. On the perceptual plan, the geometric perspective allows to interpret the difference of intensity of sounds according to different distances from the audio-viewer. Thus, perspective works the same way both for images and for sounds: it is 'a geometric space that allows us to have more or less exact information about the distance and the dimensions of the objects included in the observator's point of view' (ibidem: 220). However, in cinema perspective can be used also to create illusions, consider, for example, the use, in the early stage of cinema, of children dressed like adults in order to make them appear as a crowd walking at a longer distance from the camera than it really was (ibidem).

In television information, it may happen that the voice of the actor of the information is overlapped by a voice off either of the comment or the translation, at a similar or higher volume (ibidem: 221). In most cases, 'the need for clarity and intelligibility of words leads to record the voice off always on the foreground', at the expense of the sounds and voices produced by the actors on screen (ibidem). In the end, 'the voice of comments monopolizes the news' (ibidem); in newscasts, 'the information is hierarchically ordered by the voice'; this is also true for media events (ibidem 221-222). The effect on the audience is that it identifies itself with the voice:

Si hay un yo con el cual el espectador ansía identificarse, difícilmente será el protagonista del acontecimiento. Lo hará seguramente con la voz en *off*, porque la voz en *off* es el primer plano. Su cercanía al micrófono la convierte en una voz cuya presencia definida por la proximidad permitirá que el auditor se identifique más directamente con ella que con cualquier otro elemento de la escena. Ella, la escena, es lo englobado, lo lejano, la distancia necesaria para crear diferencias de espacio entre la voz den primer plano y el resto. El verdadero objeto del espacio de una noticia es la voz del telediario. Por ello la voz en *off* es el vehículo de la autorreferencialidad. La televisión no habla de otra cosa que de sí misma (Vilches 1989: 222).

It is worth pointing out that both Bonitzer and Vilches talk about *voice off* as both *voice off* and *voice over*, but the features that they attribute to the phenomenon defined as *voice off*, are those of the *voice over*. Casetti and Di Chio (1994:90) define sound in cinema as 'diegetic if its source is present in the space of the event being represented', and 'non-diegetic if its source has nothing to do with the space of the story'. A diegetic sound can be 'onscreen'

or ‘off-screen’, ‘depending on if the source is inside or outside the limits of the frame’; and it can be ‘interior’ or ‘exterior’, ‘depending on if its source is inside the soul of the characters or it has an objective physical reality’ (ibidem). All ‘*non-diegetic* sounds’ and ‘interior diegetic sounds are also called sounds *over*, since they do not come from the physical space of the event’ represented (ibidem):

    Riassumendo, allora, distinguiamo tre categorie di suono: il suono *in* propriamente detto (il suono diegetico esteriore, la cui fonte è inquadrata), il suono *off* propriamente detto (il suono diegetico esteriore, la cui fonte è inquadrata) e il suono *over* (il suono diegetico interiore, *in* o *off* che sia, e il suono non diegetico (Casetti, Di Chio 1994: 90).

These narratological categories related to the narrative functions of sound in cinema and television, can be applied to the context relevant to the present research study, i.e. the live broadcast interpretations of the US Presidential Debates, especially to define the narrative function of the simultaneous interpreter’s voice. In this media event, the voice of the speakers and the moderator (and the audience in the studio intervening with questions directly addressed to the candidates, according to the format of the Second Presidential Debate) are *diegetic* voices *onscreen*; the voice of the simultaneous interpreter is surely *off-screen* and *over*, since it does not come from the physical space of the event represented, because the interpreters work in the facilities of the broadcasters, i.e. in states of the target languages, and is not in the frame; however, the narrative function of the interpreter’s voice is not clearly definable with respect to the space of the story. Indeed, the interpreter’s voice is *non-diegetic* because it develops the function of framing the event for the audience, setting it in the context or situating it; however it cannot be said that it has nothing to do with the story, because the interpreter is fundamental for the audience to understand the story and follow it. The way the event is broadcast adds more relevance to the interpreter; as a matter of fact speakers’ and interpreters’ voices are overlapped, but the interpreter’s voice is transmitted at a higher volume than that of the speaker’s voice; therefore, the interpreter’s voice it is the only channel the audience has to follow the story. In addition, the interpreters know the format of the US Presidential Debates, the subjects dealt with in each debate and the speakers (as they have to study them before executing the interpretation); therefore they know part of the story in advance. For these reasons the interpreter voice can be considered *diegetic*. To conclude, it is worth considering the classification proposed by

Genette (1976; in Marchese 1978: 335-336) regarding the ‘narrative level’ (extradiegetic/intradiegetic) and the ‘narrative relationship’ (homodiegetic/heterodiegetic) of the voice (‘narrating instance’) with respect to the narrative in question, i.e. the simultaneous interpretation of US Presidential Debates. The ‘narrating instance’ is ‘the process of enunciation or narration where the narrator emerges’ (ibidem). Thus, the interpreter’s voice oscillates between the extradiegetic and intradiegetic fields, because the interpreter speaks in first person, but s/he can also speak in third person to report what a speaker has just said in case of overlapping turns; in addition, the interpreter’s voice oscillates between heterodiegetic and homodiegetic fields, because the interpreter is not physically present in the space of the story, but s/he in some way belongs to the story, and is fundamental to its understanding.

#### **4.0.3 Radiophonic information and credible voice**

According to Iglesias Fernández (2011:42), particularly relevant to the definition of voice quality in interpretation is the ‘Acoustic Model of Credibility in radiophonic communication (MAC)’, according to which the speaker’s credibility is structured by 12 variables (Prado et al. 1988-1997; in Iglesias Fernández 2011:42): ‘pleasantness, familiarity, honesty, cleverness, naturalness, nervousness, depths, responsibility, self-confidence, sympathy, sincerity and verisimilitude’ (Prado Pico 1988; in Soto Sanfiel 2000: 10). This semantic structure of credible voice was the result of a study carried out through a questionnaire on perception of radiophonic voice; and the definition of credible voice through the above mentioned adjectives comes from a series of bipolar adjectives, according to the Osgood’s psychological method of semantic differential scales (1952; in Soto Sanfiel 2000: 10, 449ff). Thus, pleasantness was only one of the variables, and not all pleasant voices showed the same level of credibility (Iglesias Fernández 2011: 42). According to Rodríguez Bravo (1989: 260; in Iglesias Fernández 2011: 42), voices considered ‘radiogenic’ are such because perceived as more ‘self-confident’ ones, since they transmit the image of a ‘clever’ speaker. The main feature of a ‘phonaesthetic voice, or credible voice on radio, is the speaker’s timbre, perceived as more ‘harmonic and transparent, more tense and brilliant’ (ibidem). According to Soto Sanfiel (2000:1), the term *phonaesthetic* was created by Fónagy (1983) and used for the project of a “Gramática de la expresión fonoestésica y de los sistemas integrales de percepción de la voz” (Prado Pico

et al. 1988). A credible voice perceived through the radio has a lower fundamental frequency ( $F_0$ ), lower intensity of low formants ( $F_1$ - $F_3$ ) and higher intensity of high formants (Rodríguez Bravo 1989: 249; in *ibidem*). The impressions of attraction and maturity, associated to a lower  $F_0$  and lower energy, contribute to the construction of a phonaesthetic voice (*ibidem*; in *ibidem*). Male radio speakers with a low-pitched voice, and female speakers with medium-to-low-pitched voice are perceived as more pleasant (*ibidem*: 247-251; in *ibidem*). High-pitched voices are tiring to listen to on the radio, especially for a long time (González Conde 2001: 87; in *ibidem*).

De Meo, Pettorino and Vitale (2012) conducted a study on the relationship between foreign accent and credibility, within the perception of Italian as a foreign language (L2). For the perceptive test, they used a corpus of 10 short bizarre-but-true news, that were read in Italian by 5 female speakers in a radio studio; afterwards, the recordings were edited to create 5 realistic radio news magazines, each of them containing all the 10 news, in the same order, but read by different voices, randomly distributed. The choice of bizarre-but-true news was due to the fact that text of the message can influence the credibility judgment (De Meo 2012: 4), while the radio was chosen to ‘hide’ the speaker in order to avoid respondents’ sociocultural prejudices related to the recognition of mother tongues (De Meo, Pettorino, Vitale 2012: 231). Of the 5 reading subjects (aged on average 25) only 1 was a native Italian speaker, the other 4 reading subjects being non-native Italian speakers with two different levels of competence: A2 for the Arabic and Japanese mother tongue speakers, and B1 for the Chinese and Vietnamese mother tongue speakers. The perceptive test was administered to 175 native Italian speakers (80 males and 95 females aged 14 to 60); each of them listened to one of the radio news magazines and then indicated firstly the level of ‘auditory comprehensibility’ (on a 3-point scale), and secondly a true/false judgment. The study was based, among other sources, on Derwing’s and Munro’s (1995, 1997, 1999; in De Meo, Pettorino, Vitale 2012: 233) proposal of speech perception of L2, structured on three interrelated levels: *accent*, *intelligibility* and *comprehensibility*; where *accent* means ‘degree of divergence between an utterance and the expected acoustic model’, *intelligibility* means ‘the real listener’s understanding of an utterance’, and *comprehensibility* means ‘the listener’s subjective evaluation of the degree of difficulty found in decoding the utterance’ (De Meo, Pettorino, Vitale 2012: 232).

The study focused on the acoustic correlates of credibility, in detail the following indices were created and then measured: speed of articulation and speech, fluency, tonal range and silent pauses. Results showed “no evidence of a correlation between degree of accent and credibility” (ibidem: 229); however, it was found that a series of “factors” had an impact on comprehensibility, namely: “disfluency, frequency of silence, pitch range variations, silent pauses and segmental errors” (ibidem). According to the authors, ‘the research confirmed the hypothesis of «processing fluency», according to which the difficulties of the acoustic stimulus induce listeners to judge the test of the message as not very ‘credible’ (ibidem: 243; based on Oppenheimer 2008, Reber and Schwarz 1999, Reber and Unkelbach 2010; in De Meo, Pettorino, Vitale 2012: 230). «Processing fluency» is a ‘cognitive strategy in processing stimuli [...] that would influence the judgment on credibility’. The authors maintained that their study ‘would demonstrate the involvement also of the acoustic-auditory level of language in «perceptual fluency»’ De Meo, Pettorino, Vitale 2012: 244). A more extensive study was conducted by De Meo (2012), including two more voices, more listeners, and 4 different tests, 2 of which related to artificial modifications of fluency and intonation for experimental reasons. Results showed that

the correlation between foreign accent and credibility is delivered by comprehensibility: poor comprehensibility generally lowers the credibility level of an utterance. When comprehensibility is high, a reduced tonal range and longer silent pauses, i.e. the suprasegmental features of the utterance, determine a significant increase of trustworthiness. (De Meo 2012: 3).

The above mentioned studies, based on voice as it was perceived through the radio, were considered relevant to this research that considers voice as perceived through audio-vision, for two main reasons: i) firstly, because results confirm the attention listeners pay to the aspects of the sound matter (of human language, in this case) when the voice is acousmatic, i.e. its source (the speaker, in this case) is invisible, and this is true in radio as in the case of simultaneous interpreter in television; ii) secondly, because, even if the studies mentioned focused on prosodic aspects, these were only *aspects* of the news selected and then modified both from Soto Sanfiel (2000) and De Meo, Pettorino, Vitale (2012). The conclusions of the latter study also considered the involvement of listeners’ systems of expectations and encyclopedic knowledge, that in our opinion should be considered implicit in “comprehension”, “perceptual fluency” and “judgment of trustworthiness”.

## 4.1 Rationale

Interpreting Studies on both quality expectations (ideal evaluation) of simultaneous interpreting (Bühler 1986; Kurz 1989, 1993; Kopczyński 1994; Chiaro and Nocella 2004; Moser 1995; Pöchhacker and Zwischenberger 2010) and quality assessment (evaluation after real experience) of simultaneous interpreting (Gile 1990; Marrone 1993; Vuorikoski 1993; Mack and Cattaruzza 1995; Garzone 2003; Russo 2005; Catana 2005; Collados Aís et al. 2007; García Becerra 2013) show that quality criteria adopted for evaluation have not changed substantially over time. In fact, they present more or less the same “linguistic” criteria devised by Bühler (1986): native accent, pleasant voice, fluency of delivery, *logical cohesion of utterance*, *sense consistency with original message*, *completeness of interpretation*, *correct grammatical usage*, *use of correct terminology*, *use of appropriate style*. These criteria have been adopted in subsequent studies on quality, sometimes with similar names, sometimes with the same criteria grouped into other categories, other times with other criteria adapted to the objective of the study (Soler Caamaño 2006: 101; García Becerra 2013: 55, 74, 84).

Nonetheless, Garzone (2003: 25) observed that “in the actual assessment of real instances of interpretation there might be interferences and interdependence between the different criteria separately submitted to, and evaluated by, respondents”. This conclusion was confirmed by the results of the study by Collados Aís et al. (2007) that showed an interrelation and interdependence among the different criteria and their incidence on the overall quality assessment of SI. García Becerra (2013: 571) observed that “it looks like that insufficient formal aspects could eclipse remaining parameters in the evaluation mechanism of subjects”.

For these reasons, Iglesias Fernández (2013: 59) concluded that “quality criteria do not seem to be processed separately, but holistically, in clusters of features”. Soler Caamaño (2006: 283) proposed that “el estudio de la calidad debe llevarse a cabo desde una perspectiva olística”.

Considering the high ratings assigned to the voice as quality criterion in the survey on quality expectations by Kurz and Pöchhacker (1995) and in the survey on quality assessment of film interpreting by Russo (2005), the methodology used to build a questionnaire to elicit a holistic perception of the quality of a television broadcast simultaneous interpretation has to take into consideration the influence that the medium may have on perception.

## 4.2 Objective

Considering:

i) the importance assigned to the criteria *pleasant voice* and *fluency of delivery* by TV professionals in the quality expectation survey on TV interpreting (Kurz and Pöchhacker 1995 - see section 1.1.1.3.);

ii) the position of the parameter *fluency of delivery* in the “perceptual map” or “interpreter’s image of linguistic criteria” plotted by Chiaro and Nocella (2004), that led the authors to state that “intonation is considered by interpreters [...] on the interface of the two dimensions” considered for the map, i.e. “structural dimension [of criteria]” and “discriminating quality dimension” (Chiaro and Nocella 2004: 290; see section 1.1.1.6.);

iii) the importance assigned to *fluency of delivery* by interpreters in the survey on quality and role by Pöchhacker and Zwischenberger (2010) (see section 1.1.1.4.);

iv) the importance of the criterion *pleasant voice* assigned by film critics, students and others in the survey on quality assessment of film interpreting (Russo 2005), considering that the relative rating almost equaled that of *general quality* (see section 1.1.2.8.);

v) the impact of *monotonous intonation* (manipulated as an experimental variable) in the ratings of: *overall quality* of interpretation; *professionalism* and *credibility* of the interpreter; her/his *ability to captivate the audience*; *voice*; *pleasantness of listening*; and *self-evaluation of comprehension*, resulted from the study by Catana (2005) (see section 1.1.2.9.);

vi) the interdependence of “non-verbal” parameters (voice, fluency, diction, accent) and their interrelation with the “verbal” ones (fidelity, cohesion, terminology, grammar, accuracy) revealed by the study by Collados Aís et al. (2007) (see section 1.1.2.10.);

vii) the impact of “formal aspects” on the formation of the “impression” of quality of an interpretation, detected by García Becerra (213: 567) (see section 1.1.2.11);

viii) the interdependence of prosodic elements that are peculiar to simultaneous interpretation speech: long pauses; short intonation units; final pitch movements marked by rising, level and rise-level contours; frequent accentuation (Ahrens 2005);

ix) the relevance, in television interpreting, of the ‘vocal’ aspect (Mack 1999), or that of ‘teleginicity’ or ‘television usability’, based on ‘form’ (Straniero Sergio 2007: 541);



this pilot study aims at providing a proposal for a gestaltic assessment of television broadcast simultaneous interpreting.

### **4.3 Hypothesis**

Perception (and not analysis) of speech, and consequently, of simultaneously interpreted speech, is based on gestalt laws, where voice plays an important role as a means of expression. The reason why we are interested in gestaltic perception is that the previous approaches (see above – sections 1.1.1 and 1.1.2) are based on analytic perception.

### **4.4 Method**

The survey is questionnaire-based; it includes 3 video excerpts (1 min each) and 3 sets of the same questions for each video excerpt; therefore, each questionnaire is made up of 3 blocks – each block being constituted by 1 video plus 1 set of questions – plus a final block of questions on personal data. Sets of questions are equal except the last 3 questions of each set, which are related to the comprehension of the text of the relative interpretation (video).

Since the corpus ORe nesit (see chapter 2) was not completed when the pilot study had to be carried out, video excerpts were selected from the Italian interpretations of 2008 US Presidential Debates (Obama vs. McCain). In fact, the only changes to the format of televised debates introduced in 2012 by the Commission on Presidential Debates (CPD) was the division of the first and third debates (i.e. the presidential ones, not the vice-presidential one) into 15-minute segments related to specific issues.

In 2000, the CPD held its first debate in which the candidates were seated at a table with the moderator, a format that further encourages candid conversation without the physical separation of podiums. In 2012, the CPD adopted a significantly different format for the first and last presidential debates: those two debates were divided into six 15-minute segments, during each of which the candidates discussed one major issue facing the country. One debate was devoted to domestic issues and one to foreign affairs. The topics for both debates were chosen by the moderators and announced several weeks beforehand (<<http://www.debates.org/index.php?page=overview>>).

In our opinion, such a change in the format would have not affected the reception of a 1-minute video excerpt, for the purpose of the pilot study, which was to test the questionnaire. Another important aspect that was not

thought to have a great impact on reception of the video excerpts was the social and political context, given the four year between the 2008 and 2012 US Presidential Debates.

## 4.5 Materials

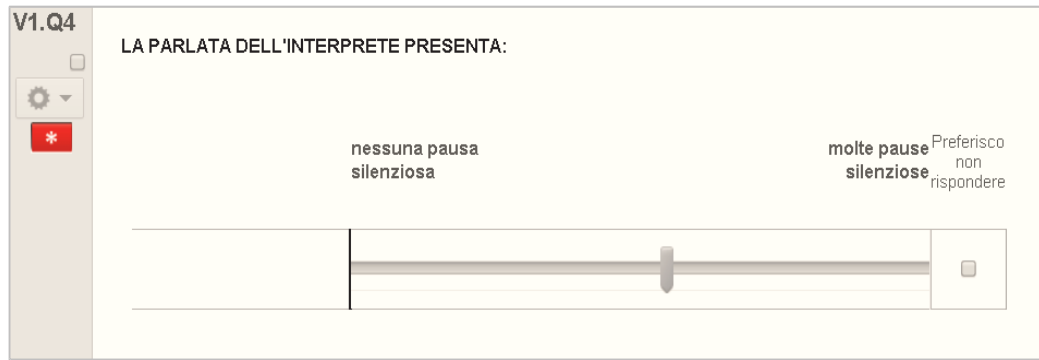
### 4.5.1 Questionnaire

The questionnaire for the pilot study was designed according to the linguistic approach described above (see chapter 3) and following the methods of social research (Bailey 1995:103-207; Gillham 2000: 1-45, 395-424; Various authors 2006: 1-57). It was administered *in praesentia* to BA and MA interpreting and translation students at the University of Trieste; in total, 101 subjects filled in the questionnaire.

The “questionnaire flow” (Various authors 2006: 47-48; Bailey 1995: 163-170) was built in such a way to help respondents to move from sound perception to sense construction, and the self-assessment of comprehension; it ends with a test on comprehension.

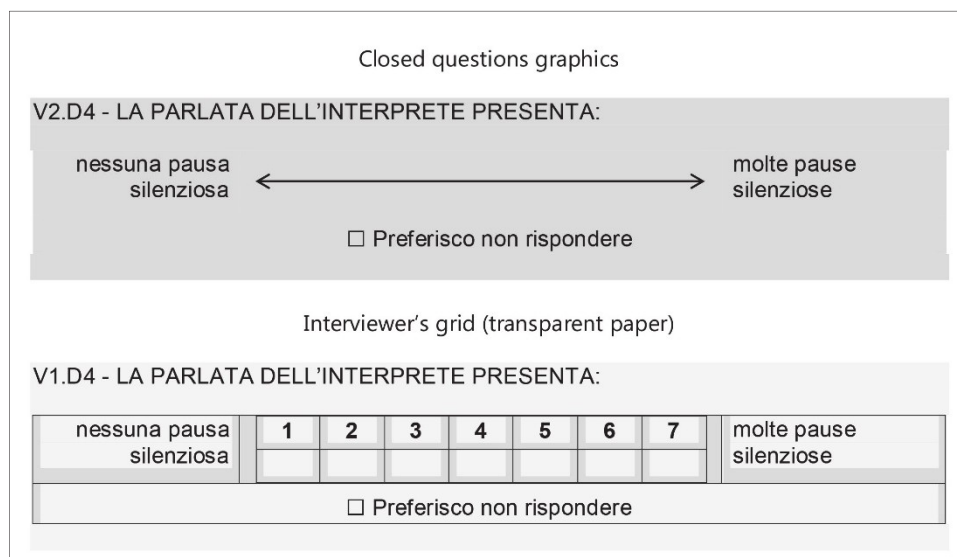
As to “questions formats” (Various authors 2006: 35-43), the majority of questions are “closed” questions. Questions on comprehension are “multiple choice”; those on possible features not considered in the questions and those on comments on the questionnaire are “open” questions, while the final questions on personal data are all “factual” question. All the closed questions are “scaled-response” questions, with “numeric (endpoint-labelled) scales” (ibidem: 40-41).

In its last stage, the questionnaire was developed on a web-based platform, since it was to be administered on-line. The web platform Qualtrics (® 2015) allowed using an evaluation pattern that is a numeric endpoint-labelled scale, but the scale was hidden to the respondent; thus the graphic solution resulted in a cursor initially positioned in the middle of an intensity bar that could be moved by the subjects leftwards or rightwards, towards the two labelled endpoints, at the ends of the bar. The hidden scale was 1 to 7 points; all the ‘negative’ labels of the questions were all positioned on the left label, corresponding to point 1 of the scale.



**Figure 4.2.** Example of the response pattern of one of the closed questions from the web Platform Qualtrics (® 2015) – screenshot reduction.

Finally, due to legal and technical reasons related to the video excerpts, the questionnaire was administered *in praesentia*, and no longer through the on-line platform. Nevertheless, the graphic trick of the hidden scale was transferred to paper; thus, respondents were asked to indicate their choice with a sign on a segment or intensity line with the labels as endpoints, the level on the 1 to 7 points scale was subsequently found by the researcher after measuring it with the a scale printed on transparent paper having the same length of the segment (see below – Figure 4.3).



**Figure 4.3.** Example (top) of the response pattern of the closed questions with the intensity bar where to indicate the of intensity of perception with a cross; and example (bottom) of grid used by the researcher to quantify the perception indicated by the respondent.

The Likert scale was ruled out, because it has a semantic anchorage; it is a semantic scale for linked to a numeric scale, which is used to quantify the results. The main problem of this conversion is that the semantic distance between an element of the scale and the preceding/following one is generally measured with one number of the numeric scale; the problem is, on what bases can one establish that the difference (distance) between «strongly disagree» and «do not agree» is «1 / 2 / 3 / ...»? How can it be quantified the distance among different semantic labels for a same response? (Delli Zotti 2014, personal communication).

I valori che si sommano sono numeri naturali (1,2,3,4,5) che si attribuiscono ai cinque livelli di (dis)accordo; questi si sommano per assegnare a ciascun oggetto un punteggio totale sulla scala, che esprime il suo (supposto) stato sulla proprietà generale che la scala pretende di misurare. Ma non si possono sommare numeri ordinali (1 + 2 fa 3 ma primo più secondo non fa terzo); se si sommano, si stanno usando come numeri cardinali le etichette numeriche attribuite a semplici espressioni di disaccordo.

Impiegare tecniche cardinali con punteggi solo ordinali non produrrebbe gravi distorsioni se – e solo se – fosse ragionevole immaginare che ai soggetti queste cinque espressioni di (dis)accordo appaiano collocate in posizioni approssimativamente equidistanti lungo il continuum (Marradi 2007: 146-147).

On the contrary, the end-point labelled numeric scale is a better evaluation scheme, since the respondent quantifies his response on his own, without the interference of semantic labels. In this case, the researcher decided to hide the numeric scale, with the aim of eliminating the interference of the habit of assigning a mark to one's response, and therefore, to reduce the influence of numbers; this is an application of the “semantic differential” introduced by Osgood (1952; in Marradi 2007: 158ff) and still used in psychology (Marradi 2007: 158). The intensity bar only shows a horizontal line to the respondent, who is forced to focus only on the recall and judgment of his/her own perception of the phenomenon indicated in the question and limited by the two labels positioned at the two ends of the bar to represent the poles (-/+ ) of the intensity. This evaluation scheme can be seen more and more often in surveys on websites, even without endpoint labels, only with the question; this format is also easily accessible from smartphones and tablets.

The web based platform Qualtrics also allowed a randomized administration of the 3 blocks linked to the video excerpts, in order to limit the “response set” phenomenon (Marradi 2007: 117-119), i.e. the tendency of a respondent to give similar or equal responses to the questions of the second

and third block, because s/he is influenced by the responses already given in the first block, especially when the questions in two or more blocks are the same and share the same response pattern. In the *in praesentia* administration, the randomization was created by distributing the possible combinations of the 3 videos (in sequence) in groups of subjects of equal number as far as possible (see below – section 4.7.1, table 4.5).

Pilot Questionnaire		RANDOMIZATION
Block 1	Video 1 Original	
	Questions	
Block 2	Video 2 <u>Manipulated</u>	
	Questions	
Block 3	Video 3 Original	
	Questions	
Block 4	Personal data	
	Free comment	

**Figure 4.4.** Structure of the questionnaire for the pilot study.

#### *4.5.1.1 Process of construction of the questionnaire based on the development of the theoretical framework*

The process of wording and sequencing of the question has been divided into seven stages, corresponding to the drafts of the questionnaire. The third draft was used for a pre-test, administered to 7 students and one musician.

##### *4.5.1.1.1 First draft*

On the basis of the questionnaires used in previous surveys on quality assessment and/or expectations of quality in simultaneous interpreting (SI), a new questionnaire for a subjective assessment of TV broadcast interpreting was designed. Evaluation parameters were arranged in the questionnaire flow (Bailey 1995: 103-207; Gillham 2000: 1-45, 395-424; Various authors 2006: 1-57) in such a way as to help respondents to move from phonic perception to sense construction. This aspect of the sequence of questions remained constant over the various versions in the process of questionnaire building.

The first part of the questionnaire, which remained unchanged in the following version, was related to the socio-demographic (anonymous) data of the respondent (age, gender, profession, education). In its first draft the body of the questionnaire, related to the assessment of SI quality, was mainly composed of closed questions with a five-point Likert scale as evaluation pattern (not at all – slightly – moderately – very – extremely) (see Appendix 2). The questionnaire flow included the following criteria (items): *pleasant voice*; *clear pronunciation*; *speed of speech* and *intonation functional to comprehension*; *telegenic voice*; *use of words adequate to the discourse* (register); *correct phrasing and sentence construction* (cohesion); *all the above aspects assessed together*; *rhythm* (functional to the clarity of expression); *self-assessment of comprehension* on the basis of clarity of expression; *self-assessment of comprehension* according to rhythm; test on *real comprehension of text*. The question *all previous aspects assessed together* had the function of eliciting the gestalt perception of speech (see above – section 3.1); the subsequent question on *rhythm* functioned as a control question for the previous one. The questions on *self-assessment of comprehension* on the basis of both expression and rhythm also functioned as control questions for the previous ones on expression and rhythm. The aim of all these questions on expression, rhythm and comprehension was to test, in perception, the notion of rhythm as form (= *Gestalt*) of language (Meschonnic, 1982:70).

Starting from the first draft, particular attention was paid to the wording and phrasing of questions, the ideal respondents being ordinary television viewers with varying levels of education. Consequently, the main difficulties in wording and phrasing were related to the translation of linguistic terms into common-language words and phrases.

#### 4.5.1.1.2 Second draft

In the second draft, the number of items increased to 19 questions, in order to elicit more specific features of simultaneous interpreting speech. This draft was based not only on the previous surveys on quality assessment and/or expectations of SI, but also on studies focused on the peculiar prosody of SI speech (Shlesinger 1994; Tissi 2000; Ahrens 2005; Tohyama and Matsubara 2006). The first item was a multiple-choice question on the voice of the interpreter, to be classified according the following attributes: *enchanting*, *pleasant*, *lovely*, *sweet*, *clear*, *boring*, *plaintive*, *irritating*, *monotonous*, *other*. This classification of voice was taken from a questionnaire for the evaluation of pathologic voice, created to consider both the objective acoustic measure and

the perceptual measure of voice quality in hypoacoustic children (Alfano; Marciano; Riccardi 2007). However, this question was omitted in subsequent drafts: firstly because the voice to be evaluated was not pathologic; secondly because the evaluation under consideration could not be based on an acoustic measure: as a matter of fact, in the audiovisual recordings of *ORenesit* it is not possible to detach the interpreter's voice (in the foreground, with a high volume) from the original speaker's voice (in the background, with a low volume), since both are merged in a single monophonic audio track. The other items of the second draft were: *comprehensibility of voice* (multiple choice: very clear, clear, not clear); *articulation*; *speed of speech*; *use of stressed words*; *use of a melodic speech*; *fluency*; *vowel and consonant lengthenings and other voiced hesitations*; *interruptions* (false starts, repetitions, restructurings); *use of words appropriate to the discourse*; *presence of technical terms*; *use of correct sentences*; *use of complex sentences*; *all previous aspects assessed together*; *overall clear expression by the interpreter*; *self-assessment of text comprehension*; *knowledge of the information given in text*. With the exception of the first and second questions, the evaluation pattern in the second draft was an eight-point numeric scale (0-7); all the questions were declarative sentences and the respondent was asked to indicate if s/he totally agreed (7) or disagreed (0) (see Appendix 3).

#### 4.5.1.1.3 Third draft (pre-test)

The third draft was used for a pre-test with 6 MA interpreting students and 1 musician, to prove the usability of the questionnaire and the smoothness of the questionnaire flow. The questions of the second draft (19) were slightly changed in their wording to include linguistic terminology in order to adapt it to the respondents, who in this case were supposed to know (or at least grasp) terms like *word stress*, *melodic curve*, *tone*, *lexical density*, *voiced hesitations* (*false starts*, *restructurings*, *repetitions*), *vowel and consonant lengthenings*, etc. Also the evaluation pattern of questions was the same as in the second draft, i.e. an eight-point numeric scale (0-7) through which respondents had to express their (dis)agreement to positive questions (see Appendix 4). Respondents were asked to assess three video excerpts (approximately 3 min each) of the 2008 US presidential debates, simultaneously interpreted from English into Italian. The average time of response was 12 min. The aggregate average of responses for the three video excerpts showed a maximum difference of 2.12 points between the highest and the lowest ratings, respectively 5.62 (*self-assessment of comprehension*) and 3.50 (*tone variation*) of 2.12. The rating obtained by the question related to gestalt perception was 4.95; the closest figure to it was the one

related to the question on pronunciation (4.80), followed by the interpreter's expression (5.12) and the interpreter's hesitations (4.68), while the overall aggregate average (all videos and all questions) was 4.54, lower than the aggregate average of the musician's assessment (5.38).

#### 4.5.1.1.4 Fourth and fifth draft

With respect of the third draft, in the fourth draft only two more questions were added, relating to the possible detection of singsong in the interpreter's speech. In Italian diction and elocution, the equivalent word for singsong, i.e. *cantilena*, may mean: (i) an unpleasant repetition of a monotonous voice; (ii) the repetition of non-standard consonant groups; (iii) the presence of regional accent in the speech (Veneziano 2001: 53-55). Therefore, a general question on the presence of singsong was added, followed by a more definite question on the presence of regional accent, in order to disambiguate the interpretation of the term by respondents. In addition, this draft presented a new evaluation pattern, which was no longer a numeric rating scale, but an end-point labelled graded scale, where the labels for all the questions were *totally agree/ totally disagree*. The graded scale allowed a lower degree of influence on subjects, since numbers were not present, but were replaced by empty spaces (14 in this case, to represent one decimal of a seven-point scale).

The fifth draft presented a step forward to reduce the circularity of questions; declarative questions of previous drafts, sometimes with positive attributes (e.g. *high* speed of speech; *pleasant* tone variation; etc.), were considered to have an excessive influence on respondents. For this reason, questions were rendered more neutral in their wording, by eliminating positive and negative attributes and by moving the *comment* part of the information structure (topic/comment) of the utterance from the sentence itself to the labels of the scale (e.g. *From the point of view of information the speech is: dense* | \_ | \_ | \_ | \_ | \_ | \_ | \_ | \_ | \_ | \_ | \_ | \_ | \_ | *not dense*) (see Appendix 5).

#### 4.5.1.1.5 Sixth draft and on-line version

The sixth draft and subsequent versions of the questionnaire were developed on the web platform Qualtrics (© 2015), due to our intention to administer the questionnaire through the web. This tool allowed to preserve the seven-point numeric graded scale of the evaluation pattern in the fifth draft; however, both grades and numbers were hidden to respondents, who could only view a segment with a cursor. This form was considered optimal



for respondents to indicate the intensity of perception related to the parameter to be assessed; thus, the evaluation pattern was represented by the following scheme: (label 1) - ←[draggable slider]→ + (label 2). In the on-line version, the questionnaire flow was based on the same principle used starting from the first draft (see above – section 4.5.1.1.1.), but items (parameters) were revised according to the theoretical paradigm purposely built up to elicit a gestalt perception of simultaneous interpreting quality, applied to a subjective assessment of TV broadcast interpretation. Thus, the last version of the questionnaire was based on: (i) previous surveys on quality assessment and/or expectations of SI; (ii) studies focused on the peculiar prosody of SI speech (see above – section 1.4); (iii) the categories extracted from the main sources of the above-mentioned theoretical framework: *Il volto fonico delle parole* (Albano Leoni, 2009); *L'Audio-vision* (Chion, 1990); *La vive voix* (Fónagy, 1983) (see above – section 3.8.1 and figure 3.1 for the structure of the questionnaire; see Appendix 6 for the questionnaire as it was administered *in praesentia*).

#### 4.5.1.2 Results from pre-test

The main purpose of the pre-test (third draft) was to check *in praesentia* the usability of the questionnaire, mainly the time of response, possible problems related to the structure of questions and the evaluation pattern. It was administered to 6 MA interpreting students (four females and two males) plus 1 musician (female). The subjects watched three video excerpts of the Italian version (SI) of the 2008 US Presidential debates; after viewing each excerpt, the respondents answered the same series of questions (19) (see Appendix 4 for the questionnaire). Each excerpt lasted around three minutes and the response time was 15 to 20 minutes.

The video excerpts were selected on the perceptual basis of “usability” of the interpretation (Viezzi 1996: 79-105; see also above – section 1.3), according to the researcher’s perception. Video 1 was judged to have a medium level of usability, video 2 the lowest level and video 3 the highest level; the interpretation of video 3 was executed by a female voice, while the interpretations in videos 1 and 2 were performed by a male voice. Since the interpretations were (and still are) televised in voice-over modality, the original version can be listened to at a lower volume with respect to the interpreted version, but only in part, therefore the assessment only concerns the interpreted version, and not a comparison between the original and the

interpretation - in the end, the target text is the only one that the TV format makes fully intelligible.

Process of construction of the questionnaire flow			
Draft 1	Draft 2	Draft 3 (pre-test) (0-7 point rating scale)	Last version
<ol style="list-style-type: none"> <li>1. Pleasant voice</li> <li>2. Clear pronunciation</li> <li>3. Speed of speech</li> <li>4. Intonation</li> <li>5. Telegenic voice</li> <li>6. Words</li> <li>7. Phrasing and sentence construction</li> <li>8. All previous aspects assessed together</li> <li>9. Rhythm</li> <li>10. Self-assessment of comprehension on the basis of clarity of expression</li> <li>11. Self-assessment of comprehension according to rhythm</li> <li>12. real comprehension of text</li> </ol>	<ol style="list-style-type: none"> <li>1. Type of voice</li> <li>2. Comprehensible voice</li> <li>3. Articulation</li> <li>4. Speed of speech</li> <li>5. Word stressing</li> <li>6. Melodic speech</li> <li>7. Fluency</li> <li>8. Vowel/consonant lengthening and other voiced hesitations</li> <li>9. Interruptions (false starts, restructurings, repetitions)</li> <li>10. Use of adequate words</li> <li>11. Presence of technical terms</li> <li>12. Use of correct sentences</li> <li>13. Use of complex sentences</li> <li>14. All previous aspects assessed together</li> <li>15. Clear expression</li> <li>18. self-assessment of text comprehension</li> <li>19. knowledge of the information given in the text</li> </ol>	<ol style="list-style-type: none"> <li>1. Male/female voice</li> <li>2. Type of voice</li> <li>3. Comprehensible voice</li> <li>4. Intelligible voice</li> <li>5. Pronunciation</li> <li>6. High speed of speech</li> <li>7. Word stressing</li> <li>8. Melodic curve with pleasant tone variation</li> <li>9. Fluency</li> <li>10. Filled pauses (vowel/consonant lengthening and other voiced hesitations)</li> <li>11. Interruptions (false starts, restructurings, repetitions)</li> <li>12. Lexicon relevant to discourse</li> <li>13. High lexical density and sentences</li> <li>14. Correct phrases and sentences</li> <li>15. Use of complex sentences</li> <li>16. All previous aspects assessed as one</li> <li>17. Clear expression</li> <li>18. Self-assessment of text comprehension</li> <li>19. knowledge of information given in the text</li> </ol>	<ol style="list-style-type: none"> <li>1. Articulation</li> <li>2. Hesitations</li> <li>3. Audible breaths</li> <li>4. Silent pauses</li> <li>5. Speed of speech</li> <li>6. Melodious /monotonous voice</li> <li>7. Sung speech</li> <li>8. Aggressive/tender voice</li> <li>9. Self-confident /insecure voice</li> <li>10. Active/self-defeating behaviour of the interpreter</li> <li>11. Expressivity of voice</li> <li>12. Comprehensibility of voice</li> <li>13. Synchronicity of voice with respect to the image</li> <li>14. Informativity of voice with respect to the image</li> <li>15. Skilled/unskilled interpreter</li> <li>16. Complex/simple words</li> <li>17. Complex /simple syntax</li> <li>18. Natural/non-natural syntax</li> <li>19. Self-assessment of text comprehension</li> <li>20. real comprehension of text (test)</li> </ol>

**Table 4.1.** Stages of the process of construction of the questionnaire flow for the pilot study.

The very small number of respondents, the fact that female and male voices cannot be deemed comparable due to physical differences are sufficient reasons not to consider an ‘analysis’ of the results but just ‘impressions’ from the results.

A comparison of the frequencies of ratings (0-7 point scale) assigned to the type of voice (question 2) showed that the voice in video 3 was the most appreciated, since it was defined by all the 7 respondents with positive

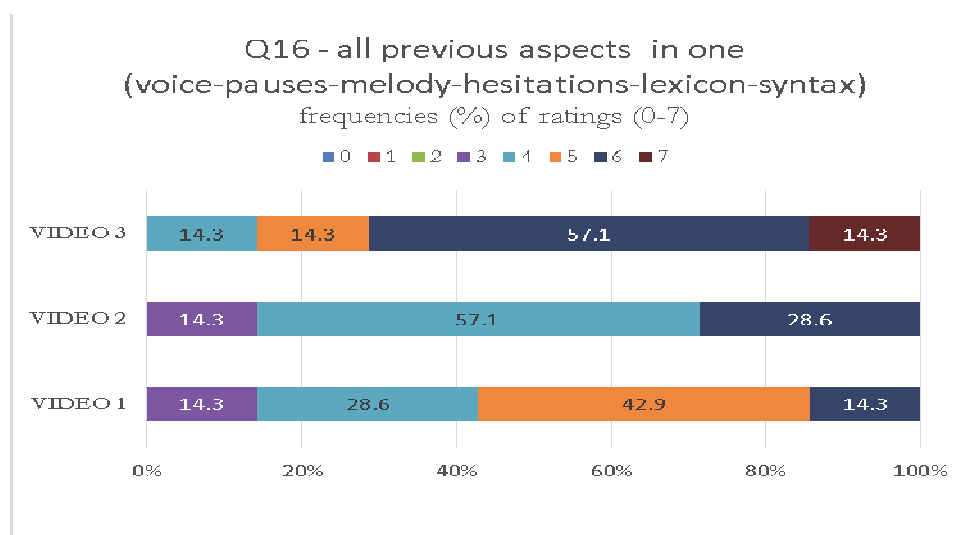
adjectives: *clear* (1 respondent); *pleasant* (2 respondents); *pleasant, clear* (2); *pleasant, clear* and *natural* (1). The second most appreciated type of voice was the one related to the video 1, which was considered: *clear* by 2 respondents; *pleasant* and *clear* by 1; *pleasant, clear, monotonous* and *self-confident* by 1; *monotonous* by 1; *monotonous* and *unnatural* by 1; *boring* by 1. The type of voice related to the video 2 was the less appreciated one, since it was judged to be *pleasant* and *clear* just by 1 respondent, while others respondent considered it: *monotonous, calm* and *without emotion* (1); *monotonous* (1); *irritating* (1); *boring* (1); *between sweet and clear, between grumbling and irritating, insecure* (1).

V1-Q2-type o fvoice (pleasant/lovable/sweet/ clear/boring/other)	Frequency VIDEO 1		V2-Q2-type o fvoice (pleasant/lovable/sweet/ clear/boring/other)	Frequency VIDEO 2		V3-Q2-type o fvoice (pleasant/lovable/sweet/ clear/boring/other)	Frequency VIDEO 3	
	Subjects	%		Subjects	%		Subjects	%
Boring.	1	14.3	No response	1	14.3	Clear.	1	14.3
Clear.	2	28.6	Between sweet and clear; between grumbling and irritating. Insecure.	1	14.3	Pleasant.	2	28.6
Monotonous	1	14.3	Boring.	1	14.3	Pleasant.	1	14.3
Monotonous, unnatural.	1	14.3	Irritating.	1	14.3	Pleasant, clear.	1	14.3
Pleasant, clear.	1	14.3	Monotonous.	1	14.3	Pleasant, clear.	1	14.3
Pleasant, clear, monotonous. Self- confident.	1	14.3	Monotonous. Calm, without emotion.	1	14.3	Pleasant, clear. Natural.	1	14.3
Total	7	100.0	Pleasant, clear.	1	14.3	Total	7	100.0
			Total	7	100.0			

**Table 4.2.** Frequencies of ratings in assessment of Question 2 (type of voice) in pre-test (0-7 point scale).

These results showed that even with a small number of respondents it is possible to elicit a pleasant voice when the subject has the possibility to evaluate it through a number of labels. The results were consistent with those from the question on the *comprehensibility of voice* (Q3) and the following (control) question on the *overall intelligibility of voice* (Q4), since in both the voice of video 3 obtained the highest frequency of the highest ratings, followed by the voices in videos 1 and 2 (see Appendix 4). The question on *clear pronunciation* (Q5) obtained the same pattern of responses, while the question on the *overall high speed of speech* (Q6) obtained a different pattern, since the speech judged to be the fastest was that in video 1, followed by that in videos 3 and 2. Questions on *quantity of words stressed* (Q7) and *melodic curve with pleasant tone variation* (Q8) confirm the response pattern of the voice in video 1 as the

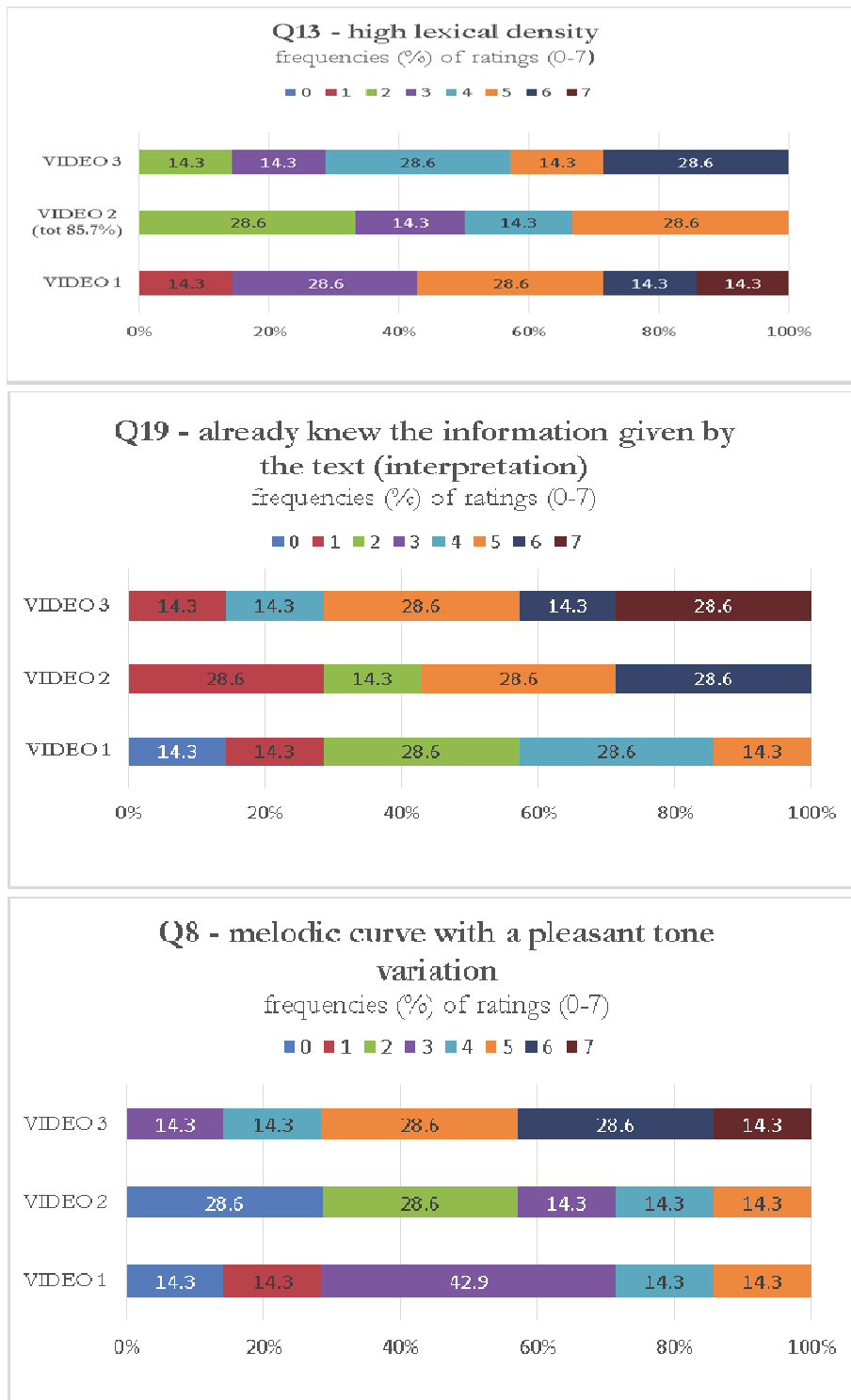
most appreciated, followed by the voices in video 2 and video 3. Questions related to fluency showed an apparently inconsistent response pattern, since the results of question on *fluent speech* (average speed of speech) (Q9) and those of question on *filled pauses* (*vowel an consonant lengthening, voiced hesitations*) (Q10) confirmed the primacy in video 1, but this was followed by video 2 and then video 3, while the results of the question on *interruptions* (*false starts, repetitions, rephrasing*) (Q11) showed the most common pattern, i.e. video 3 followed by video 1 and video 2. According to this few data, the aspect that most influenced the judgment on fluency was *filled pauses*, and not *interruptions*. The *lexicon* judged to be most *relevant to discourse* (Q12) was that related to video 1, followed by those in videos 3 and 2; same response pattern for the *high lexical density* (Q13). The interpretation judged to have the highest *number of grammatically correct sentences* (Q14) was that in video 3, followed by those in video 2 and video 3. The speech judged to have the highest *number of complex sentences* (Q15) was video 1, followed by video 3 and video 2. The results of the question on the assessment of *all the previous aspects as one* (*voice-melody-pauses-hesitations-lexicon-syntax*) showed that the interpretation in video 3 was the most appreciated, followed by video 1 and video 2.



**Figure 4.5.** Compared frequencies of ratings of question 16 of pre-test questionnaire, thought to elicit perception as a unique form (Gestalt) of SI speech.

Also the results from the questions on the *overall clarity of* [the interpreter's] *expression* (Q17) and *overall comprehension* of the respondent (Q18) showed the same response pattern: video 3, followed by videos 1 and 2. While the question on the subjects' *knowledge of the information given by the interpretation*

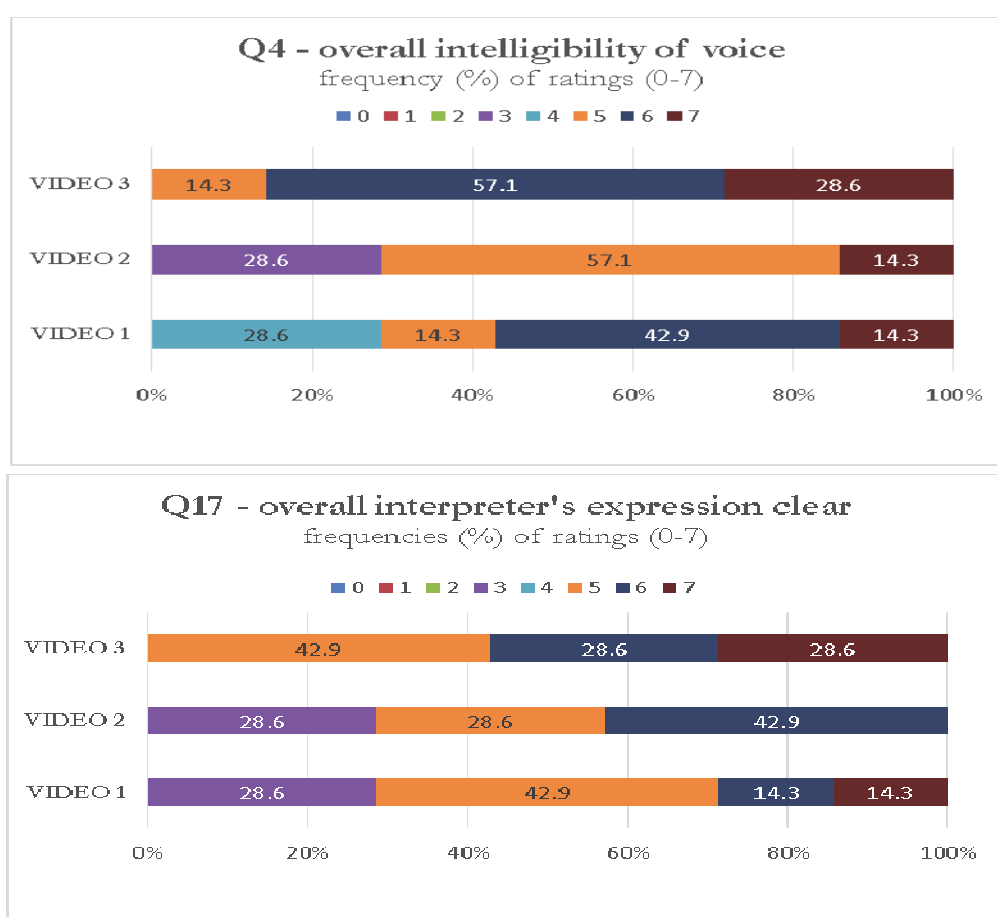
(Q19) obtained results according to which the most known information was the one related to video 3, followed by the one in video 2 and video 3.



**Figure 4.6.** The three questions of the pre-test that obtained the widest range of different ratings (4 to 5 out of 8, in a 0-7 point scale) in the three videos.

According to the frequency (in percentage) of ratings (0-7 point scale), the questions that obtained the widest range of ratings assigned by respondents, i.e. four to five different ratings out of eight points, in two out of three videos, were: i) *melodic curve with pleasant tone variation* (Q8); ii) *high lexical density* (Q13); iii) *knowledge of the information given by the interpretation* (Q19) (see Figure 4.6 below). These questions were immediately followed by those on the *overall speed of speech* (Q4) and on the *filled pauses* (Q10), whose results showed, respectively, a variety of 5 and 3 different ratings, and of 4 and 5 ratings (see appendix 7).

On the contrary, the questions that obtained the narrowest range of ratings; i.e. three to four different ratings out of eight in two out of three videos were: *overall intelligibility of voice* (Q4); and *overall clarity of interpreter's expression* (Q17) (see below – Figure 4.7) .



**Figure 4.7.** The two question of the pre-test that obtained the narrowest range of different ratings (3 to 4 out of 8, in a 0-7 scale) in the three videos.

These are but few attempts to find common response patterns in the pre-test, a survey with only 7 subjects, mainly administered to check the time and fluency of response, and the comprehensibility of questions. As mentioned before, some questions had a wrong wording and phrasing, it was the case of questions like “overall high speed of speech” (Q6) or “high lexical density” (Q13), which both contain a positive element (“high”), instead of being neutral. This is also true for the question on “melodic curve with pleasant tone variation” (Q8) that included a positive word (“pleasant”). Since the assertive mode of all the questions may influence the assessment slightly, in the pilot study the positive element of the questions were moved to the right-end of the scale, as a label, while its opposite negative element was introduced and placed at the left-end, to create a polarity in each close question (see section 4.5.1.1.1.3). In addition, the scale was changed as well, since a 1-7 point scale was introduced.

## **4.5.2 Video excerpts of simultaneous interpretations for the pilot survey**

### *4.5.2.1 Selection of three video excerpts*

Since the corpus described above (see above – chapter 2) was not yet fully transcribed when the pilot study was carried out, video excerpts were selected from the Italian interpretations of 2008 US Presidential Debates (Obama vs. McCain). It is worth pointing out that there are no notable differences in the “format” between the 2008 and 2012 Debates; while the differences among the “topics” dealt with in each debate are more important. In fact, the first 2008 debate was based on “foreign policy and national security”, while the first debate in 2012 was dedicated to “domestic policy”; moreover, the third debate of 2008 was dedicated to “economy and domestic policy”, while the “topic” of the third debate of 2012 was “foreign policy” (cf. Commission of Presidential Debate > Debate History > 2008 Debates vs. 2013 Debates: on-line [www.debates.org](http://www.debates.org)).

The main criterion for the selection of the three video excerpts was the classification in three categories of difficulty: low, medium and high. The problem was finding three videos that fitted each one into one of these categories. Initially, some criteria of perceptual analysis, namely lexicon, syntax, intonation, disfluencies, articulation were defined; each had to be rated as low, medium and high difficulty.

The audio-video recording of the Italian interpretation of the third 2008 US Presidential Debate had already been split into 5-minute clips, in order to be transcribed through *WinPitch*, as required by the standards of the *CorIT* (see chapter 2). Thus, a grid with the above mentioned criteria for each clip was built and filled in with the corresponding ratings (see Appendix 8).

Since the very beginning of the analysis of the clips through the grid, it was clear that it would not have been easy to assign the ratings, because of the interrelations of the criteria; also considering that these could not be assigned to a single clip, but had to be assigned to the speakers, namely the two candidates (Obama and McCain) and the moderator. For this reason, the names of the speakers were added next to each clip. For the same reason, it was decided that each video excerpt to be used for the pilot survey would contain only one speaker, and this could not be the moderator, since s/he was not the protagonist in that context.

Moreover, at the beginning, the topics of the debate, were not considered. As can be seen from the grid, in many cases, it was not easy to rate the criteria, since in several cases both rates of “low” and “high difficulty” were assigned to the same criterion. In the end, the selection of the excerpts did not depend merely on the single criteria, nor on the sum of these, but on the interrelation between them. For the same reason, the grid was not completed, and the decision as to which video clip (or part of it) was of low, medium or high difficulty was made on the basis of a total perception of the interpretation. In this case, the criterion of “usability” of the interpretation (Viezzi 1996: 100-105; see above – section 1.3), according to the researcher’s perception, could also be applied, to justify the method of selection.

Initially, the duration of each video excerpt was of 3 minutes, but finally this duration was reduced to 1 minute, for three reasons: i) the videos were submitted to 2 professors of simultaneous interpreting, and both stated that 3 minutes was too long a duration for a perceptual analysis, one of them said that 1 minute would have been enough; ii) considering that respondents would need time and energy to respond to a series of questions three times, the duration of 1 min for each video excerpt would have permitted an estimated total time of 15-20 minutes, an optimal time to keep the attention high during the whole process; iii) considering the context of reception of the video excerpts, i.e. the television watching, it was imagined that one minute was enough for an ordinary television viewer to decide if to keep audio-



viewing the interpretation of US Presidential Debate, mainly thanks to the interpreter's performance, or to zap to another channel.

#### 4.5.2.2 *Experimental variable on video 2*

According to the objective and the hypothesis of this pilot survey (see above – section 4.3), in order to isolate the vocal aspect related to the audiovisual reception of television interpreting, an experimental variable was introduced. The second video excerpt selected for the survey was artificially modified as follows: the original Italian interpretation was replaced with an imitation, executed by an actor and dubber while reading the original transcript and listening to the original English speech. The execution was recorded in a studio and this sound track was then mixed with the sound track of the original speaker and then edited with the original images (for further details about the execution see below – section 4.5.2.3.5). This experimental variable was different from the simple reading of the transcript, used in other similar surveys (e.g. Christodoulides and Lenglet 2014), and from the manipulation of fluency and intonation using software programs for the analysis of speech, also used in other similar studies (e.g. Holub 2010; Rennert 2010). In this way, the dubber, with acting training and expertise in audiovisual products, created his own natural (and human) speech – an imitation of a simultaneous interpretation speech – without focusing on one or more aspects (pauses, hesitations, intonation, accentuation, etc.).

This interest for a natural human sound of speech was related to the ecological validity of the experiment. For the purpose of ecological validity consistent with the aim of the research, it was decided not to isolate one aspect of speech, but to consider all of them without a specific alteration; so they were considered in function of a speaker with a professional telegenic voice, not in function of a simultaneous interpreter.

Nevertheless, a change in the natural rhythm of the dubber's speech had to be made, in order to artificially create *décalage*, and it was done by lengthening and cutting the silent pauses. This was necessary to make the experimental variable as credible as the original video excerpt. It was executed through the software program Audacity®, where both the English track of the original speaker (Obama) and that of the imitation were imported. The English track was extracted from the digital recording of the original debate as broadcast by the CBS, while the imitation track, as stated before, was recorded in studio in stereo mode. Since it was recorded in a sound studio, this

recording had a high definition, higher than that of the original Italian interpretation, as broadcast by the Italian state-owned channel Rainews24, in monophonic mode. Thus, the imitation track was changed into monophonic mode through Audacity. Afterwards, through Adobe Premiere Pro CS6 program<sup>2</sup> some audio filters were added to render the sound less clear and less defined at perception.

Once the two tracks were imported into Audacity, a first short *décalage* at the beginning was created by sliding the track of the imitation rightwards. In addition, the volume of both tracks was adjusted in order to artificially create the same difference in intensity between the high volume of the Italian interpretation and the low volume of the original English speech.

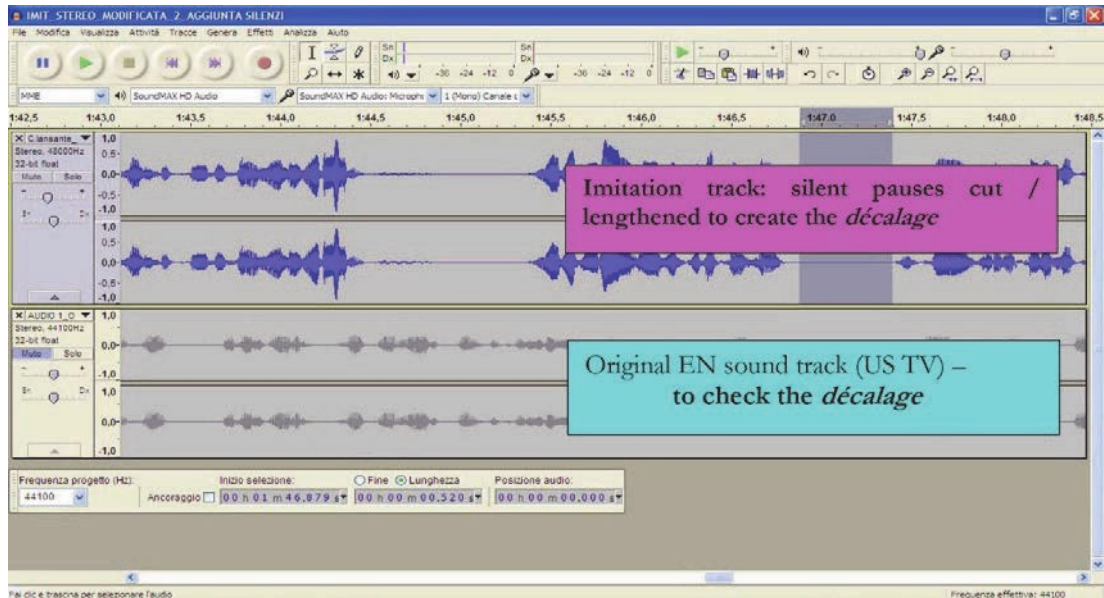
Afterwards, the artificial creation of *décalage* was developed in the imitation track, through the functions of ‘pause creation’ and ‘deleting’ of the software. However, audible breaths were maintained, in order to preserve the biological mark of ~~the~~ natural speech, and so improving the ecological validity of the manipulation. It goes without saying that for the creation of artificial *décalage* it was not possible to intervene in the track of the English speaker, because this track was linked to the image through the lip synchronization, as the speaker was on-screen, while the interpreter was off-screen.

Once the *décalage* was created, the two audio tracks were merged into one and exported in a single audio file (.wav). This audio file was later imported into Adobe Premiere Pro CS3, where the audiovisual file (.avi) of the Italian interpretation had already been imported too. The editing process was developed by synchronizing the new sound track created (imitation plus the original English) with the sequence of images from the Italian broadcaster.

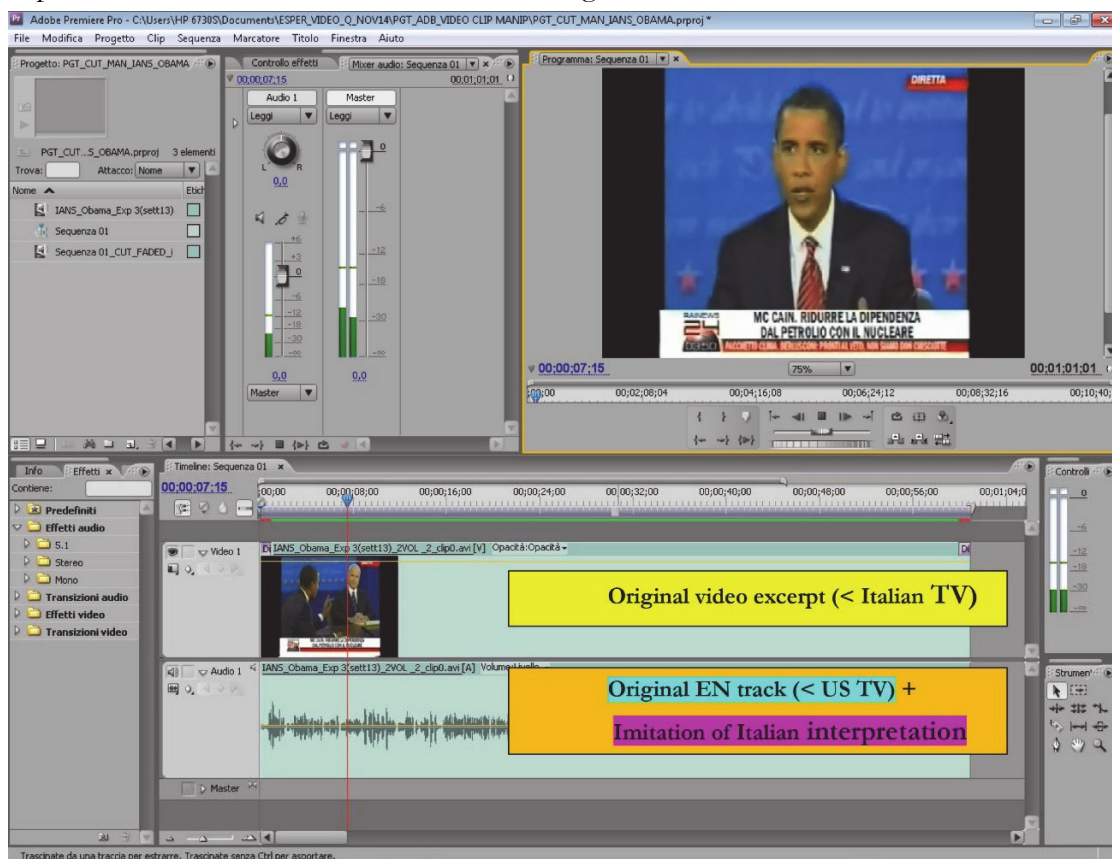
It was fundamental to preserve the images from Rainews24, in order to fool the Italian respondents, making them believe that the imitation was the actual interpretation, and the on-screen graphic of the Italian channel would play a fundamental role in helping the subjects mistake the experimental video for an original one. As for the video editing process on Adobe Premiere Pro CS3, the last step was to delete the original audio track of the Italian broadcasting, no longer needed for the synchronization, and the result was the new sound track (imitation-with-*décalage* merged with the original English) edited with the video excerpt from the Italian broadcaster. All this work was done on a 3 min video excerpt, which was finally cut to get a 1 min excerpt.

---

<sup>2</sup> Collaboration with José Manuel Framit Campaña from the *Centro de Instrumentación Científica – Tratamiento de la imagen* (Universidad de Granada).



**Figure 4.8.** Screenshot from Audacity® reporting an image of the activity of cutting / lengthening of silent pauses, developed - through two functions of the program - on the imitation track, but using the English speaker’s track to check the modifications. At the end of the process the two tracks were merged into a single track to be later imported in Adobe Premier Pro CS3 for editing.



**Figure 4.9.** Screenshot from Adobe Premiere Pro CS3 showing an image of the last part of the editing session, when, after the sound-image synchronization between the original Italian video and the audio track with the imitation and the original English speech, the original soundtrack had been deleted.

<b>WinPitch Transcript of VIDEO 1</b>
<p>(RESPIRO =0.798) è vero che nessuno è completamente innocente qui (RESPIRO =0.552) negli ultimi otto anni abbiamo avuto il più g- il più grosso aumento di spesa: (RESPIRO =0.47) di deficit e di debito nazionale dena nostra storia-e-il senatore McCain ha votato (RESPIRO =0.614) per (COLPO DI GLOTTIDE) quattro su cinque di questi: bilanci di: eh [PAUSA =0.226] eh (RESPIRO =0.563) eh leggi finanziarie di Bush quindi io spenderò dei soldi sulle questioni chiare su cui dobbiamo lavorare (RESPIRO + PAUSA =1.979] eh [PAUSA =0.196] lei ha visto un aumento dei premi sull'assistenza sanitaria dobbiamo riformare: l'assistenza sanitaria dobbiamo affrontare l'energia perché non possiamo continuare a prendere in prestito d- (RESPIRO =0.188) dai cinesi dando: i soldi all'Arabia Saudita (RESPIRO =1.18) [PAUSA =0.253] eh ci vuole un piano energetico diverso dobbiamo investire /sulle,su le-/ sulla scuola: (RESPIRO =0.411) dobbiamo fare degli investimenti ma dobbiamo fare anche dei tagli alla spesa e la mia proposta (RESPIRO =0.728) (i)l senatore McCai(n) discse che queste: le mie pro(po)ste sono tutte nuove spese ma in realtà io sto tagliando più di quanto spendo (RESPIRO =0.542) quindi ci sarà un taglio sulla spesa netta allora la chiave è se noi abbiamo delle priorità che funzionano per voi (RESPIRO =0.394) rispetto (RESPIRO =0.412) a coloro che hanno [PAUSA =0.378] dettato le politiche di Washington negli ultimi otto anni cioè soprattutto le lobby e gli interessi speciali noi dobbiamo porre fine a tutto questo senatore McCain</p>
<b>WintPitch Transcript of VIDEO 2 - ORIGINAL VERSION</b>
<p>voglio dire un ultima cosa (RESRPIRO =1.136) perché: il senatore McCain ha parlato: [PAUSA = 0.501]del NAFTA e delle questioni commerciali una questione che voglio sottolineare (RESPIRO =0.25) io credo nel libero scambio (RESPIRO =0.199) ma credo anche (RESPIRO =0.229) [PAUSA =1.359] che per troppo tempo sicuramente nel corso dell'amministrazione Bu:sh [PAUSA =0.143] con l'appoggio del senatore McCai:n (RESPIRO =0.615) l'a- l'atteggiamento è stato che qualsiasi: (RESPIRO =0.1) accordo commerciale fosse giusto [PAUSA =2.561] eh!- nel NAFTA non c'erano degli accordi commerciali [PAUSA =0.829] e e di lavoro e: secondo me do(ve)vamo introdurli (RESPIRO =0.643) così come avremmo dovuto applicare delle: (RESPIRO =0.213) [PAUSA =0.444] regole contro la Ci:na: (RESPIRO =0.787) [PAUSA =0.629] per ren- (RESPIRO =0.401) [PAUSA =0.701] per rendere le loro [PAUSA =0.099] eh: importazioni (RESPIRO =0.428) hm più economiche (RESPIRO =0.966) [PAUSA =0.727] adesso abbiamo [PAUSA =0.478] adesso loro stanno inviando centinaia di migliaia di: macchine da Corea del Sud negli Stati Uniti da Corea del Sud questo va benissimo hm (RESPIRO =0.176) ma noi invece riusciamo soltanto a mandarne quattromila cinquemila in Corea del Sud questo non è libero scambio noi dobbiamo avere un presidente (RESPIRO =0.808) [PAUSA =1.146] che lavori</p>
<b>WinPitch Transcript of VIDEO 2 - MANIPULATED VERSION (IMITATION)</b>
<p>voglio dire un ultima cosa: [PAUSA =0.518] perché: il senatore McCain ha parla:to del NAFTA e delle questioni commerciali (RESPIRO =0.363) la questione che voglio sottolineare: (RESPIRO =0.269) io credo: nel libero scambio ma: [PAUSA =0.202- COLPO DI GLOTTIDE] cre:do . anche . che: [PAUSA =0.256] per troppo t:empo (RESPIRO =0.448) sicuramente nel corso dell'amministrazione Bush con l'appoggio del senatore McCai:l'atteggiamento è stato che qualsiasi accordo (RESPIRO =0.284) commerciale fosse giusto (COLPO DI GLOTTIDE) il: (RESPIRO =0.355) nel NAFTA non c'erano accordi commerciali e . e . e di lavoro e secondo me: (RESPIRO =0.683) eh: [PAUSA =0.519] dovevamo introdurli così come avremmo dovuto app:licare delle regole (RESPIRO =0.118) contro la Ci:na (RESPIRO =0.409) per rendere: [PAUSA =0.901] per rendere le loro: [PAUSA =0.402] importazioni più economiche [PAUSA =2.277] eh: adesso: abbiamo (RESPIRO =0.263) adesso loro stanno: (RESPIRO =0.227) [PAUSA =0.464] inviando centinaia di migliaia di macchine: da Corea del Sud negli Stati Uni:ti (RESPIRO =0.322) la Corea: del S:ud questo va benissimo ma noi invece (RESPIRO 0.228) riusciamo solta:nto a mandarne quattromila cinquemila in Corea del Sud questo (RESPIRO =0.237) non è libero scambio noi dobbiamo avere un presidente che lavori <i>[per gli Stati Uniti]</i></p>
<b>WinPitch Transcript of VIDEO 3</b>
<p>ma qual è il vantaggio [PAUSA =0.18] in una: zona: (RESPIRO 0.2) povera di mandare i lo- i figli [PAUSA =0.201] in scuole che stanno fallendo (RESPIRO 0.963) la concorrenza tra le scuole quindi [PAUSA =0.119] è uno d- degli elementi chiave [PAUSA =0.2] si è già visto . il successo . a New Orleans a: New York (RESPIRO =0.852) in cui abbiamo [PAUSA =0.379] delle scuole con ottimi insegnanti (RESPIRO =0.082) che vengono: eh:m premiati (RESPIRO =0.864) e troviamo [PAUSA =0.224] ai agli insegnanti meno bravi [PAUSA =0.237] altri posti di lavoro (RESPIRO =0.591) dobbiamo dare la stessa scelta [PAUSA =0.189] eh: [PAUSA =0.461] ai genitori di tutta l'America eh le stesse scelte che hanno avuto il senatore Obama e sua moglie e quella che ho avuto io e mia moglie (RESPIRO =0.426) di mandare i nostri figli alle scuole migliori (RESPIRO =0.898) [PAUSA =0.532] bisogna garantire la concorrenza tra le scuole bisogna: (RESPIRO =0.804) garantire quel tipo di concorrenza che ha migliorato sia le scuole pubbliche che quelle private (RESPIRO =0.615) ora [PAUSA =1.253] (RESPIRO =1.182) eh: gli investimenti il finanziamento non poss- ehm essere l'unica soluzione</p>

**Table 4.3.** Transcripts of the video excerpts used for the pilot survey. The original version in video 2 is reported to show the differences with respect to the imitation.

### 4.5.2.3 *Analysis of video excerpts*

#### 4.5.2.3.1 *Transcripts*

The transcription of the text related to the video excerpts was executed through WinPitch. The conventions of transcription were those adopted for the transcription of the corpus *CorIT* (see above – section 2.2.2); thus, the transcription was mainly orthographic. Pauses and audible breathes were reported; vocalized hesitations were represented with an accurate representation through orthography (e.g. “ehl:- nel”) and not with a standardized representation (eg. ehm, hm); cases of unstressed syllable deletion were also represented (e.g. “storia-e-il senatore”, “do(ve)vamo”). Transcripts of the video excerpts are shown in the table 4.3 above.

#### 4.5.2.3.2 *Identification of linguistic units and indices of analysis*

Before analysing the oral texts of the video excerpts, a series of linguistic units were identified. These units mainly refer to the context, the text, and its rhythmical-prosodic features. The units are listed and explained below.

**Context:** the communicative situation where the oral text originates and all other factors involved in the communication act, i.e. time, space, people, instruments; in this case, the communicative situation is the Third 2008 US Presidential Debate. This debate, as well as the first, the second and the Vice-presidential debate, is transmitted by television, live. It is broadcast live in the USA, but also in other countries.

In Italy (and similarly in other countries) it is transmitted live with a service of simultaneous interpretation from English into Italian and presented – in the broadcasting – in a voice-over modality, i.e. with the original English speech at a lower volume than that of the interpreters who are translating. In Italy (and similarly in other countries, like Spain, for example), the broadcaster that transmits the debates organizes and is responsible for the service of live simultaneous interpretation. In Italy (and Spain) the images that can be seen on television are taken from the international circuit of news by the broadcaster and then retransmitted in ~~its~~ screen format, i.e. with ~~its~~ screen graphic.

**Television genre:** both in USA and in Italy (and in other countries) the broadcasting of Presidential Debates is included in a television information programme; it is announced and presented as an event.

**TV interpreting format:** according to the classification of “formati del

discorso e generi mediali nell'interpretazione televisiva”, Straniero Sergio (2007:11) ascribes ‘deabtes’ to the ‘dialogic format’.

**TV interpreting mediatic genre:** in the same classification, Straniero Sergio (2007:11) ascribes ‘historical-political-celebrative events’ to both “breaking news” and ‘mediatic events’ as sub categories of “news” and “entertainment” respectively; being “news” and “entertainment” both forms of mediatic genres of television interpreting.

**Speaker → source-language listeners; interpreter → target-language listeners; client:** this is a theoretical model proposed by Gile (1995: 21-27; in Bendazzoli 2010: 52) ‘where the dynamics of the protagonists of the communication mediated by interpreters (and translators) are highlighted’. In this model, interpretation was considered as a “professional act of communication” where the main actors are:

- *[the] speaker*, cioè colui che trasmette il messaggio nella lingua di partenza (LP);
- *[the] source-language listeners*, cioè gli ascoltatori a cui lo speaker rivolge direttamente il suo messaggio poiché sono in grado di comprendere la LP;
- *[the] interpreter*, che trasmette il messaggio dello speaker nella lingua di arrivo (LA);
- *[the] target-language listeners*, ovvero gli ascoltatori ai quali lo speaker si rivolge “indirettamente” attraverso il supporto dell'interprete;
- *[the] client*, cioè la persona o l'ente che ha richiesto il servizio e ha affidato l'incarico professionale, provvedendo inoltre al pagamento della parcella dovuta (il client potrebbe anche coincidere, o meno, con lo speaker o con uno dei listeners). (Bendazzoli 2010: 52)

All these subjects are involved in the television broadcast simultaneous interpretation of US Presidential debate.

**Text:** the simultaneous interpretation from English into Italian of the video excerpts.

**Topic:** the subject developed by the speaker (either Obama or McCain, in our case), and interpreted by the interpreter.

**Phonetic chains:** “the sequence of phonetic segments (including filled pauses and disfluencies) delimited by two silent pauses” (Magno Caldognetto, Zmarich, Ferrero 1997: 779).

**Perceptual syllables:** the number of syllables perceived by the ear of the researcher. The procedure of definitions started from the syllabic scan of the text according to Italian standards, i.e. orthographic syllabification; subsequently, while listening to the text, this was adjusted considering the

cases of contraction (e.g. «storia-e-il senatore McCain»; «pro(po)ste»); vocalizations were also counted as syllables.

**Audible breaths:** mainly breathing-in between two successive phonetic chains. The breaths counted were *audible* according to the researcher's attentive perception: i.e. they were detected, and consequently counted, on the basis of careful attention by the listener; this could mean that some of them would pass unperceived by a listener who was not focused on them. If the researcher detected no sound of breathing, then it was counted as a silent pause.

**Silent pauses:** silent interval after a phonetic chain where no breathing was perceived by the researcher.

**Words:** totally uttered words, including those with a bad articulation or unstressed syllable deletion; words pronounced in part, as interruptions (restructurings and false start) were not counted as such, but as syllables.

**Disfluencies:** include filled pauses (vocalized hesitations, vowel and consonant lengthenings) and interruptions (repetitions, restructurings, false starts), according to the classification by Tissi (2005).

**Non-fluencies:** include filled and silent pauses and audible breaths, according to the classification by Tissi (2005).

**Total text:** “the whole speech production, i.e. the sum of speech chains” (Magno Caldognetto, Zmarich, Ferrero 1997: 779), and non-fluencies.

**Minimum and maximum values of the Fundamental frequency ( $F_0$ ):** the two values that mark the limit of the tonal range, inside this range the melodic curve (pitch contour or  $f$ ) develops in time lapse. Values are reported in semitones (and not in hertz) because this unit of measurement is a closer representation to the human perception (Domenico Di Russo 2015, personal communication). Some of the measurements related to rhythmical-prosodic features were computer-assisted, i.e. values were extracted through WinPitch: this is the case of the time of phonetic chains, silent pauses, audible breaths; and values of  $F_0$ . The values of  $F_0$  are related to both the  $F_0$  of the original text in English (in a low volume) and the  $F_0$  of the translated text in Italian (in higher volume). In fact, both sound tracks are merged in a unique mono digital track, and it is not possible to separate them, not even for acoustic analysis purposes. This is a great limitation to this study, due to the nature of the original material. Nevertheless, it was decided to measure the value of  $F_0$ , since it represents a physical aspect that has an impact on the perception of phonic matter, also because the condition of the two melodic

curves overlapped is equal in all three video excerpts, as well as in the corpus of the interpretations of the 2012 Presidential Debates (ORenesit).

On the basis of the measures of the above-mentioned units, a series of rhythmical-prosodic indices have been calculated, they are listed below.

**Speech rate index 1 – words per minute:** the number of words uttered each 60 seconds, as a value of speed of speech.

**Speech rate index 2 – syllables per second:** “the number of syllables divided by the total text time (syll./s)” (Magno Caldognetto, Zmarich, Ferrero 1997: 780), even if in this study this index is called “fluency index” (ibidem).

**Articulation rate index:** the number of well-articulated syllables (all syllables minus disfluencies and mispronunciations) divided by the phonetic chains time (syll/s). This index differs from the one adopted by Magno Caldognetto, Zmarich and Ferrero (1997: 780), i.e. “the number of syllables divided by the articulated sequence time”. It is also different from that adopted by De Meo, Pettorino and Vitale (2012: 234), called “*velocità di articolazione [...] equivalente al rapporto tra il numero di sillabe e il tempo impiegato a produrle*”. What are the reasons for this difference? Our approach does not consider the possibility of establishing the exact limits between an “articulated sequence”, as defined in Magno Caldognetto, Zmarich and Ferrero (1997: 779), i.e. “any phonetic chains excluding all filled pauses”, where filled pauses correspond to “any occurrence of hesitations, interjections, abnormally lengthened vowels, repetitions etc.” (ibidem). In our opinion, the problem is: how can the limit(s) of a vowel or consonant lengthening be established on a perceptual basis? Similarly: how can an “abnormally lengthened vowel” be objectively identified? How can the borders of unstressed syllable deletions be defined? A vocalized hesitation can be identified if it occurs between two silent pauses or breaths, but what if it occur in the middle of a phonetic chain? As to the speed of articulation calculated by De Meo, Pettorino, and Vitale (2012: 234), this is based on the number of syllables divided by the total time needed to produce them; in this case: if a vocalized hesitation was considered as a syllable, would it be also considered as a unit of articulation? Especially considering that the verb “to articulate” means “to give clear and effective utterance to”, “to utter distinctly”, or “to give shape or expression to” (Merriam-Webster Dictionary – on-line version)? In fact, it is true that from a pragmatic point of view, filled pauses (ehm, mm, ee) are often used to maintain the turn, but also indicate a difficulty in discourse planning, and co-occur with silent pauses, lengthenings



of the preceding vowel and correction markers (Bazzanella 1995: 225-257). Nonetheless, if filled pauses are considered ‘indices of difficulty’ in discourse planning; then, from the cognitive point of view, they could also be considered as “linguistic epiphenomena” (Domenico di Russo 2015, personal communication), i.e. secondary or collateral mental phenomena that are caused by and accompany a physical phenomenon (cf. Merriam-Webster Dictionary – on-line version; Vocabolario Treccani – on-line version). Whatever the interpretation of the filled pauses, the problem is that the a phonic signal is continuous and variable (Albano Leoni 2009: 85), and it is not possible to mark a limit in time and space – at least perceptually – between a simple vowel or consonant and its lengthening (e.g. “eh:”); and sometimes between a sequence of words (e.g. “ehl:- nel”; “storia-e-il senatore”). For this reason, in this analysis only the occurrences of disfluencies were considered, and not their duration (time). Consequently, the indices of disfluencies and non-fluencies were based on the number of occurrences and not on the duration of the units; the indices are listed below.

**Disfluencies rate index:** the number of disfluencies divided by the number of syllables x 100.

**Non-fluencies rate index:** the number of non-fluencies divided by the number of syllables x 100.





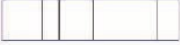


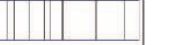
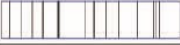



**Number of syntactical errors:** agreements, unfinished phrases or sentences. Agreement, cohesion errors and unfinished phrases or sentences are typical of conference interpreting speech.

**Tonal range / Delta  $F_0$ :** the maximum value of  $F_0$  minus the minimum one. As explained above, the limitation of this index is due to the melodic curves of both the speaker and the interpreter that are merged.

**AntConc analysis:** in order to consider the lexical density of each text of the video excerpts, an analysis through the software programme AntConc was executed. It was, however, limited to the following parameters, namely:

1. word types and word tokens and minus (BREATHS) and [PAUSES];
2. Type/Token Ratio (TTR);
3. the plot displays of (BREATHS), [PAUSES] and (BREATHS) + [PAUSES] in order to have a visual report of distribution of both elements of fluency.

The table below reports all the values of the units and the indices described above.

Linguistic units and indices of analysis	VIDEO 1	VIDEO 2 ORIGINAL	VIDEO 2	VIDEO 3
Context	US 2008 Presidential Debates			
Debate	2nd Debate	3rd Debate	3rd Debate	3rd Debate
Topic ( <a href="http://www.debates.org">www.debates.org</a> )	All Topics	The Economy and Domestic Policy		
Format ( <a href="http://www.debates.org">www.debates.org</a> )	town hall meeting	candidates seated at table with moderator		
TV genre	TV information			
TV interpreting mediatic genre	political event			
TV interpreting format	debate			
Speaker	Obama	Obama	Obama	McCain
Source text listener	English-comprehending TV audience			
Interpreter	Interpreter 1 (male)	Interpreter 1 (male)	Dubber (imitation) (male)	Interpreter 2 (male)
Target text listener	Italian comprehending TV audience			
Client	RAI - Italian State-owned broadcaster <i>Rainews24</i> - All-news channel			
Text	EN-IT simultaneous interpretation of political discourse			
Topic of the excerpt	US Budget policies	Free Trade agreements	Free trade agreements	US public schools' quality
Total number of phonetic chains	19	21	21	21
Total number of silent pauses	5	12	8	11
Total number of audible breaths	13	14	13	11
Total number of perceptual syllables	400	301	289	273
Total number of words	199	142	142	136
Total number of syntactical errors	1	2	1	1
Total time of phonetic chains (s)	58.9	45.3	50.6	44.6
Total time of silent pauses (s)	2.4	9.5	5.4	4
Total time of audible breaths (s)	8.6	7	4.1	7.5
Total time of pauses + breaths (s)	11	16.6	9.5	11.5
Total time of text (s)	69	61	61	58
Average duration of phonic chains (s)	3	2.2	2.4	2
Average duration of pauses + breaths (s)	0.5	0.5	0.5	0.4
Total number of disfluencies	18	23	48	17
filled pauses	12	14	26	13
vocalised hesitations	5	5	4	6
vowel and consonant lengthenings	7	9	22	7
Interruptions	6	9	7	4
repetitions	3	3	6	1
restructurings	1	2	1	0
false starts	2	4	0	3
Total number of non-fluencies	36	49	69	39
Min - Max values of F0 (semitones)	-4.3 - 12.7	-5.6 - 18.8	-3.8 - 14.7	-5.6 - 10.8
Speech rate index 1: words/min	173	140	140	136
Speech rate index 2: syll/s	5.8	4.8	4.6	4.6
Articulation rate index	6.4	6	4.6	5.74
Disfluencies rate index	4.5	7.5	16.6	6.1
Non-fluencies rate index	9	16.1	23.9	14.3
Tonal range / Delta F0 (semitones)	17	18.8	18.5	16.4
Word Types minus ( <b>BREATHS</b> ) and [ <b>PAUSES</b> ]	146	115	118	99
Word Tokens minus (BREATHS) and [PAUSES]	226	160	167	150
Type/Token Ratio (ITR) (%)	65	71.9	70	66
Plot Display of (BREATHS)				
Plot Display of [PAUSES]				
Plot Display of (BREATHS) + [PAUSES]				

**Table 4.4.** Values of linguistic units and indices in the three video excerpts of the pilot survey, plus the original video 2, reported to show the differences with the imitation.

4.5.2.3.3 Video 1 (compared to videos 2 and 3)

As for the context, this excerpt is taken from the second debate of the 2008 Presidential Debates, characterized by the format “town hall meeting debate”:

Candidates questioned by uncommitted voters identified by the Gallup Organization. In addition, moderator has discretion to include questions submitted online. Candidates questioned in turn with two-minute responses, followed by one-minute open discussion for each question.

(Commission on Presidential Debates > Debate History > 2008 Debates:  
<<http://debates.org/index.php?page=2008-debates>>)

The topic of this excerpt was the national budget policy, a domestic issue. The original speaker was Barak Obama, Senator of the United States. The interpreter was an employee of the Italian state-owned broadcaster *RAI* (Italian Television Radio); he worked as a television interpreter for the all-news channel *Rainews24*.

The oral text of this excerpt has the highest number of syllables (400), words (199); but the lowest number of phonetic chains (19) and silent pauses (5) than the other two video excerpts; the number of audible breaths is the same as in video 2 (13), but higher than that in video 3 (11).

This excerpt also presents the highest duration of all phonetic chains summed up (58.9); moreover, the average duration of these is the highest of the three videos (3). The total time of audible breaths is also the highest, but the total time of silent pauses is the lowest.

The duration of the whole text is the highest, with a difference of 8 seconds and 11 seconds compared to the second and the third videos respectively.

This text presents the lowest number of non-fluencies (disfluencies, audible breaths, silent pauses), even if the number (36) is close to that in video 3 (39). The same can be said about the disfluencies (filled pauses + interruptions), since Video 2 has just one unit more (18) than video 3 (17).

There is a notable difference between the types of the components of disfluencies: there are 12 filled pauses (vocalised hesitations, vowel and consonant lengthenings) and 6 interruptions (repetitions, restructurings and false starts). As to the subcomponents of filled pauses: there are 5 vocalised hesitations, and 7 vowel and consonant lengthenings – not a great difference in numbers. The same can be said for the subcomponents of interruptions: repetitions equals to 3, restructurings equals to 1 and false starts equals to 2.

As a consequence of the above described units, this excerpt presents the highest speech rates (173 words per minute and 5.8 syllables per second); and the highest articulation rate (6.4). However, it has the lowest disfluencies index (4.5) and non-fluencies index (9). The tonal range is not so different from those in videos 2 and 3 (17 semitones against 18.5 in video 2 and 16.4 in video 3).

According to the analysis executed through the AntConc program, this excerpts has the highest number of word types (146) and word tokens (226), pauses and breaths excluded; nevertheless, the type/token ratio (TTR) is the lowest (65) if compared with those in video 2 (70) and video 3 (66). This means that the lexical density, measured by the TTR, is not so high: in fact, it is the lowest of all the videos, despite the high speech rate and number of words.

The plot display of breaths shows a first movement (series) of breaths that progressively become more and more spaced out, up to two thirds of the entire duration. Subsequently, there is the largest distance between breaths, followed by another series of breaths progressively spaced out, slightly longer than the first one; its end does not coincide with the end of the excerpt.



**Figure 4.10.** Pitch contour of the interpretation in Video 1. The x-axis indicates the length of the audio excerpt (69 sec), while the y-axis reports, on the top, the tonal range (blue line), expressed in semitones (-12 to 24); and on the bottom the intensity (volume – blue graph) and the waveform (green graph). On the line corresponding to semitone -12 there are some traces of the original stereophonic track, converted into a monophonic track through Audacity®.

The plot display of pauses show few pauses that mostly occur at the end of the first series of breaths, at the beginning and end of the second series. From a glance at all the plots, it is clear that this excerpt presents the lowest number of pauses, given the high speed of speech.

The melodic curve of the video 1, from a single monophonic audio track where the sound of the original English speaker (Obama) was merged with the track of the Italian Interpreter 1, appears as an almost continuous line, with few interruptions (see Figure 4.10).

#### 4.5.2.3.4 Video 2 (original vs. experimental variable)

This section has the aim of highlighting the differences between the original version of the interpretation in video excerpt 2 and the imitation of the interpreted text made by an actor and dubber; the latter is the official version of video 2, used in the questionnaire as an experimental variable.

The context of this excerpt is the third debate of 2008, characterized by the following format: “90-minute debate with candidates seated at table with moderator. Candidates questioned in turn with two-minute responses, followed by five minutes of open discussion for each question” (Commission on Presidential Debates > Debate History > 2008 Debates: <http://debates.org/index.php?page=2008-debates>).

The topic of the debate is “The Economy and Domestic Policy” (ibidem). The interpreter is the same that in video 1. In the manipulated version, the interpreter is an actor and dubber who imitated the original oral text. The topic of this excerpt is free trade agreements between the USA and other countries.

With respect to the original version, the imitated one presents the same number of words (142), phonetic chains (21), and only one audible breath less (13 against 12), as well as one syntactical error less (1 against 2). If the average duration of silent pauses plus breaths is the same (0.5 seconds), the average duration of phonetic chains in the imitated version is slightly higher (2.4 against 2.2 seconds) However, the imitated version has a lower number of perceptual syllables (289 against 301) and silent pauses (8 against 12). Despite the same number of phonetic chains, in the imitated version the total utterance time is higher (50.6 against 45.3 seconds).

Notwithstanding almost the same number of audible breaths in both versions, the duration in the imitated version is lower (4.1 against 7 seconds of the original). The total time of text is the same (61 seconds). A remarkable

difference between the two versions is represented by the difference in the number of disfluencies: 48 in the imitated version against 23 in the original one – more than double. Looking more in detail at the components of disfluencies, one finds that the difference is marked by the number of filled pauses (48 against 23 of the original), and not by the number of interruptions, which are even less in the imitated version than in the original one (7 and 9, respectively).

Looking even more in detail, i.e. at the subcomponents of filled pauses, vowel and consonant lengthenings in the imitated version double in number those of the original version (26 against 14); while vocalised hesitations in the imitated version are less than in the original one (4 against 5). As for the subcomponents of interruptions, in the imitated version there are 6 repetitions against 3 of the original one, 1 restructuring against 2, and 0 false starts against 4; the latter is a notable difference.

As concerns the minimum and maximum values of fundamental frequency ( $F_0$ ), it is important to remember that this is the result from the blended tracks of both the speaker and the interpreter. In addition, the imitated version was the product of an editing process executed by the researcher through software programs Adobe Premiere Pro CS3 and Audacity (see above). The minimum and maximum values related to the original version are -5.8 and 18.8 semitones respectively; while those of the imitation are -3.8 and 14.7 semitones. The difference of the two values ( $\Delta F_0$ ) is 18.8 semitones for the original version and 18.5 of the imitated version – not a big difference.

Nevertheless, considering that the imitated version is the one used for the perceptual test, it is worth noting that the tonal range of this version is the highest of all the three video excerpts: 18.5 against 17 in video 1 and 16.4 in video 3. In addition, the minimum value of  $F_0$  of this version is the lowest (-3.8 against -4.3 in video 1 and -5.6 in video 3); while the maximum value is the highest (14.7 against 12.7 in video 1 and 10.8 in video 3). A more detailed analysis of the intonation related to this video is described below (see next section: 4.5.2.3.5).

As concerns the other indices, the number of words per minute is the same for the two versions (140), but the number of syllables per seconds is slightly lower in the imitation (4.6 against 4.8) due to the lower number of vocalised hesitations and interruptions. For the same reason, the articulation rate index is lower in the imitation than in the original version (4.6 against 6),

but the disfluencies index is higher (16.6 against 7.5). Differences in disfluencies between the two versions also explain the higher values in the imitated version of both word types (118 against 115) and word tokens (167 against 160), considered that the number of words in both versions is the same. Consequently, the type/token ratio is lower in the imitated version, mainly due to the absence here of false starts and to the higher number of repetitions.

As for the distribution of breaths, in the imitated version one can notice a uniform distribution along the whole text, against an irregular distribution in the original version. In the imitated version, silent pauses are concentrated in one part of the text. In this same part of the text, in the original version there is a concentration of pauses as well, but in the imitation text one can notice the presence of other pauses in the previous parts of the text more or less regularly distributed.

It is worth remembering that in the imitated version, many pauses were not natural, but artificially created by the researcher through the program Audacity®, in order to *adjust* the décalage of the interpretation (with respect to the original English text); for the same reason, with the same technique, other natural pauses produced by the dubber were deleted; However, with the aim of preserving the natural sound of human speech, the actor's audible breaths were not deleted.

#### 4.5.2.3.5 Video 2 (the performance of the dubbing actor)

The dubbing actor Christian Iansante was selected for his professionalism and especially for his peculiar speech. Perceptually, the peculiarity of his speech consists of pushing the limits of intonation and rhythm with respect to syntax. This means that – perceptually – he creates additional suspensive pauses where they are not expected to occur, considering the syntax or the text of scripts, i.e. not in correspondence, for example, of boundaries of noun with verbal groups, or in correspondence of colons, semicolons or full stops. In addition, Iansante's intonation appears to be characterized – again, perceptually – by a slight flattening of the rising movement in correspondence of comas and colons; as well as a slight flattening of falling movement in correspondence of full stops<sup>3</sup>.

---

<sup>3</sup> Christodoulides and Lenglet (2014) measured the “mismatch between prosodic and syntactic boundaries and pauses within syntactic units”, in read vs. simultaneous interpretation from German into French, to study the impact of fluency in the perception of quality in SI.

This is the reason why he is very much in demand among the radio recording studios and TV post-production studios in Rome. For this same reason, he mainly works as narrator (in voice-over mode) for documentaries and docu-dramas. These perceptual features of Iansante's readings of scripts perhaps constitute an added value that is at the basis of his success, which, therefore, could be due to the aesthetics of his speech, considered telegenic and radiogenic. Fónagy (1983: 256-266 – see section 3.6.4 and 3.6.5) demonstrated that intonation, in artistic voice, may induce the listener to perceive more silent pauses than in reality, i.e. after physical analysis of intonation, do not exist.

Considering that simultaneous interpretation has peculiar rhythm and prosody different from ordinary speech (cf. Ahrens 2005; Shlesinger 1994; see above – section 1.4.2.), Iansante, was selected for the imitation of an authentic interpretation. The imitation was selected as an experimental variable because it was deemed to have better ecological validity with respect to the reading, used in other similar studies (see above – section 4.5.2.2). The speaker listened to the original Italian interpretation while reading the transcript and afterwards imitated the interpretation while listening to the original English speaker through the headphones.

The dubber was instructed about the transcription conventions, especially about the signs that represented vowel and consonant lengthenings (:.) and silent pauses (/); in addition, vocalized hesitations, false starts, repetitions and restructurings were also shown in the script. Despite being told to respect (i.e. reproduce) all the disfluencies marked in the script, in the real execution of the imitation the speaker, produced his own disfluencies and silent pauses.

The reading-imitation was recorded in a professional sound recording studio<sup>4</sup>. The recording session was similar to a voice-over recording session, because the dubber received the English original (the speaker understands English) in his headphones. The dubbing actor read the transcript trying to reproduce a *décalage*, but he succeeded doing so only in the first recordings of two video excerpts; however, in these cases he did not manage to respect the

---

<sup>4</sup> We take this opportunity to express our deep gratitude to Christian Iansante and the staff of the studio General Jingles, in Rome, for having permitted the recordings used for this study.

We also thank Stefano Gaiani Billi and the studio European Television Service Srl (ETS), in Rome, for having permitted other recordings that were not used for this study.



time limit of the original sequence; in other words, his pace was too slow and the *décalage* was too long.

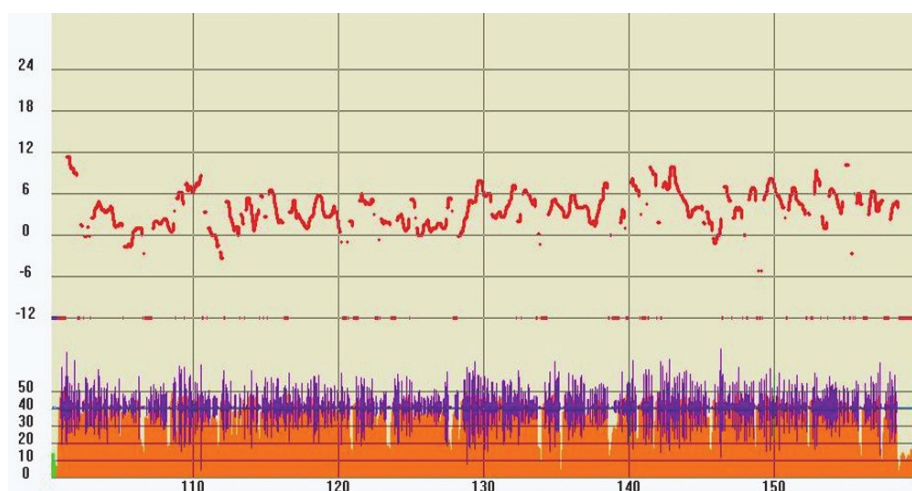
1 \_ Iansante

in diec'anni noi potremo ridurre la nostra dipendenza in modo da non dover più importare petrolio dal Medio Oriente o dal Venezuela penso che: que:sto è una tempistica realistica questa è la questione più importante che: la nostra economia futura dovrà affrontare adesso c'è una una crisi impellente ma non c'è nulla di più importante / del fatto di non prendere più in prestito settecento miliardi di dollari o più dalla Cina per poi mandarli in Arabia Saudita, questo significa ipotecare il futuro dei nostri figli, dall'inizio della nostra campagna io ho identificato questa come una delle mie principali priorità, ecco quello che dovremmo fa:re prima di tutto, abbiamo bisogno di espandere / la produzione interna, il che significa ed esempio: dire alle società petrolifere che i: sessantotto milioni di a:cri che attualmente non sono / hanno a noleggio, li dovranno utilizzare altrimenti li perderanno e sicuramente ci sarà la pref- la trivellazione: a largo / da attuare in modo da aver / ul:teriore petrolio, ma dobbiamo capire che abbiamo soltanto tre quattro percento di riserve petrolifere e noi utilizziamo venticinque percento del petrolio mondiale, il che significa che non possiamo uscire dal problema trivellando, quindi dobbiamo concentrarci su l'investire risorse nel sola:re l'eo:lico biodiesel / il geote:rmico que:ste sono le priorità che ho portato in Sena:to ed è assolutamente cruciale sviluppare / delle / macchine non costruite in Giappone o in Corea del Sud che risparmino energia, ma costruite qui negli Stati Uniti d'America, noi

commerciali una questione che voglio sottolineare / io cre:do nel libero scambio / ma credo anche / che per tro:ppo tempo sicuramente nel corso dell'amministarzione Bush con l'appoggio del senatore McCain l'a- l'atteggiamento è stato che qualsiasi: accordo commerciale fosse giusto / il: nel Nafta non c'erano accordi commerciali / e e di lavoro e: secondo me dovevamo introdurli così come avremmo dovuto applicare delle: regole contro la Ci:na / per rend- per rendere le lo:ro eh: importazioni più economiche / adesso abbiamo, adesso loro stanno inviando centinaia di migliaia di macchine da Corea del Sud negli Stati Uniti, <sup>IA Corea del sud,</sup> ~~la Corea del Sud~~ questo va benissimo ma noi invece riusciamo soltanto a mandarne quattromila cinquemila in Corea del Sud questo non è libero scambio noi dobbiamo avere un presidente / che lavori / eh in nome delle: società americane.

**Figure 4.11** Transcript of the original interpretation with the original notes written by the dubber after listening to the original interpretation and before the recording of his performance. The part highlighted in grey corresponds to the text of the Video excerpt used for the experiment, the first part was cut.

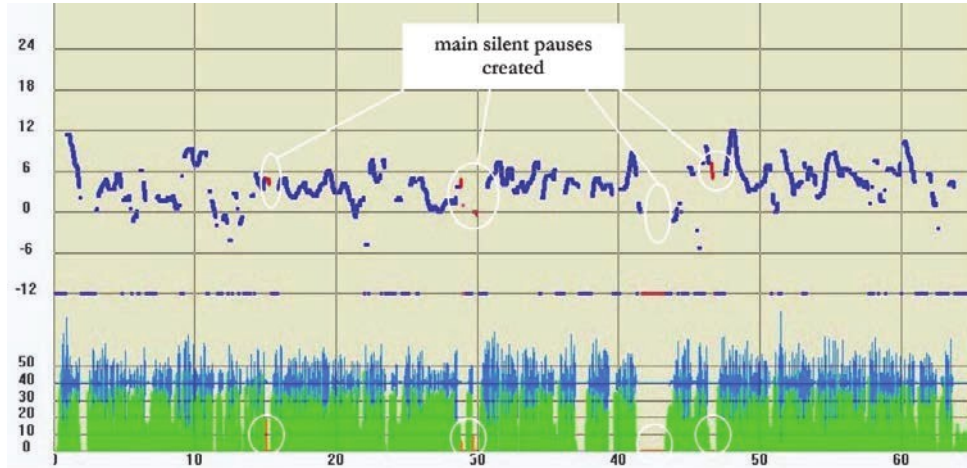
Therefore, in the subsequent recordings, he paid more attention to the time synchronization between the original speech and his own speech, and he managed to control the *décalage* albeit only in the first parts of subsequent recordings. In total, 7 takes were recorded for two different texts (interpretations/video excerpts), for a total of 791 seconds. The recording used in this pilot study needed 3 takes, the last one was thought to be acceptable, because in the previous takes the speaker had used a lower speed of speech, thus the time synchronization was not respected. In the figure below is reported the original script used by the speaker with his own annotations; the text highlighted in grey is the one that was cut and then used for the pilot survey (video 2).



**Figure 4.12.** Pitch contour (from WinPitch) of the dubbing actor’s imitation of the original interpretation recorded in stereophonic mode and later rendered monophonic through Audacity®.

At the end of the recording session, the dubbing actor, in a free comment on his performance, stated that his main effort focused on intonation: “everything depends on intonation” (Iansante 2012, personal communication). The melodic curve of the original recording of the dubber’s performance (corresponding to the part highlighted in grey in the figure 4.11 – above) has a tonal range of 18.6 semitones (min= -4.9; max= 13.7) (see below – figure 4.12).

As explained above, the original recording was manipulated through Audacity®, in order to artificially create the *décalage*; hence, some silent pauses were cut and others were added. The melodic curve resulting from this work is shown below (figure 4.13).



**Figure 4.13.** Pitch contour (from WinPitch) of the dubbing actor’s imitation, after the session of *décalage* settlement through Audacity®. Main silent pauses lengthened were highlighted through WinPitch function and then indicated in the pitch contour (high) and the waveform (low).

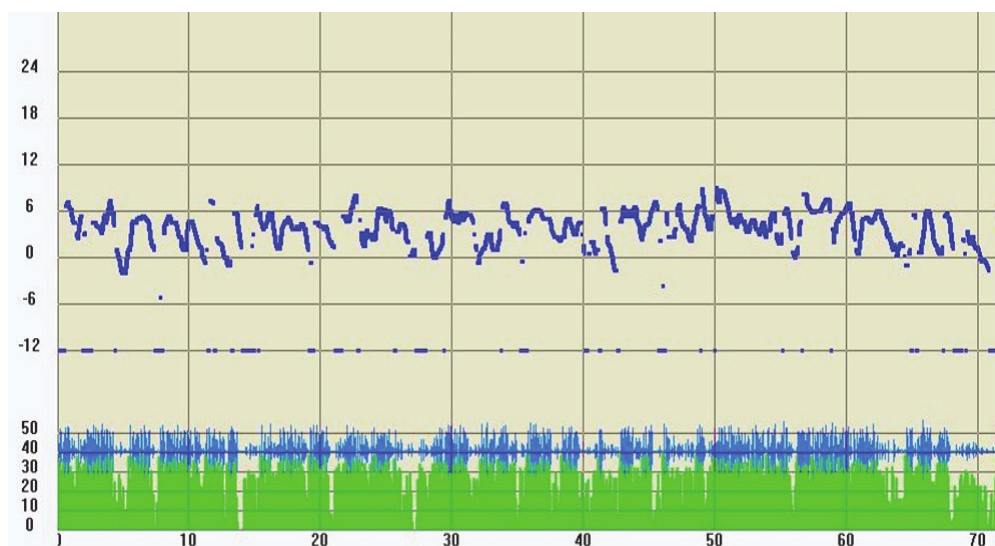
The imitation audio track was then merged with the audio track of the original English – through Audacity® – to obtain a single audio track that was added to the video excerpt from the Italian TV channel – through Adobe Premiere Pro CS3. The melodic curve of the two voices blended, then used for the perceptual test, is shown below (figure 4.14).



**Figure 4.14.** Pitch contour (from WinPitch) in video 2 (after the experimental modification).

A comparison between the manipulated track (imitation + original speaker) and the audio track of the original video excerpt (original

interpretation + original speaker), with the voice of interpreter 1, and its pitch contour is shown below (figure 4.15).



**Figure 4.15.** Pitch contour (from WinPitch) of the original video 2 (speaker + original interpretation).

#### *4.5.2.3.6 Video 2 (compared to videos 1 and 3)*

Video 2 has the same number of phonetic chains as video 3 (21), more silent pauses than video 1 but less than video 3 (8 against 5 and 11, respectively). It has the same number of audible breaths as video 1 (13), but less than video 3 (11). The number of perceptual syllables here is lower than those in video 1, but higher than those in video 3 (289 against 400 and 273, respectively). There is one syntactical error as in video 1 and video 3.

The total time of phonetic chains is 50.6 seconds: higher than that in video 1 (58.9), and lower than that in video 3 (44.6). However, the total time of silent pauses is the highest in the three videos: 5.4 seconds against 2.4 in video 1, and 4 in video 3. The total time of audible breaths is the lowest in the three videos: 4.1 seconds against 8.6 in video 1 and 7.5 in video 3. The same proportion applies to the total time of silent pauses plus breaths: 9.5 against 6.9 in video 1 and 11.5 in video 3.

The total time of text of this video excerpt is the second longest among the three (61 s), in between video 1 (69 s) and video 3 (58 s). This same ranking pattern is valid for the average duration of phonetic chains: 2.4 seconds against 3 in video 1 and 2 in video 3. The average duration of pauses plus breaths is almost equal in the three video excerpts: 0.5 seconds for videos 1 and 2, and 0.4 for video 3.

The total number of disfluencies is by far the highest of the three videos: 48 against 18 in video 1 and 17 in video 3. Looking more in detail at the structure of disfluencies, i.e. at the components and subcomponents of disfluencies, this number is mainly due to the high number of vowel and consonant lengthenings (22 against 7 of both video 1 and 2). As concerns the other subcomponent of filled pauses, i.e. vocalised hesitations, in this video there are 4 occurrences (against 5 in video 1 and 6 in video 3), the lowest number of the three videos.

On the contrary, there is the highest number of repetitions (6 against 3 in video 1 and 1 in video 3). Since repetitions constitute a subcomponent of interruptions, the number of interruptions in this video excerpt is also the highest (6 against 3 in video 1 and 1 in video 3), because the difference in number of the other subcomponents, i.e. restructurings and false starts, is not high. Both in this video and in video 1, there is only 1 restructuring, and 0 in video 3; as to false starts, in this video there are none, against 2 in video 1 and 3 in video 3.

The speech rate index 1 (words per minutes) is 140, i.e. an intermediate rate between video 1 (173) and video 3 (136). However, the speech rate index 2 (syllables per seconds) is the same as that in video 3 (4.6), and lower than that in video 1 (5.8). The articulation rate index is the lowest of the three videos: 4.6, against 6.4 in video 1 and 5.74 in video 3. As a consequence of the values described above, disfluencies and non-fluencies rate indices are the highest of all videos: respectively 16.6 (against, 4.5 in video 1 and 6.1 in video 3), and 23.9 (against 9 in video 1 and 14.3 in video 3).

As for the minimum and maximum values of the fundamental frequency ( $F_0$ ), this video, compared to other videos, has the lowest minimum value (-3.8 semitones, against -4.3 in video 1 and -5.6 in video 3) and the highest maximum one (14.7 semitones, against 12.7 in video 1 and 10.8 in video 3); as a consequence of this, the tonal range is the highest of all videos: 18.5 semitones, against 17 in video 1 and 16.4 in video 3.

The number of word types (118) is an intermediate value between those in video 1 (146) and video 3 (99). The same is true for the word tokens (167, against 226 in video 1 and 150 in video 2). Despite these values, the type/token ratio is the highest of the three videos: 70, against 65 in video 1 and 66 in video 3. This means the highest lexical density lies in the text of video 2, if compared with the texts of videos 1 and 3.

The plot display of audible breaths shows a uniform or regular distribution in video 2, if compared to those of other videos. However, the same cannot be said about the distribution of silent pauses, since the display plot shows a concentration of pauses at the beginning of the second part of the text; while in video 3, silent pauses appear more regularly distributed along all the first part of the text.

#### 4.5.2.3.7 Video 3 (compared to videos 1 and 2)

The context of this excerpt is the Third Presidential Debate of 2008, the same as that of video excerpt 2 (see above – section 4.5.2.3.4). Therefore, the format is the following: “90-minute debate with candidates seated at table with moderator. Candidates questioned in turn with two-minute responses, followed by five minutes of open discussion for each question” (Commission on Presidential Debates > Debate History > 2008 Debates: <http://debates.org/index.php?page=2008-debates>).

The topic of the Debate is “The Economy and Domestic Policy” (ibidem); while the topic of the excerpt is US public school quality. The interpreter is different from that of video excerpts 1 and 2; the excerpt is taken from a different broadcaster, i.e. the Italian private all-news channel Sky-TG24.

In this video excerpt, the number of phonetic chains is the same as that in video 2 (21, against 19 in video 1). The number of silent pauses is the highest of all videos (11, against 5 in video 1 and 8 in video 2). The same is true for the number of audible breaths (11, against 5 in video 1 and 8 in video 2). On the contrary, the number of perceptual syllables is the lowest of all videos (273, against 400 in video 1 and 289 in video 2). The same is true for the number of words (136, against 199 in video 1 and 142 in video 2).

There is only 1 syntactical error, just as in videos 1 and 2. The total time of phonetic chains is the lowest of all videos (44.6, against 58.9 in video 1 and 50.6 in video 2). The total time of silent pauses is 4 seconds: higher than that in video 2 (2.4) and lower than that in video 2 (5.4). While the total time of audible breaths is 7.5 seconds: higher than that in video 2 (4.1) and lower than that in video 1 (8.6). However, the total time of silent pauses plus breaths is the highest of all videos (11.5, against 11 in video 1 and 9.5 in video 2).

The total time of text is the lowest in all videos (58 seconds, against 69 in video 2 and 61 in video 2). The same proportion applies to the average duration of phonetic chains (2 seconds, against 3 of video 1 and 2.4 of video

2), and the average duration of silent pauses plus breaths (0.4 seconds, against 0.5 of both videos 1 and 2).

The total number of disfluencies in this video is the lowest of all videos (17, against 18 in video 1 and 48 in video 2). This number is mainly due to the high number of filled pauses (13, against 12 in video 1 and 26 in video 2), more than the number of interruptions (only 4, against 6 in video 1 and 7 in video 2). The number of subcomponents of filled pauses is almost equal; in fact, there are 6 vocalised hesitations (against 5 in video 1 and 4 in video 2) and 7 vowel and consonant lengthenings (just as many as in video 1, but in video 2 are 22).

As to the subcomponents of interruptions, there are: only 1 repetition (against 3 in video 1 and 6 in video 2); 0 restructurings (against 1 in both videos 1 and 2); and 3 false starts (against 2 in video 1 and 0 in video 2). As a result of all these values, the total number of non-fluencies is 39, slightly higher than that in video 1 (36) and definitely lower than that in video 2 (69).

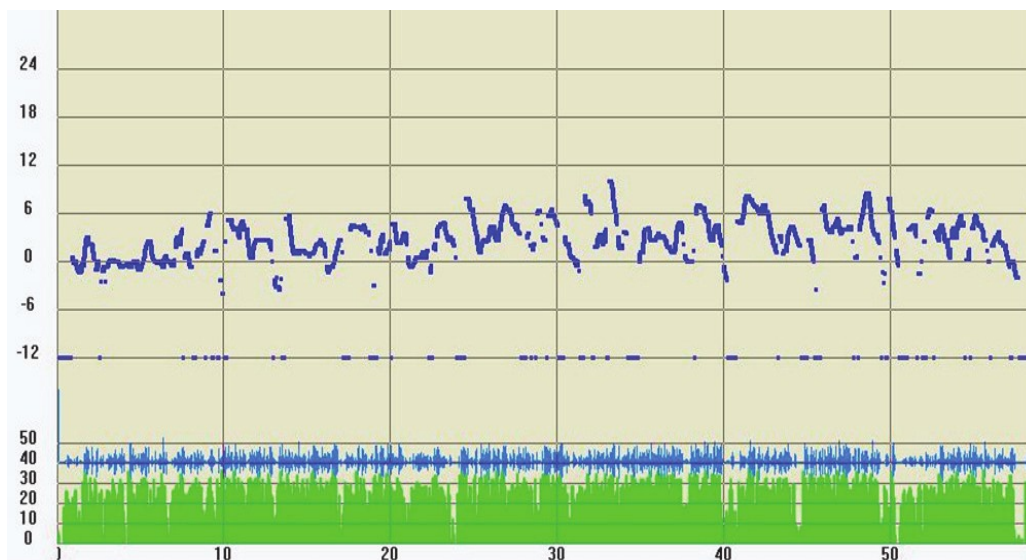
The speech rate index 1 (words per minute) in this video is the lowest of all videos (136, against 173 in video 1 and 140 in video 2). While the speech rate index 2 (syllables per second) is the same as that of video 2 (4.6), but lower than that in video 1 (5.8). The articulation rate index is 6.1: an intermediate value between that in video 1 (5.74) and that in video 2 (4.6). The same proportion applies to the disfluencies rate index (6.1, against 4.5 in video 1 and 16.6 in video 2), and the non-fluencies rate index (16.4, against 17 in video 1 and 18.5 in video 2).

As for the minimum and maximum values of fundamental frequency ( $F_0$ ), this excerpt shows the lowest minimum value (-5.6 semitones, against -4.3 in video 1 and -3.8 in video 2), and the lowest maximum value (10.8, against 12.7 in video 1 and 14.7 in video 2). The tonal range is the lowest of all videos (16.4 semitones, against 17 in video 1 and 18.5 in video 2).

The melodic curve of both the original English speaker and the Italian interpreter sound track is shown below (figure 4.16).

The number of word types is the lowest of all videos (99, against 146 in video 1 and 118 in video 2). The same is true for the number of word tokens (150, against 226 in video 1 and 70 in video 2). Despite these values, the type/token ratio, indicator of the lexical density, is not the lowest in all videos; in fact, its value is 66%, almost the same as that in video 1 (65%) and lower than that in video 2 (70%).

The plot display of audible breaths shows a distribution which appears less regular than that in video 2. While the plot display of silent pauses shows a relatively regular distribution only in the first part of the video, because in the second part there are only 2 occurrences of pauses.



**Figure 4.16.** Pitch contour (from WinPitch) of the video 3 (speaker + original interpretation).

#### 4.5.2.3.8 Comparison of Videos 1, 2 and 3

From the analysis of each video compared with the other two, it seems that the differences among the three excerpts perceptually detected in function of a scale of difficulty (low, medium and high) were confirmed by the values of linguistic units and indices of analysis.

Video 1 has the most difficult topic, i.e. US budget policies; it is a topic that requires – on the part of the interpreter and of the audience – at least a general knowledge of US domestic politics, especially those carried out by the US admin still in office at the time of the Debate.

In addition to the technical subject, this video has the highest speech rate and articulation index, but the lowest non-fluencies index. Therefore, it is dense, even if the type/token ratio is the lowest in the three videos and is almost the same than that in video 3; to this respect, it should be considered that the type/token ratio in video 2 is higher because of the high number of vocalised hesitations.

The topic of video 2, free trade agreements (foreign affairs and commercial relations), is less difficult than that of video 1; this can be seen



from a comparison between the two texts (see above: transcripts – figure 4.3); for example, the subject matter in video 1 was developed by the interpreter through the use of a more technical words with respect to what was done by the interpreter in video 2.

Video 2 has the same speech rate as video 3, and both are lower than that in video 1; nevertheless, the articulation rate is the lowest, but the non-fluencies rate is the highest, mainly due to the high number of vowel and consonant lengthenings. This may be the trace of a marked intonation used by the dubbing actor, who stressed a higher number of syllables to better communicate what he was saying; in addition, tonal range is the highest of all videos, and the distribution of audible breaths is the most regular. The type/token ratio of this video is the highest, but not so far from those of other videos.

The topic of video 3, US public school quality, is not an easy topic; however, it turned out to be the easiest of the three videos, because it was not dealt with technically (see above; transcript – figure 4.3). In fact, the type/token ratio is almost the same as that in video 1. In addition, this video excerpt has the lowest number of words, but the highest total time of breaths plus pauses, even if it is almost the same than that in video 2; the same is true for the speech rate. The non-fluencies index is five points higher than that in video 1; in fact, there are not as many vowel and consonant lengthenings as in video 2. The easy topic, the many pauses and breaths, more or less regularly distributed, the low number of words, which are not so difficult, are all elements that should render the interpretation easy to be followed and comprehended.

In conclusion, compared to other videos, video 1 has a difficult subject, with complex words; however it has a medium tonal range, a high fluency with few silent pauses: all elements that should help comprehensibility, counterbalancing the relative difficulty of its topic; nevertheless, its credibility should not be high (see above – section 4.0.3); the interpreter's performance could be judged to be high, given the ability to translate such a dense passage, at a high speed.

Video 2 has a subject of medium difficulty, with a high tonal range, a high total time of silent pauses, but not an high number of them, and a high number of sillable lengthenings; therefore it should be considered of medium comprehensibility and credibility, and high phonaesthetic voice (see above – section 4.0.3); as to the interpreter's performance, actually there is no

interpreter's performance but only a speaker's performance, and this aspect could influence assessment, even if respondents did not recognize the nature of the performance.

The nature of the performance in video 2 does not represent the only anomaly; another abnormality is represented by the proportion among i) the number of silent pauses, ii) the total time of silent pause, and iii) vowel and consonant lengthenings. In fact, vowel and consonant lengthenings co-occur with silent pauses, in spontaneous speech (Fónagy 1983), and also in simultaneous interpreting speech (as experienced through transcription; see above – section 4.5.2.3.1); however, in the experimental variable of video 2, many silent pauses were cut to create the *décalage*; for the same reason, few silent pauses were lengthened, as it was the case of the long central pause (2.3 s), having a duration notably above the duration of other pauses (0.2 to 0.5 s). This long pause could have an impact of the assessment of parameters, considering the overall duration of the video excerpt.

Video 3 has an easy subject, with a high number of silent pauses, well distributed along the text, a low number of non-fluencies; therefore, it should result in medium comprehensibility and credibility, even if the voice is not phonaesthetic, given the low tonal range; the interpreter's performance is not that good, given the easy topic and the low speed of speech of the source text.

#### **4.6 Subjects**

The subjects chosen for the pilot survey are MA and BA students in conference interpretation and specialized translation at the University of Trieste. They represent a homogeneous group of around 100 people; this group cannot be defined a representative sample, according to survey's theory, given the low number of subjects (Delli Zotti 2015, personal communication). The main purpose of the pilot survey is to test the questionnaire, to check the validity of its items, and not to define a profile of a statistically representative group of people in society. The subject chosen know what it is an interpretation and a translated text; therefore, they could evaluate it. In its definitive version, the survey was thought to be administered to professional interpreters, TV experts, ordinary TV viewers, actors and musicians.

## 4.7 Results and analysis of the pilot survey

### 4.7.1 Responses

The questionnaire was administered to 101 students attending the courses of Applied Interlinguistic Communication BA Programme (17 subjects) and Specialised Translation and Conference Interpreting MA Programme (84 subjects); both courses are offered by the University of Trieste (Section of Studies on Modern Languages for Interpreting and Translation). The 17 BA students already attended the courses of Liaison Interpreting (between at least two language pairs); while the 84 MA students attended Specialised Translation courses (36) or Conference Interpreting courses (48). Of all the subjects, 14 were males, 86 were females, 1 did not reply. The average age was 23.

RANDOMIZATION	
Sequence of videos	Subjects (N = 101 = %)
1 - 2 - 3	15
1 - 3 - 2	13
2 - 1 - 3	15
2 - 3 - 1	14
3 - 1 - 2	21
3 - 2 - 1	22

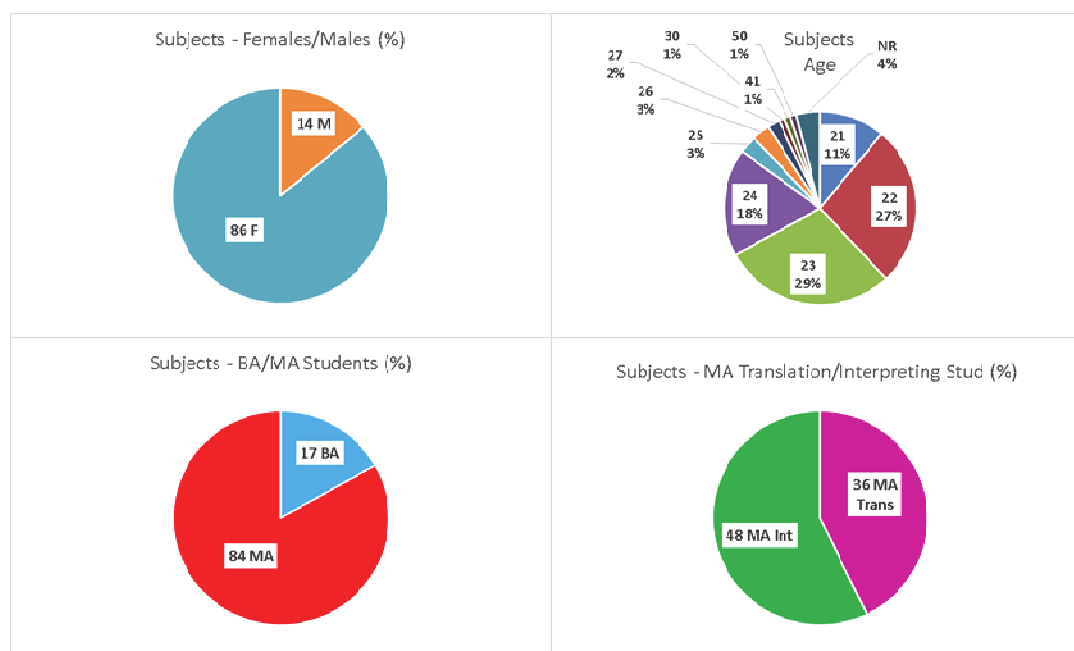
**Table 4.5.** Randomization of the videos 1, 2 and 3.

### 4.7.2 Responses to open-ended questions

The questionnaire included an open-ended question at the end of each of the three series of questions related to the three video excerpts: a question asking respondents a free comment on aspects of the interpretation not included in the previous closed questions. In addition, at the end of the block on personal data, there was another open-ended question on possible free comments on the questionnaire or aspects not included in other questions.

Of all respondents, 48 out of 101 subjects answered the open-ended questions, both those related to each video and the final one on the questionnaire. Of these, 35 answered the questions related to the three videos: 18 respondents answered the question on video 1; 18 answered the question on video 2 and 17 answered the question on video 3.

Of all the subjects, 33 answered the question related to the whole questionnaire.



**Figure 4.17.** Social and demographic profile of the subjects of the pilot survey.

#### 4.7.2.1 Open-ended question on Video 1

In one case the interpreter was considered ‘constantly anxious, worried about lagging behind’; it is not the only case, since another subject wrote that the interpreter ‘had a notable time-lag’. According to one respondent, ‘there were excessively long pauses followed by sentences pronounced too fast’. In one case ‘decelerations and accelerations’ in interpreter’s speech made difficult for one subject to ‘follow the thread of what the interpreter was saying’. A respondent found ‘too many filled pauses and dragged out vowels’. A subject wrote that the interpreter ‘seemed to talk about energy without understanding what he was talking about’. In the same line, a respondent said that ‘my impression is that the interpreter didn’t always understand the speaker’s message; therefore, some restructurings turned out very awkward’. Another subject had the impression that the interpreter ‘was gasping for words, probably because of President Obama’s speaking rate; it is as if the interpreter wanted to say everything at the same speed’. According to one respondent ‘the interpreter sounded very confident, considering the tone of his voice and his aptitude. However, in some cases, I detected some sentences that were not duly finished’. A respondent said that ‘the interpreter corrected himself by

changing the words and the structure of the sentence’. Another one noted that ‘there was a long *décalage* and false starts’. However, in one case the interpretation of video 1 was considered ‘the best interpretation of the three’. A subject found that the ‘interpreter uttered disconnected sentences’, for that reason s/he found it ‘difficult to grasp the overall meaning – the very short sentences could have made sense as separate sentences’. In one case the interpreter was considered ‘chaotic, confused. He articulated properly, but the content was confused and not very comprehensible’.

#### 4.7.2.2 *Open-ended question on Video 2*

A subject wrote that the interpreter “had a pleasant and captivating speech”; however, “the interpretation was not good, the use of language was not correct”, and “the interpretation was not natural for such a type of discourse”. Another subject said that “the interpretation was much more usable than the first one, especially because the rhythm was slower”. In one case it was noted that “some sentences were not finished, hampering the general comprehension of the speech, especially in the second part. Nevertheless, this same subject had “a good impression of credibility and confidence towards the interpreter”. Another respondent detected “self-repairs – especially the replacement of a personal pronoun at the beginning of the sentence”. In one case, the interpreter of this video excerpt was judged “the best of the three interpreters (maybe thanks to the more pleasant voice)”. In another case, it was noted that “the interpreter didn’t use a boring tone. Despite a couple of repairs, the delivery was not too bad”. According to a respondent, “the grammar structures of the Italian interpretation were not always perfect; however, the overall message was very clear”. Another respondent wrote that “the logical links among the different parts of speech were not very clear”. According to a subject, “the interpreter here used a monotone and boring voice and, in the interpretation, tried to reproduce the source text without executing a real interpretation in the target language, since he often used unusual Italian structures”. In one case “the content wasn’t fully clear”, because the “communicative intention” and “the message” was not understood by this subject, perhaps because of a “poor planning of the source text and the cryptical structures used by the speakers, which might have made it difficult to restructure the interpreted text”. According to a respondent, the “interpretation was clear, despite some hesitations on the part of the interpreter”, who “stuttered a little bit, but this did not really hinder

comprehension”. “Unfortunately”, says this subject, “I only listened to the interpreter’s voice and tone, without paying attention to the content of Obama’s speech”. A respondent noted the use of “a lot of filled pauses and stock phrases”. A subject noted a “good use of prosody”. Another subject underlined the presence of “some false starts and restructurings in the middle of the sentences”. A respondent said that “the interpretation was very good, the interpreter was at ease and even if he hesitated a bit at a certain point, he did not lose the thread of the speech and rendered the sentence in a coherent way”.

#### *4.7.2.3 Open-ended question on Video 3*

A respondent “found it difficult to grasp the words” at the beginning, “then the interpretation improved, but worsened again towards the end”. In one case “this was the most comprehensible interpretation and the one that inspired more confidence, because it did not present neither the excessive speed of the first one nor the hesitations of the second one”. A subject noted that “the redundancy of the interpretation made me assume that there was a generalization, because in some points the interpreter repeated himself but remained vague”. A respondent wrote that “sometimes the interpreter’s voice sounded a bit subdued, which could have given the audience a feeling of insecurity”. In the same line, another respondent was “struck” by the “interpreter’s voice”, because s/he found it “very unexpressive, monotonous and little communicative”. Similarly, another subject stated that “the voice was a little bored/not very engaging, sometimes annoying”. Two subjects underlined the “poor pronunciation” of “words”. According to a respondent, “the interpreter didn’t articulate with distinction, making it difficult to understand and affecting the assessment of the content: some sentences were incomprehensible and were pronounced too fast. Maybe for these reasons, or maybe not, another respondent stated that “following the interpreter’s speech turned out to be a difficult task – especially considering a longer duration than that considered in this study”. According to a subject, “though the interpreter’s sounded less clear than the previous video, the content [in this video] was more comprehensible, and there were less hesitations”. In one case, “the first part of the video was difficult to follow and understand [...], because of the “[high] volume” of the English original speech, and the “fragmented” Italian interpretation. On the contrary, another subject found that “the interpreter appeared hesitant only at the end; while in the initial and

central parts managed to follow the speaker and didn't have notable hesitations". A subject found that "the interpreter omitted several pieces of information, especially in the final part, where an entire sentence by McCain was interpreted using few words". A respondent "noticed the presence of calques from the English which were apparent, even without listening to the source text".

#### *4.7.2.4 Open-ended question on the whole questionnaire*

According to a respondent, "for a better evaluation, listening to different interpretations of the same excerpt would have helped".

A subject stated that "some questions were a little vague", namely the "questions 7 and 14", s/he also wondered the "difference between the question 14 and question 13". Another subject "found it difficult to answer the questions on: i) the monotonous/melodious voice of the interpreter and; ii) the synchronization of images with the interpreter's speech".

A respondent recognized that s/he "did not answered the questions on the content" (comprehension test) related to the first video excerpt s/he watched (video 2), because s/he "focused on the interpretation only and not because the interpreter was not clear"; nonetheless, s/he answered the same questions related on the following videos (1 and 3, because s/he "knew that [s/he] would have found questions on the content", therefore "[s/he] also focused on that aspect while listening".

Three subjects complained about the intensity bar as evaluation scheme. One found it "too confusing", saying s/he would have preferred "a multiple choice question". One considered it "a bit misleading"; therefore, "a numerical rating scale would have been better off". One found that the "written explanation of the intensity bar was not very clear; this same respondent stated that the "question no. 13 did not appear to be related to the videos", and that "the part related to the text comprehension was very interesting (and brilliant)".

Three subjects stated that the duration of the video excerpts was "too short" for such an evaluation, meaning the number of criteria to be assessed. Two respondents also complained about the "too short" duration of video excerpts, but only for the difficulty to "understand" the "meaning of the speech". One of these two respondents also admitted that "maybe it was difficult to answer correctly to all of them"; nevertheless, s/he stated that "the questions were complete and exhaustive". Two more subjects noted that the

“questionnaire” or the “research”, “appear[ed]” or “seem[ed]” to be “exhaustive”. One of these wrote that “the questions on comprehension were well focused on the main topic of the interpretation”; while the other one noted that “the questions on comprehension should [have been] placed before the others, and not at the end of the questionnaire”, because “the attention [was] absorbed by the previous questions”. In Agreement with this subject, another respondent also suggested “to move the questions on comprehension at the beginning of the questionnaire”.

A subject admitted that “some answers to the questions on comprehension might appear to be inconsistent with the assessment of the interpretation, because the videos were very short”; therefore, “it was difficult [for her/him] to follow the thread of the speech in the middle of the debate”. In the same line, a subject recognized that “it was not always easy to understand immediately the topic of the video excerpts”.

A respondent found it “difficult to distinguish between *aggressive* and *sweet* with reference to the interpreter’s voice; and that a second listening might have allowed her/him to be more precise in selecting the answers”.

According to a subject, “none of the interpretations turned out fully comprehensible”.

A respondent wrote that s/he “would have paid more attention to the coherence and the cohesion of the interpreter’s speech”. Another subject noted that “a question on the consistency of the sentences should [have been] included”.

A respondent complained about the lack of a question on the “tone (volume) of the interpreter’s voice”, because s/he “found it too loud in the first video and too low in the second one”.

A subject complained about the “lack of questions on factors of distraction in the video”.

Another respondent “[would have added] questions on the content, the completeness and usability of the message”.

### **4.7.3 Assessment of parameters**

Considering the nature of variables, which were all ordinal, mode and median were considered first; afterwards, also means were considered. It is worth to remember that the questions related to the effective comprehension were multiple-choice questions, but the relative answers were artificially turned into ordinal data, to make them comparable with the ordinal scale of



the evaluation scheme of other question, i.e. a scale from 1 to 7 points. Thus, of the three choices that respondents could select, the correct one was arbitrarily assigned a value of 7 (the maximum of the scale); the less correct one was assigned the value of 4 (central in the scale); the wrong choice was assigned the value of 1 (the minimum of the scale). However, results from the statistical analysis of data (see below) show the limits of such procedure, because of the dissimilarity of the data obtained through this artificial and arbitrary scales (1/4/7 points), and those obtained through the natural scales (1-7 point).

#### 4.7.3.1 Modes and medians

The analysis of ratings was based on measurements of descriptive statistics, i.e. modes, medians and means. The mode is the most frequent value of the scale (in this case, 1-7 point) assigned by respondents to each variable, represented by the item of the questionnaire. The median is the central value among those selected by subjects, which does not necessarily coincide with the central value of the scale presented; for example, in the 1 to 7 point scale the central value is 4, but subjects may assess the parameter choosing value from 4 to 6, consequently, in this case the central value (median) is 5.

Considering the nature of variables, which were all ordinal, including those “rendered” ordinal (see above), mode and median were taken into a greater consideration for analysis. Mode and median for each variable were compared with the values of linguistic units and indices reported in table 4.4 and described in section 4.5.2.3.2 (see above).

Figures show that *articulation* in video 1 received the same most frequent rating as video 2 (mode=6), but the central value of video 2 (median=6) is one point higher, meaning that the latter was considered slightly better by respondents. These figures appear in contrast with the *articulation rate index* for the three videos, because the lowest value was that in video 2 (4.6), while the highest value was that in video 1 (6.4); this datum seems consistent with assessment. It may be supposed that perception of articulation was influenced by intonation, since the value of the tonal range in video 2 was the highest (18.5), followed by that in video 1 (17) and finally that in video 3 (16.4).

The variable *hesitations* was rated in the same way in the three video excerpts (mode=5); with only a difference represented by video 3, which could have been perceived as little less hesitant, because its median (=4) was

one point lower than that of videos 1 and 2 (=5). Also these data do not seem consistent with the *disfluencies rate index*, which was 4.5 in video 1, 16.6 in video 2 and 6.1 in video 3.

Item / variable	Video 1			Video 2 (experimental)			Video 3			Means all videos
	Mode	Median	Mean	Mode	Median	Mean	Mode	Median	Mean	
Articulation	6	5	4.78	6	6	5.36	3	4	3.98	4.71
Hesitations	5	5	4.2	5	5	4.45	5	4	3.68	4.11
Audible breaths	2	2	2.55	2	2	2.26	2	2	2.72	2.52
Silent-pauses	2	3.5	3.64	2	3	3.31	5	4	3.55	3.51
Speed of speech	6	6	5.82	4	4	4.49	4	4	3.66	4.66
Melodious voice	4	4	3.84	5	5	4.49	2	3	3.05	3.80
Same melody repeated	6	4.5	4.50	5	5	4.32	5	5	4.49	4.43
Sweet voice	2	3	3.13	5	4.5	4.45	5	4	4.53	4.03
Self-confident voice	5	4.5	4.31	6	4	4.17	5	4	4.17	4.22
Active personality	5	5	4.78	5	5	4.83	3	3	3.29	4.31
Expressive voice	5	5	4.47	5	5	4.88	3	3	3.18	4.18
Comprehensible voice	5	5	4.99	5	5	5.16	5	5	4.41	4.85
Sound-image synchrony	3	4	3.99	5	5	4.60	4	4	4.53	4.36
Interpretation accounting for image	4	4	4.31	5	5	4.77	4	4	4.41	4.48
Credible interpreter	5	5	4.71	6	5	4.67	5	5	4.61	4.65
Skilled interpreter	5	5	4.77	5	5	4.75	5	5	4.85	4.79
Complex words	3	3	3.26	3	3	3.29	3	3	2.87	3.13
Natural syntax	4	4	4.03	3	4	4.16	5	5	4.56	4.26
Complex sentences	3	3	3.26	4	3	3.35	4	3	3.09	3.23
Comprehensible interpretation	5	5	4.49	6	5	4.58	5	5	4.95	4.68
Effective comprehension 1	7	7	6.30	7	7	4.72	7	7	6.32	5.78
Effective comprehension 2	4	4	5.01	4	4	4.12	7	7	5.87	5.0
Effective comprehension 3	7	7	5.96	7	7	6.33	7	7	6.59	6.30
Means all variables			4.40			4.41			4.22	4.34

Standard deviation:  $0.84 < \sigma < 1.98$

**Table 4.6.** Values of descriptive statistics (mode, median and mean) related to the assessment of parameters in the pilot survey.

The variable *audible breaths* presents the same predominant ratings in the three videos (mode=median=2). These data appear to be consistent with the *total number of audible breaths* perceptually identified by the researcher, which were 13 for videos 1 and 2, and 11 for video 3; However, considering that the perceptual data of respondents is a low value (=2), it may be assumed that audible breaths were not perceived, or at least were not perceived as such, but as silent pauses, given the similarity between the values of mode and median.

The *Silent pauses* were assessed almost in the same way both in videos 1 and 2 (mode=2; median video 1=3.5; median video 2= 3), and differently in video 3 (mode=3; median=4); it may be supposed that in video 3 a higher number of silent pauses was perceived. This perception appears to be in part consistent with the *total number of silent pauses* counted in the three videos, which were 5 in video 1, 8 in video 2 and 11 in video 3.

The *Speed of speech* was rated in the same way in both videos 2 and 3 (mode=median=4) and differently in video 1 (mode=median=6). This perception appears to be totally consistent with: i) the *speech rate index 2 (syllable/sec)*, the values of which were 4.6 for both videos 2 and 3, and 6 for video 1; and ii) the *speech rate index 1 (words/min)*, the values of which were 136 in video 3, 140 in video 2 and 173 in video 1.

The *melodious voice* received different most-frequent rates in the three videos: the voice considered the most melodious one was that in video 2 (mode=median=5), followed by those ones in video 1 (mode=median=4) and video 3 (mode=2 and median=3). These results appear to be consistent with the values of *tonal range* measured in the three videos, where the value of video 2 was the highest (18.5), followed by that of video 2 (17) and video 3 (16.4).

The variable *same melody repeated* received the same predominant ratings in both videos 2 and 3 (mode=median=5); while that in video 1 was considered less melodious (mode=median=4). These rating pattern is similar to that of *speed of speech* (see above).

The parameter *sweet voice* was rated in the same way in both videos 2 and 3 (mode=5; median video 2=4.5; median video 3=4); while voice in video 1 was considered less sweet (mode=2; median=3).

The variable *self-confident voice* received almost the same ratings in both videos 1 and 3 (mode=5; median video 1=4.5; median video 3=4); while the voice of video 2 was considered more self-confident (mode=6; median=4).

The variables *active personality* and *expressive voice* received equal predominant ratings in the three videos; in detail, videos 1 and 2 received

higher predominant ratings (mode=median=5) than video 3 (mode=median=3). These data show that the two variables are strictly related.

The variable *comprehensible voice* received the same predominant rating for the three videos (mode=median=5). Contrary to what could be expected, results from *expressive voice* do not coincide with those from *comprehensible voice*, even though the only difference is represented by video 3.

The predominant ratings of variables *sound-image synchrony* and *interpretation accounting for image* show almost the same values in the three videos; in detail, the highest values were received by the video 2 (mode=median=5, in both variables), followed by the values in video 3 (mode=median=4, in both variables) and finally in video 1 (mode *sound-image synchrony*=3; mode *interpretation accounting for image*=4; median in both variables=4). The slight difference (1 point) in the mode of the two variables related to video 1 could be notable. Indeed, the variable *sound-image synchrony* was thought to elicit the acceptability of *décalage*; while the variable *interpretation accounting for image* was thought to elicit the interpreter's ability to manage the turn-taking of speakers, their body language, and the sequence of camera takes or shots. In this case, the duration of video excerpts (1 min) did not contain any sequence of camera takes, but a single shot, a medium close-up, of the speaker; this could be the reason why the two variables received almost the same ratings in the three excerpts. The rhythm of subsequent camera takes or shots in a sequence could influence perception of speech rhythm (Veneziano 2013, personal communication). Considering the variable *sound-image synchrony*, the video that received the highest dominant rating was the video 2, i.e. the experimental one, where the synchrony was artificially created; video 2 was also the video with the highest *tonal range*; the second dominant rating was that in video 3, while the lowest dominant rating was that in video 1, i.e. the one with the highest speech rates (both word/min and syllable/sec), but also with the lowest non-fluencies rate index.

Considering the predominant rating, the most *credible interpreter* was that in video 2 (mode=6; median=5), followed by those in the other two videos (mode=median=5, in both). This rating pattern was similar to that of *self-confident voice* (see above).

The variable *skilled interpreter* received the same dominant rating in the three videos (mode=median=5) and the same rating pattern of the variable *comprehensible voice*.

The variable *complex words* also received the same predominant rating in the three videos (mode=median=3).

Differently, *natural syntax* received three different predominant ratings in the three videos; the syntax considered the most natural one was that in video 3 (mode=median=5), followed by those in videos 1 (mode=median=4) and 2 (mode=3; median=4).

The variable *complex sentences* received the same dominant ratings in both videos 2 and 3 (mode=3; median=4); while the sentences in video 1 were considered simpler (mode=median=3).

The variable *comprehensible interpretation* received the same dominant ratings in both videos 1 and 3 (mode=median=5), while the interpretation of video 2 was considered more comprehensible (mode=6; median=5). These figures are the same as those of variable *credible interpreter*.

For what concerns the comprehension test, in all the videos, the questions related to the general comprehension of the whole interpretation (*effective comprehension 1*), and the second part of it (*effective comprehension 3*), received the highest predominant value of the scale (7), assigned to the correct answer of the multiple choice question by the researcher. Differently, the questions related to the comprehension of the first part of the interpretation (*effective comprehension 2*), received a lower predominant value (4), the one assigned to the less correct answer in the multiple choice question by the researcher. However, this is true only for videos 1 and 2, because video 3 received a predominant rating of 7 also in this question. This difference in the results of the questions of comprehension might be due to the fact that the comprehension of the first part was difficult to recall, more than to comprehend, considering that a few subjects, in the final free comment question, complained about the final position of the comprehension test in the questionnaire flow, because some subjects found it difficult to recall the topic after answering all the other questions.

#### 4.7.3.2 Means

The analysis of means of the ratings assigned to the questions, for each of the three videos, shows that video excerpt 1 obtained the highest means for: *speed of speech* (5.82); *same melody repeated* (4.51, slightly above the value of video 3: 4.49); *self-confident voice* (4.31, slightly above the values of videos 2 and 3: both 4.17); and *credible interpreter* (4.71, slightly above the value of videos 2 and 3: 4.67 and 4.61 respectively). The video excerpt 2 obtained the highest

means for: *articulation* (5.36); *hesitations* (4.45); *melodious voice* (4.49); *active personality* (4.85; slightly above the value of video 1: 4.78); *expressive voice* (4.88); *comprehensible voice* (5.17); *sound-image synchrony* (4.62); *image-interpretation synchrony* (4.77); *complex words* (3.29, almost the same as the value of video 1: 3.27); *complex sentences* (3.36, slightly above the value of video 1: 3.26). The video excerpt 3 obtained the highest means for: *silent pauses* (3.57, almost the same as the value of video 1: 3.55); *sweet voice* (4.53, slightly above the value of video 2: 4.46); *skilled interpreter* (4.85, slightly above the value of video 2: 4.75); *natural syntax* (4.56); *self-assessment of comprehension* (4.95); and all the three questions on *effective comprehension*: 6.33 for the first question (almost the same as the value of video 1: 6.31), 5.88 for the second question and 6.59 for the third question.

According to the above-mentioned data, the interpretation of video 1 presents: the highest *speed of speech*; the *same melody repeated*, possibly due to the peculiar “singsong” of SI detected by Ahrens (2005): the most *self-confident voice*; and the most *credible interpreter*, even though the three last criteria mentioned do not show a remarkable difference from the interpretations in the other two videos. However, the interpretation of video 2 was perceived to have: the best *articulation*, notwithstanding the many *hesitations*; the best *voice* (melodious, expressive, comprehensible), revealing the most *active personality* of the interpreter, even though it was not more active than that in video 1; it seems that the best *voice* of video 2 also allowed the respondents to perceive the most *complex words* and *sentences*, with respect to those in video 1. This *voice* was also the more suitable one for the video, considering the best *sound-image-interpretation synchrony*. According to respondents, the interpretation in video 3 had almost the same number of *silent pauses* than that in video 1, and a voice slightly sweeter than that in video 2; the interpreter in video 3 was considered slightly more *skilled* than that of video 2, and used the most *natural syntax*; all these aspects led respondents to perceive the interpretation in video 3 as the most *comprehensible interpretation*, and effectively it was the one that was most comprehended.

It is worth to notice that, although some respondents in the open-ended questions claimed that the duration of video excerpts was too short to assess all the criteria of the questionnaire (see above – sections 4.7.2.1 to 4.7.2.4), some linguistic aspects of the interpretations that were measured (see above – table 4.4) were detected by subjects. This is the case of the highest speed of speech in the video 1; the highest tonal range in the video 2; the simplest topic, the highest number of silent pauses, and the lowest tonality of the

intonation in the video 3, which was also the most comprehended of the three.

#### 4.7.3.3 *Similar response patterns (analysis of frequencies of ratings)*

For a deeper understanding of the results from the assessment, an analysis of frequencies of ratings was carried out. Frequencies of ratings show the percentages of respondents who assigned a value of the rating scale (1-7 points) to each variable in each of the three videos. The ratings of each variable related to the three videos were grouped together and a series of graphs were plotted (stacked bars charts), so that each plot shows the ratings of each variable in each the three videos; in total 23 graphs were plotted, corresponding the number of variables, including the three questions of the comprehension test<sup>5</sup>. The graphs are reported in Appendix 10. In this section, only the plots showing similar response patterns are reported. The similarity of response patterns was identified graphically on the basis of the histograms of the graphs. This approach, based on the homogeneous forms identified among the 23 graphs, is consistent with the theoretical paradigm of the research, based on the gestalt perception of speech (see chapter 3).

As hypothesized in the analysis of modes and medians, the variables *audible breaths* and *silent pauses* show a similar pattern of response (see below – figure 4.18); in fact, the main difference is in a higher percentage of ratings of the lower values of the scale (mainly points 1 and 2) for *audible breaths*, against a higher percentage of higher values (points 3, 4 and 5) for *silent pauses*; however the assessment trend for the three videos is quite similar.

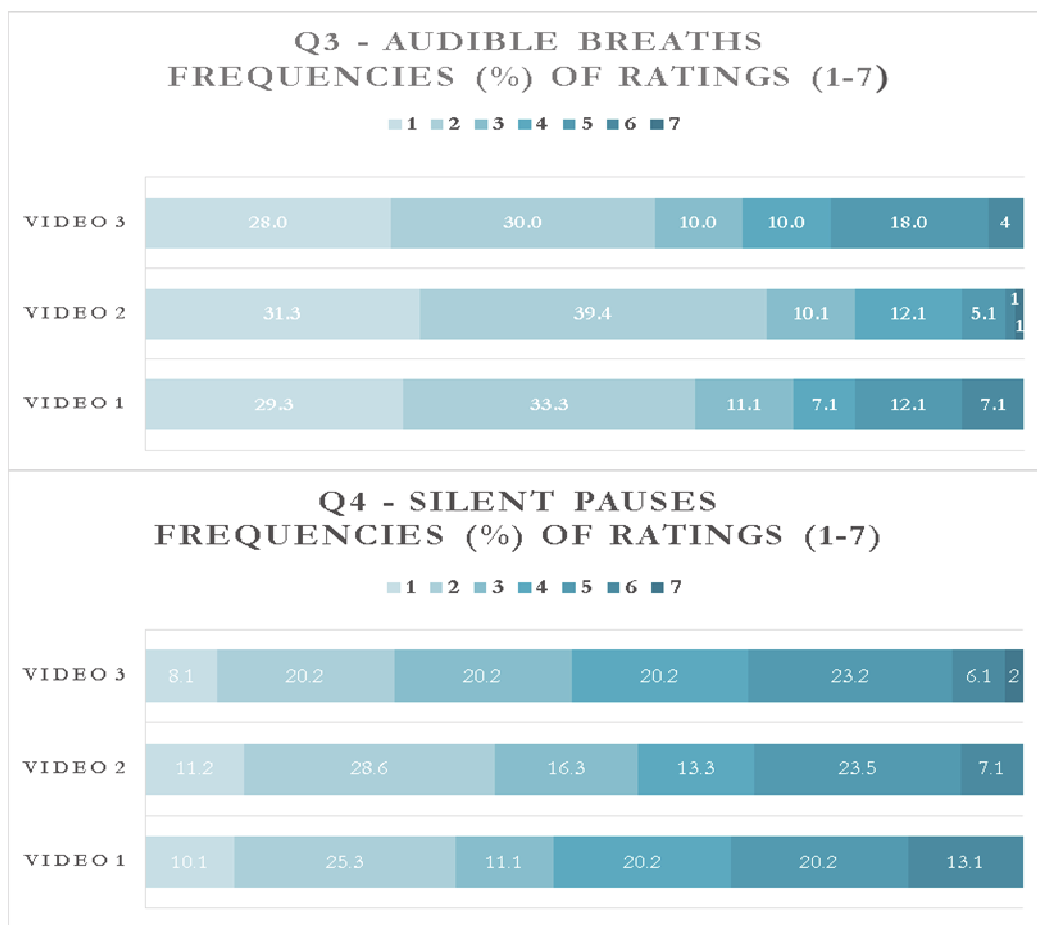
Another similarity of response patterns was found among the variables *melodious voice*, *active personality* and *expressive voice* (see below – figure 4.19). The similarity between *melodious voice* and *active personality* can be seen in the distribution of the percentages of ratings in the 7 points of the scale, where there is a slight difference between (in *active personality*) the highest percentage (37.4) in the point 5 in video 2, and the highest percentage (34) in the point 4 in video 1. Both these highest percentages, with respect to the same percentages of the same points in the same videos, but in variable *melodious voice*, ‘push to the left’ the frequencies of lower points of the scale.

The response pattern of the variable *expressive voice* is more similar to that of *active personality* than to that of *melodious voice*. There are some very little

---

<sup>5</sup> The plotting of the frequencies graphs was realized with the help of Doctor Chiara Zanetti (Università di Trieste).

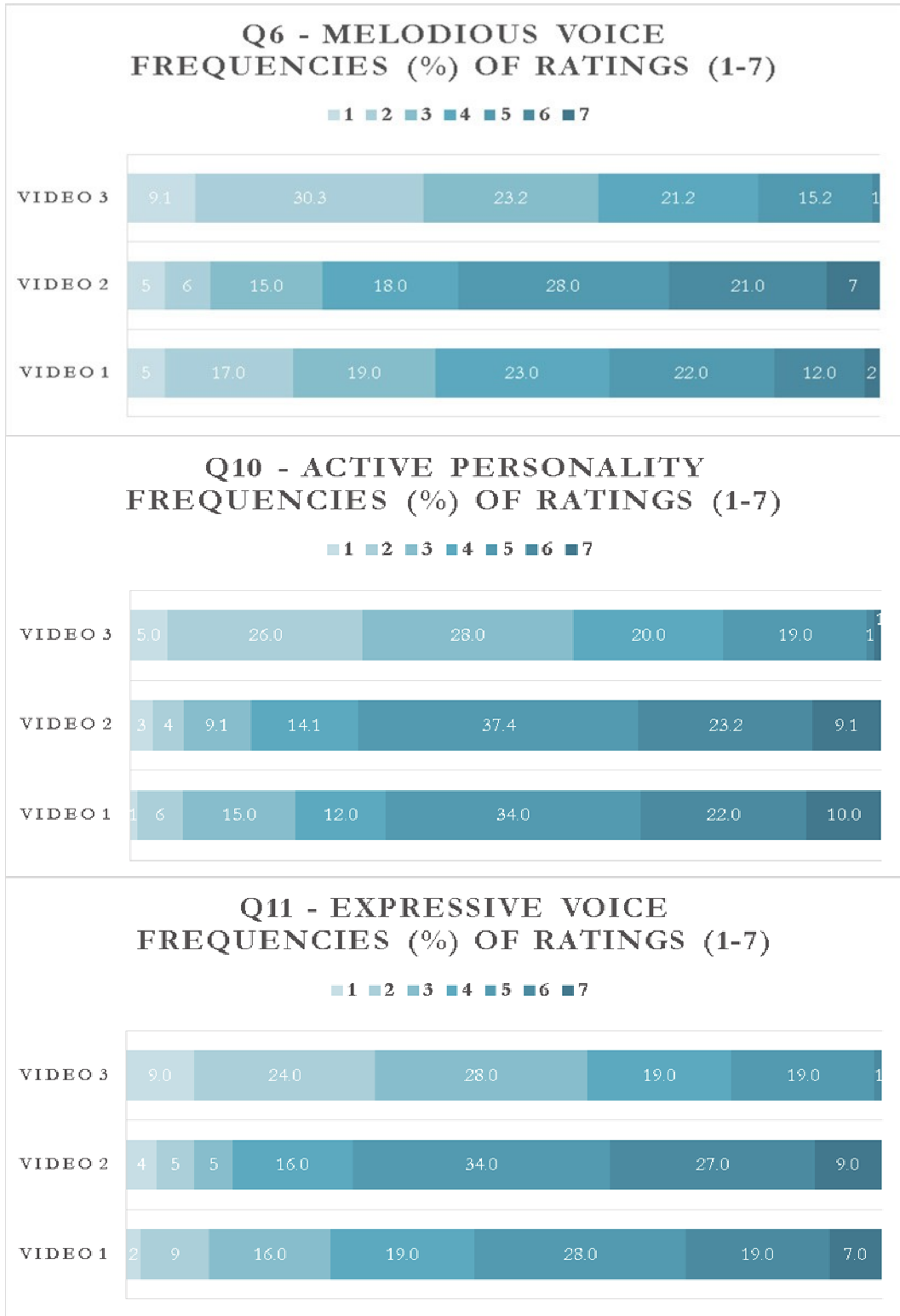
differences, one of them is represented, in *expressive voice*, in video 3, by the higher percentages of the points 1, 2, 3 and 4, with respect to the same points, in the same video, in the variable *active personality*. The variables *active personality* and *expressive voice* received the same predominant ratings, as observed in section 4.7.3.1 (see above).



**Figure 4.18.** Similar response pattern found in 2 out of 23 plots showing the ratings assigned to each variable for the three videos.

The variables *complex words* and *complex sentences* received different predominant ratings in the assessment of the three videos (see section 4.7.3.1 – above); however, frequency of ratings show a similar response pattern (see below – figure 4.20), even though the distribution of ratings among the different points of the scale presents some differences. In detail, in *complex sentences*, in the three videos, the percentages of the central point of the scale, i.e. 4, are higher than those of *complex words*.

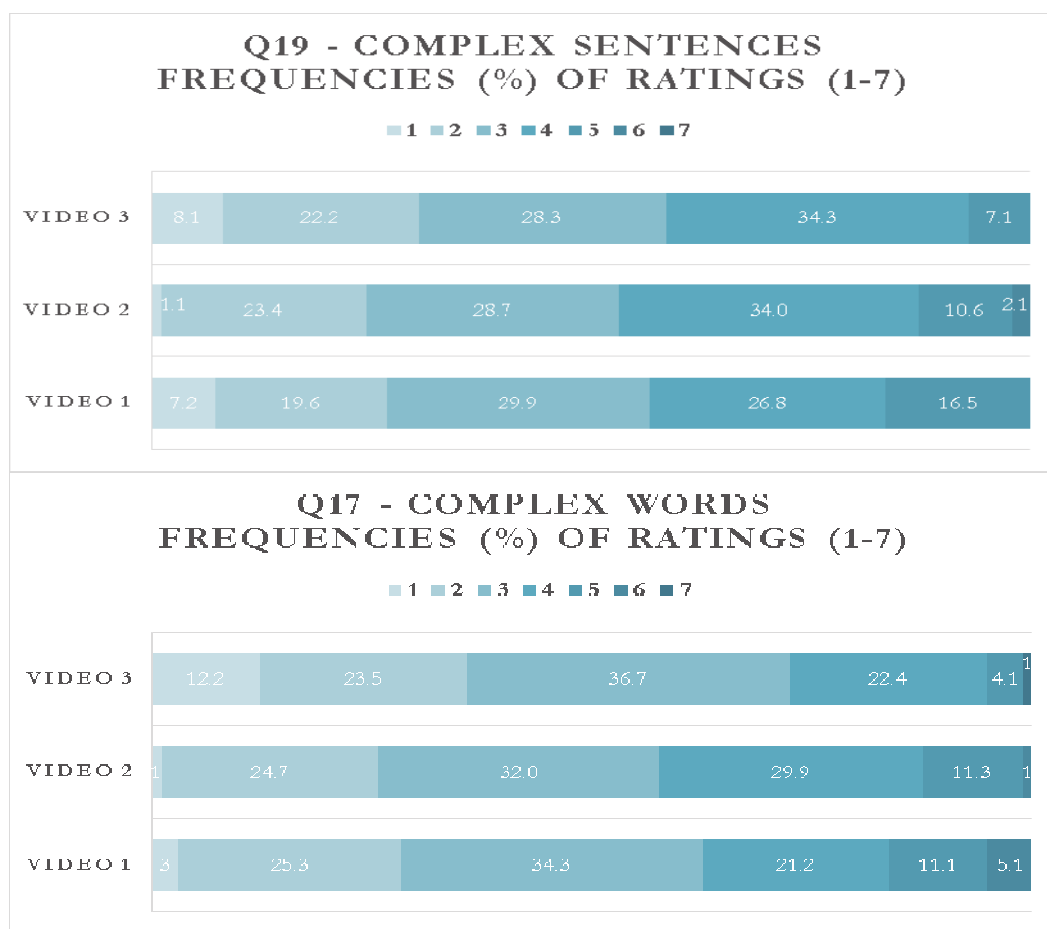




**Figure 4.19.** Similar response pattern found in 2 out of 23 plots showing the ratings assigned to each variable for the three videos.

The variables *credible interpreter* and *skilled interpreter* present a similar response pattern (see below – figure 4.21), where the main difference is

represented by the highest percentage of point 5 in video 2 in the variable *skilled interpreter*, while in *credible interpreter* the highest percentage is in point 6.

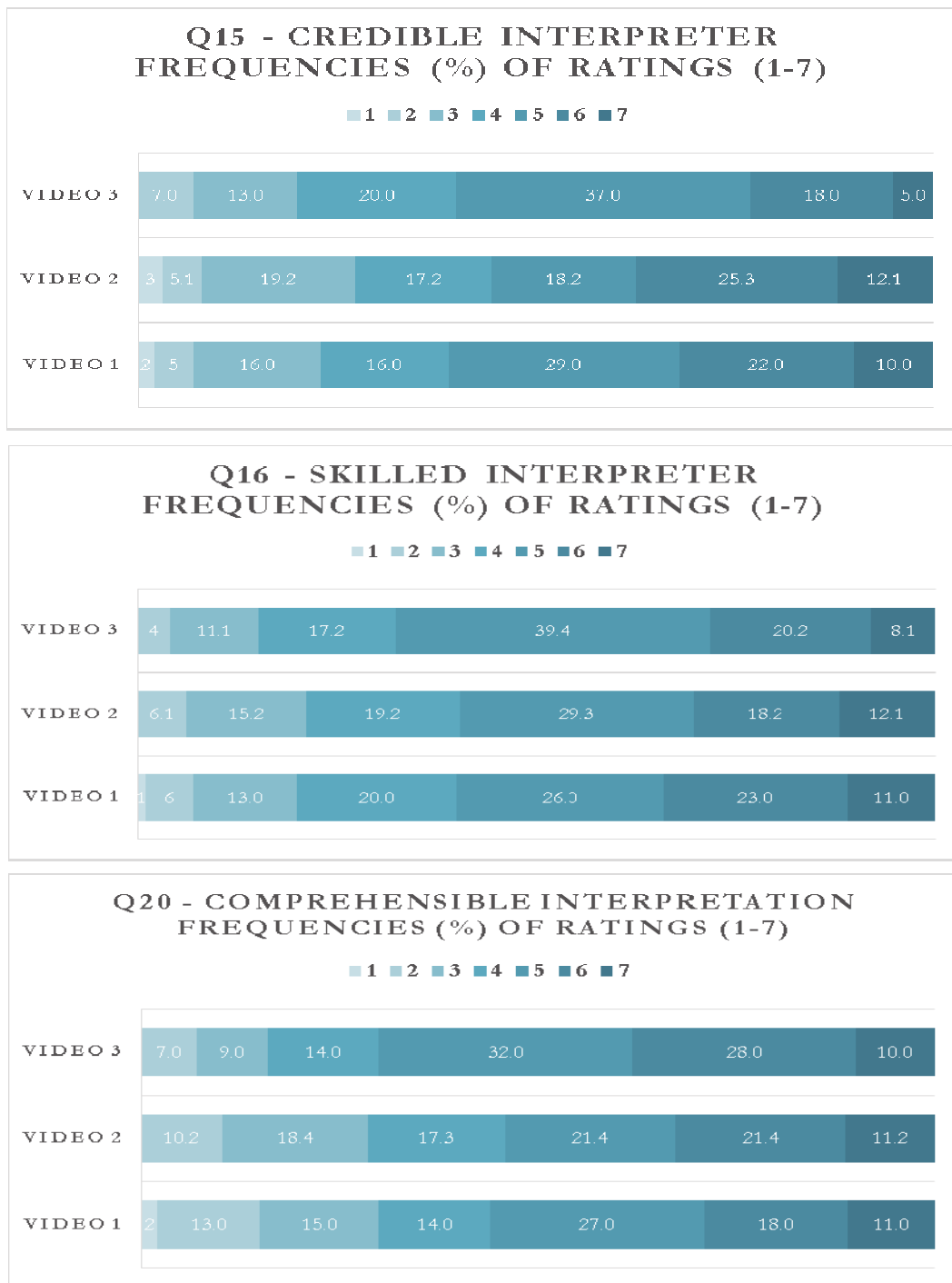


**Figure 4.20.** Similar response pattern found in 2 out of 23 plots showing the ratings assigned to each variable for the three videos.

The variables *credible interpreter* and *comprehensible interpretation* present the same predominant and central values (see above – section 4.7.3.1). They also present similar frequencies of ratings, where the main difference is represented by the highest percentage of the point 5, in video 3 (28%), in *comprehensible interpretation*, against the same percentage, in the same video, in *credible interpreter* (18%).

As stated before, the comprehension test was made of three multiple-choice questions, and the answers were artificially adapted by the researcher to an ordinal scale made of three points, where the point 1 corresponded to the incorrect answer, the point two corresponded to an answer that could be considered right but not as correct as the answer corresponding to the point

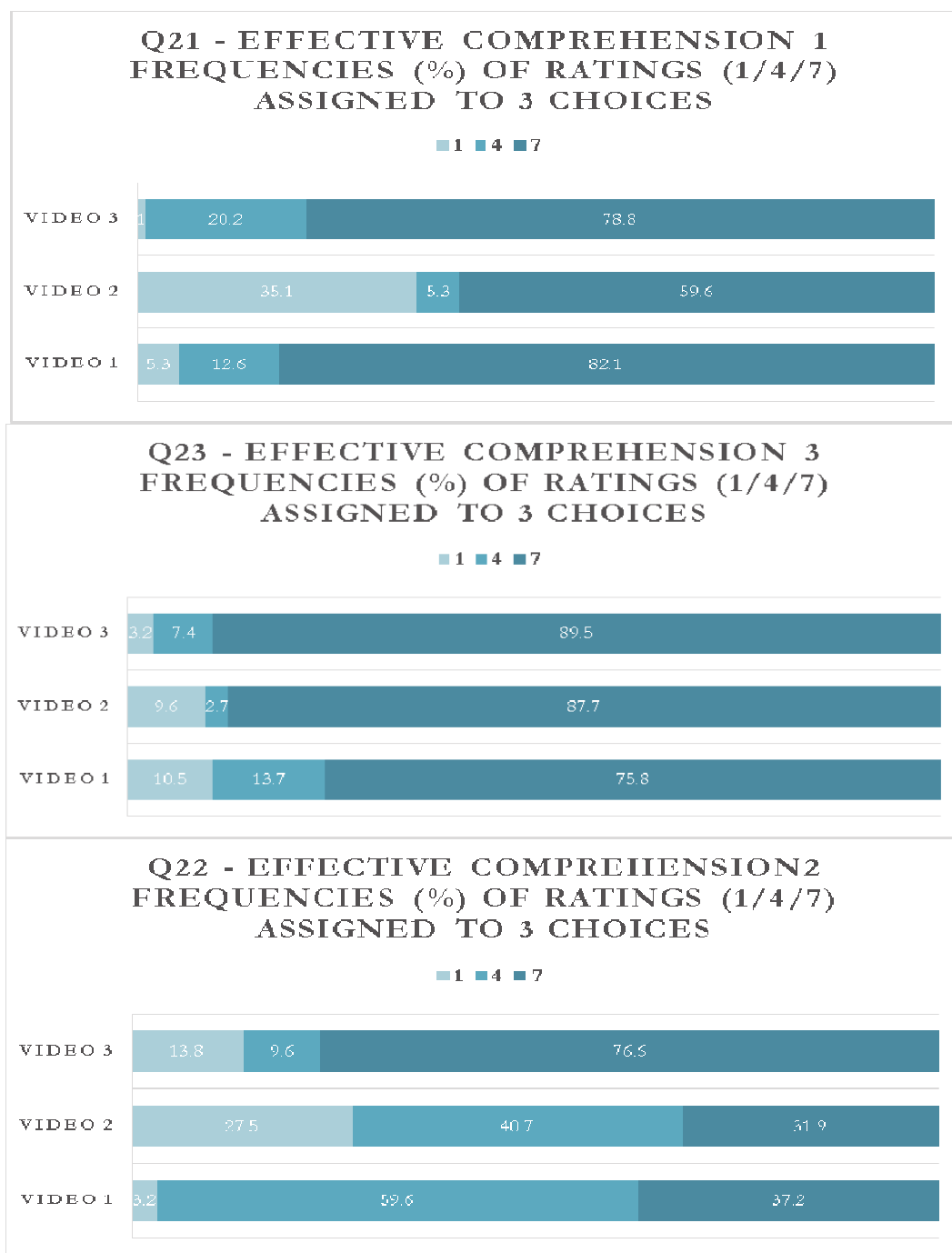
3. Hence, the difference with the previous questions, which had a 1 to 7 point ordinal scale as evaluation pattern.



**Figure 4.21.** Similar response pattern found in 3 out of 23 plots showing the ratings assigned to each variable for the three videos.

With this in mind, it can be observed that the questions *effective comprehension 1* and *effective comprehension 2* show a similar response pattern; even

though *effective comprehension 2* presents as predominant value the point 4 in both videos 1 (59.6 %) and 2 (40.7 %) (see below – figure 4.22). Differently, the answer *effective comprehension 3*, just as *effective comprehension 1*, shows the percentage of point 7, corresponding to the correct answer, as the predominant value.



**Figure 4.22.** Response patterns among *effective comprehension 1*, *2* and *3*, found in 3 similar questions out of 23 plots showing the ratings assigned to each variable for the three videos.

The percentages of the point 7 in *effective comprehension 3* shows that an overwhelming majority of respondents selected the correct answer for the three videos; these data poses serious questions about the validity of the question (Delli Zotti 2015: personal communication).

#### *4.7.3.4 Interrelations among variables*

The analysis of interrelations among variable was carried out within the statistical treatment of data, operated by the software *SPSS*; in particular, through the methods of bivariate correlations of variables, perceptual map by multidimensional scaling, hierarchical clusters and analysis of principal components.

##### *4.7.3.4.1 Bivariate correlations of variables*

Considered that some variables in the three videos show similar or homogeneous response patters, as seen through the analysis of mode and median, and the analysis of frequencies; in order to find statistically significant one-to-one variable relations, an analysis of bivariate correlations was conducted, through *SPSS*. A matrix scatter graph was plotted for each of the three videos<sup>6</sup> (see Appendix 11). From a general overview of each single scatter graph, it was noted that the majority of relationships between two variables, and not all of them, were linear. Therefore, both table of bivariate correlations based on Pearson's coefficient and a table based on Spearman's coefficient were produced with *SPSS*, for each of the three videos. In each of the six tables (two multiplied by three videos), Pearson's and Spearman's correlations significant at 0.01 and 0.05 levels (respectively  $p < 0.01$  and  $p < 0.05$ ) were highlighted and checked against each other, for each video. It was found that the majority, and not all, of the two kinds of significant (both at 0.01 and 0.05 levels) correlations coincided. The cases that did not match, significant or not, were considered, and this time the two correlations coefficients were checked against the scatterplot: if this showed a linear relationship between variables, then Pearson's correlation was selected, being that significant or not; conversely, if the scatterplot showed a non-linear relationship between the two variables, then Spearman's coefficient was selected, being that significant or not. The result of this work was summarized

---

<sup>6</sup> The plotting and the interpretation of scatter graphs was realized with the help of Prof. Giovanni Delli Zotti (Università di Trieste), who also collaborated to the design of the questionnaire.

in the table below (4.23), while the matrix scatterplots and the correlations tables are reported in Appendix (11).

**Correlations of variables**

**Bivariate (2-tailed) - Pearson's and Spearman's coefficients both significant at 0.01 and 0.05 levels - SPSS**

**Positive and negative ("+" / "-") correlations of two variable in each of the three videos ("1 / 2 / 3"), significant at 0.01 level ('underlined') and 0.05 level**

Melodious voice	Same melody repeated	Sweet voice	Self-confident voice	Active personality	Expressive voice	Comprehensible voice	Sound-image sync	Image-Int sync	Credible interpreter	Skilled interpreter	Complex words	Natural syntax	Complex sentences	Comprehensible interpret	Effective compreh 1	Effective compreh 2	Effective compreh 3	
1 2 3		2	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3	1 2	1 2 3	1 2 3	1	1 2 3		1 2 3		2	2	Articulation
1			-1 -2 -3	-1 -2 -3	-1 -2	-1 -2 -3	-1 -2 -3	-1 -2	-1 -2 -3	-1 -2 -3	-1 -2	-1 -2 -3		-1 -2 -3			-2	Hesitations
-2			-1 -2 -3	-1 -2	-2	-1 -2 -3	-2		-2	-2		-2		-2			-2	Audible breaths
1		-1 -2	-2 -3			-1 -2	-1	-2	-1 -2 -3	-1 -2 -3		-1 -2 -3		-1 -2				Silent pauses
		-1 -2	2 3	2	-1				2	-2				-1				Speed of speech
		1 2 3	1 2 3	1 2 3	1 2 3	1 2 3	1	1	1 2 3	1 2 3		1 2 3		1 2				Melodious voice
			2				1	1			2		2					Same melody repeated
1 2 3				2	1 2	1 2 3 1	1	1	1	1		1 2 3		1 2				Sweet voice
1 2 3				1 2 2	1 2 3	1 2 3 1 2 2 1	1 2 2	1 2	1 2 3	1 2 3	1 2	1 2 3	1	2 3				Self-confident voice
1 2 3		2	1 2 3		1 2 3	1 2 3 1 2 3	1 2 3	1 2	1 2 3	1 2 3	1	1 2 3	1	2 3			2	Active personality
1 2 3		1 2	1 2 3	1 2 3		1 2 3	1 2 3	1 2 3	1 2 3	1 2 3	1	1 2 3	1	2 3				Expressive voice
1 2 3		1 2 3	1 2 3	1 2 3	1 2 3	1 2 3	1 2	1 2	1 2 3	1 2 3		1 2 3		1 2 3			2	Comprehensible voice
1	1	1	1 2 3	1 2 3	1 2 3	1 2		1 2 3	1 2 3	1 2 3	2	1 2 3		1 2 3				Sound-image sync
1	1	1	1 2	1 2	1 2 3	1 2	1 2 3		1 2 3	1 2 3		1 2 3	1	2 3				Image-Int sync
1 2 3		1	1 2 3	1 2 3	1 2 3	1 2 3 1 2 3	1 2 3	1 2 3		1 2 3	1 2	1 2 3	1	2 3				Credible interpreter
1 2 3		1	1 2 3	1 2 3	1 2 3	1 2 3 1 2 3	1 2 3	1 2 3	1 2 3		1 2	1 2 3	1	2 3				Skilled interpreter
	2		1 2	1	1		2		1 2	1 2		1 2	1 2 3	2				Complex words
1 2 3		1 2 3	1 2 3	1 2 2	1 2 3	1 2 3 1 2 2 1	1 2 2	1 2 3	1 2 3	1 2 3	1 2		1	2 3				Natural syntax
	2		1 3	1 3	3			1	1	1	1 2 3							Complex sentences
1 2		1 2	1 2 3	1 2 3	1 2 3	1 2 3 1 2 3	1 2 3	1 2 3	1 2 3	1 2 3	2	1 2 3					2	Comprehensible interpretation
																		Effective compreh 1
																	3	Effective compreh 2





From a first general overview to the table of bivariate correlations of variables in the three videos, it is evident that the variables *articulation*, *hesitations*, *self-confident voice*, *active personality*, *expressive voice*, *comprehensible voice*, *sound-image synchrony*, *image-interpretation synchrony*, *credible interpreter*, *skilled interpreter*, *natural syntax* and *comprehensible interpretation* present a high number of correlations in the three videos; these variables are strongly interrelated. There are only few cases in which the correlations among the above mentioned variables are not present in all the three videos, but only in videos 1 and 2 (and not in video 3); these cases are the correlations between: *expressive voice* and *hesitations*, *comprehensible voice* and *sound-image synchrony*, *comprehensible voice* and *image-interpretation synchrony*, *self-confident voice* and *image-interpretation synchrony*, *active personality* and *image-interpretation synchrony*, *image-interpretation synchrony* and *articulation*, *image-interpretation synchrony* and *hesitations*. Taking each of these highly related variables as a point of reference, the correlations of each of them with other variables, in each video, is described in detail in the following paragraphs.

The variable *self-confident voice*, both in videos 2 and 3, was negatively correlated to *silent pauses*, and positively to *speed of speech*. Only in video 2, it was positively correlated to *same melody repeated*. Both in videos 1 and 2, it was positively correlated to *image-interpretation synchronism* and *complex words*. Both in videos 1 and 3, it was positively correlated to *complex sentences*.

The variable *active personality*, both in videos 1 and 2, was negatively correlated to *audible breaths*, and positively to *image-interpretation synchronism*, while in both videos 1 and 3, it was positively correlated to *complex sentences*. Only in video 2, it was positively correlated to *speed of speech* and *sweet voice*; only in video 1, it was positively correlated to *complex words* and *complex sentences*.

The variable *expressive voice*, both in videos 1 and 2, was negatively correlated to *hesitations* and *sweet voice*. Only in video 2, it was negatively correlated to *silent pauses*; only in video 1, it was positively correlated to *speed of speech* and *complex words*; only in video 3, it was positively correlated to *complex sentences*.

The variable *comprehensible voice*, both in videos 1 and 2, was negatively correlated to *silent pauses*, and positively to sound-image synchronism and *image-interpretation synchronism*.

The variable *credible interpreter*, both in videos 1 and 2, was positively correlated to *complex words*; only in video 2, it was negatively correlated to

*audible breaths*, and positively to *speed of speech*; only in video 1, it was positively correlated to *sweet voice* and *complex sentences*.

The variable *skilled interpreter* presents the same correlation as the variable *credible interpreter*, with only one exception, i.e. the negative correlation to *audible breaths*, related to videos 2 and 3.

The variable *natural syntax*, both in videos 1 and 2, was positively correlated to *complex words*; only in video 2, it was negatively correlated to *audible breaths*; and only in video 1, it was negatively correlated to *complex sentences*.

The variable *comprehensible interpretation*, both in videos 1 and 2, was negatively correlated to *silent pauses*, and positively to *melodious voice* and *sweet voice*; only in video 2, it was negatively correlated to *audible breaths*, and positively to *complex words*; only in video 1, it was negatively correlated to *speed of speech*.

Both the variables *complex words* and *complex sentences* present few correlations; the former in videos 1 and 2, the latter mainly in video 1. In detail, the variable *complex words*, both in videos 1 and 2, was negatively correlated to *hesitations*, and positively to *self-confident voice*, *credible interpreter*, *skilled interpreter*, and *natural syntax*; only in video 1, it was negatively correlated to *articulation*, *active personality* and *expressive voice*; only in video 2, it was positively correlated to *same melody repeated*, *sound-image synchrony* and *comprehensible interpretation*. The variable *complex sentences*, beside the positive correlation to *complex words* in the three videos, only in video 2, it was positively correlated to *same melody repeated*; in both videos 1 and 3, it was positively related to *self-confident voice* and *active personality*; only in video 1, it was positively correlated to *image-interpretation synchrony*, *credible interpreter*, *skilled interpreter* and *natural syntax*; only in video 3, it was positively correlated to *expressive voice*.

Opposite to the above mentioned group of interrelated variables, analyzed in the previous paragraphs, there is a second group of variables that show few correlations, the majority of them are present in two or just one video; these are: *audible breaths*, *silent pauses*, *speed of speech*, *melodious voice*, *same melody repeated*, *sweet voice*, *complex words* and *complex sentences*. Also the variables *effective comprehension 1*, *2* and *3* show few correlations. As stated before, these three variables, related to the comprehension test (multiple choice), were artificially rendered ordinal, in an attempt to compare the results with those of other variables. The correlations computed by the software were those

between the variable *effective comprehension 3* and other six variables; however, the question *effective comprehension 3* could be considered not valid due to the results of frequency (see above – section 4.7.3.3).

With this in mind, the description of correlations from now on will take into considerations only those cases where correlations of each variable were found in two videos or only one video, in order to pay attention to differences among videos.

#### 4.7.3.4.1.1 VIDEO 1

Only in this video, the variable *melodious voice* was negatively correlated to *hesitations* and *speed of speech*. This variable was also negatively correlated to *sweet voice*, *expressive voice* and *comprehensible interpretation*. From these data, it could be assumed that the high *speed of speech* of interpretation had a negative impact on perception of the interpreter's *expressive voice* and *comprehension of the interpretation*. Nevertheless, if *speed of speech* was negatively correlated to *sweet voice*, this variable, only in this video, was positively correlated to *sound-image synchrony*, *image-interpretation synchrony*, *credible* and *skilled interpreter*. This may suggest that, even if the *speed of speech* somehow impeded the comprehension, it was considered natural of the interpreter and also required from the high speed of speech of the speaker. Only in this video, the variables *sweet voice*, *melodious voice* and *same melody repeated* were positively correlated to both *sound-image* and *image-interpretation synchrony*. This may suggest that all these aspects were associated in perception to the telegenic voice of the interpreter and to the fact that the melodiousness (tonal range) of his voice was influenced by his high *speed of speech*, which was evident that depended on that of the speaker. The comprehension may have been impeded not only by the high *speed of speech*, but also by the high informativity of the text, considering the high number of correlations in this video between *complex words*, *complex sentences* and other variables. In detail, *complex words* was positively correlated to *articulation*, *self-confident voice*, *active personality*, *expressive voice*, *credible interpreter*, *skilled interpreter* and *natural syntax*; similarly, the variable *complex sentences* was positively correlated to *self-confident voice*, *active personality*, and, only in this video, to *image-interpretation synchrony*, *credible* and *skilled interpreter* and *natural syntax*. In the end, it may be supposed that the interpreter's performance was appreciated, and that the comprehension of the text depended more on its high informativity and the *speed of speech* of the speaker, than on the bad performance of the interpreter.

## 4.7.3.4.1.2 VIDEO 2

Only in this video, the variable *articulation* was negatively correlated to *silent pauses* and positively to *sweet voice*. This video also presents the highest number of correlations between the variable *audible breaths* and other variables. In detail, *audible breaths* was negatively correlated to *melodious voice*, *expressive voice*, *sound-image* and *image-interpretation synchrony*, *credible* and *skilled interpreter*, *natural syntax* and *comprehensible interpretation*. The reason of so many correlations related to *audible breath*, with respect to other videos, may be found in the higher definition of the audio, due to the recording of the imitation in a studio, to create the experimental variable. It is worth to remember at this point that this experimental variable was devised to isolate the vocal traits of speech, at the expenses of a non-authentic simultaneous interpretation. The marked vocal traits and the non-authentic simultaneous interpretations seem to emerge in the perception of respondents, even if none of the subjects recognized the non-authentic interpretation in video 2. The variable *sweet voice* was positively correlated to *expressive voice*, and, only in this video, the variable *active personality* was positively correlated to *active personality*. Moreover, only in this video, the variable *same melody repeated* was positively correlated to *self-confident voice*, *complex words* and *complex sentences*. Considering these correlations, together with those of *articulation*, it may be assumed that the trained voice helped the perception of words and sentences. In addition, the variable *complex words*, only in this video, besides being positively correlated to *same melody repeated*, it was also positively correlated to *sound-image synchronism* and *comprehensible interpretation*. Only in this video the variable *silent pauses* was negatively correlated to *image-interpretation synchrony*, and not to that of *sound-image synchrony*. This may suggest that the silent pauses created in laboratory to build an artificial but plausible *décalage* was perceived as acceptable for a simultaneous interpretation, but not for a TV broadcast interpretation. This correlation may be due to the longest pause of the three video excerpts, which in video 2 occurred (was created) in the second part of the video and lasted 2.277 seconds (see figure 4.3). The variable *speed of speech* was negatively correlated to *silent pauses* and positively to *self-confident voice*, just as in video 3, which had a similar speech rate index (see table 4.4); it was negatively correlated to *sweet voice*, just as in video 1, which had a similar tonal range. Only in this video, *speed of speech* was positively correlated to *active personality*; and, only in this video, *speed of speech* was positively correlated to *credible*

*interpreter* and negatively to *skilled interpreter*; this is the only case where the variables *credible* and *skilled interpreter* show an opposite correlation to another variable. It may be supposed that the interpreter was considered not skilled for the silent pauses.

If we considered valid the correlations to the variables *effective comprehension 1, 2 and 3* (for the doubt on validity see above, section 4.7.3.3), then we should also mention that the video 2 shows almost all the correlations concerning these three variables. In fact, only in video 2, *effective comprehension 2* was positively correlated to *articulation* and *effective comprehension 3*, while *effective comprehension 3* was positively correlated to *articulation*, *hesitations*, *audible breaths*, *active personality*, *comprehensible voice* and *comprehensible interpretation*.

To conclude, the good articulation, the melodiousness and melodicity (see above – section 3.6.4) of voice and the speech rate facilitated the perception of *complex words* and *sentences*; however, comprehension was made difficult by *hesitations*, *silent pauses* and *audible breaths*, which also could have been perceived more clearly thanks to the speaker's telegenic voice and the better definition of the audio take, within the context of the experimental variable. Indeed, *hesitations*, *silent pauses* and *audible breaths* were correlated to a great number of other variables in this video.

#### 4.7.3.4.1.3 VIDEO 3

This video shows the lowest number of isolated correlations among variables, i.e. the lowest number of correlations concerning only video 3, or video 3 and either video 2 and video 1. Conversely, this mean that video 3 shows the highest number of correlations among variables that also concerned the other two videos. In fact, the correlation that concern only this video is the one, positive, between *expressive voice* and *complex sentences*. The other correlations concerning *complex sentences* were linked to *self-confident voice* and *active personality*, just as in video 1, while the only correlation concerning the variable *complex words* was that with *complex sentences*, but this one was also found in the other two videos. Moreover, there is no correlation between *natural syntax* and *complex words* and *sentences*, contrary to what happened in video 1 (both) or in video 2 (only *complex words* and *natural syntax*). These data may suggest that the interpretation in video 2 was not considered difficult to understand.

#### 4.7.3.4.1.4 VIDEOS 1-2, 1-3 AND 2-3

This section deals with the analysis of correlations of pair of variables concerning two of the three videos, in order to identify the similarities among videos, beyond those similarities shared by the three videos, and that were listed above (section 4.7.3.4.1), where they were defined as the “most interrelated variables”, precisely because their correlations were common to the three videos.

As for the couple of videos 1 and 2, both share a negative correlation between: *speed of speech* and *sweet voice*; *hesitations* and *image-interpretation synchrony*; *hesitations* and *active personality*; *hesitations* and *complex words*; *silent pauses* and *comprehensible voice*. Conversely, they share positive correlations between the following pairs of variables: *comprehensible voice* and *sound-image synchrony*; *sweet voice* and *expressive voice*; *articulation* and *image-interpretation synchrony*; *self-confident voice* and *image-interpretation synchrony*; *active personality* and *image-interpretation synchrony*; *comprehensible voice* and *image-interpretation synchrony*; *complex words* and *self-confident voice*; *complex words* and *credible interpreter*; *complex words* and *skilled interpreter*; *natural syntax* and *skilled interpreter*; *melodious voice* and *comprehensible interpretation*; *sweet voice* and *comprehensible interpretation*.

As to the pair of videos 1 and 3, both share negative correlations between the pair of variables *silent pauses* and *sound-image synchrony*, while they share negative correlations between the following pairs of variables: *complex words* and *self-confident voice*, *complex words* and *active personality*.

For what concerns the pair of videos 2 and 3, these share negative correlations in the following pairs of variables: *audible breaths* and *skilled interpreter*, *silent pauses* and *speed of speech*, *self-confident voice* and *speed of speech*, while they share a positive correlation only in the pair of variables *speed of speech* and *self-confident voice*.

#### 4.7.3.4.2 Perceptual maps of variables by multidimensional scaling

With the aim of going beyond the one-to-one correlation of variables, three perceptual maps, one for each videos, were plotted through the multidimensional scaling method (SPSS – MDS) (see below – figures 4.24 – 4.26). The original matrix of ordinal data (obtained from the ratings assigned to each variable by each respondent), was automatically converted into a distances matrix by the software; from this matrix of distances among variables, a bi-dimensional representation of distances was plotted (see Appendix 12). Following the correct methodology for this measurement,

three scree plots (one for each video) were produced, in order to discover the best number of dimensions suitable for the representation of distances (see below – figures 4.30-4.32). It was found that, for the three videos, the best number of dimensions was higher than 2. Therefore, a bi-dimensional representation of distances would have not been the optimal representation; however, a 4- or 3-dimensional plot hardly would have been readable. The observation of the Shepard diagrams, automatically plotted by the software together with the perceptual maps (see Appendix 12), confirmed the limits of a bi-dimensional representation. In conclusion, the bi-dimensional representation of distances is the result of an adjustment of vectorial distances data, hence the name “multidimensional scaling”; however, it represents a step forward with respect to bivariate correlations, since it allows to see the interrelations *among* more than two variables, and no longer only *between two* variables. The labels of the two dimensions of the graphs may be assigned (or not) by the researcher, on the basis of the research that is being conducted; in this case, a clue was given by the perceptual map in Chiaro and Nocella (2004). The indication of clusters of variables on the maps depends on the research, and is the result of an interpretation of data. Considering: i) the different nature of data related to the variables *effective comprehension 1, 2 and 3*, which are not ordinal (see above – section 4.7.3.3); ii) the results from the bivariate correlations concerning these variables (see above – section 4.7.3.4.1), which differed notably from those related to other variables, which were all ordinal; and iii) (at least for the perceptual maps) the nature of the multidimensional scaling, based on the calculation of the vectorial distances among elements, where the presence/absence of an element changes the distances among all elements; then the variables in question were excluded from the multidimensional scaling, the hierarchical cluster analysis and the analysis of principal components.





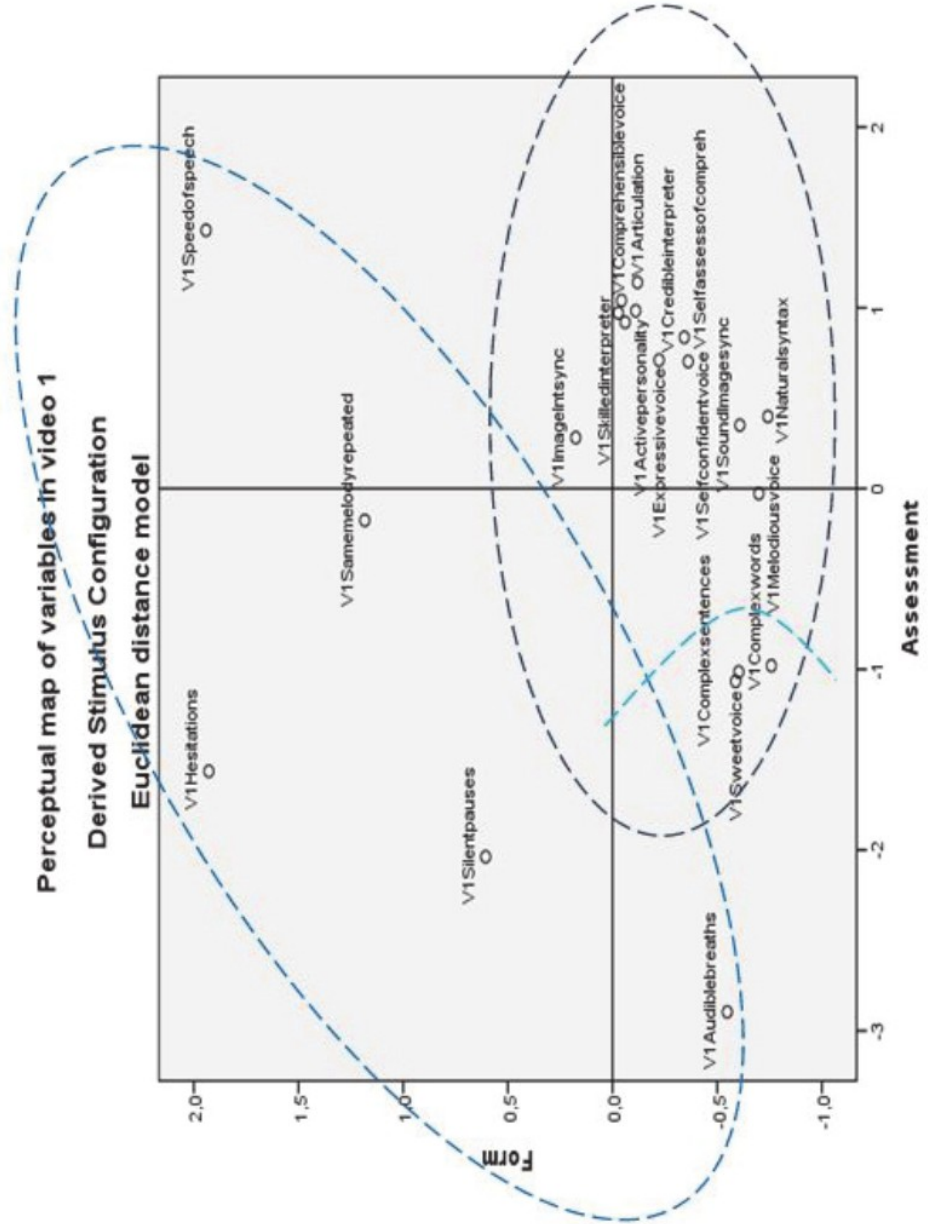


Figure 4.24. Multidimensional scaling of variables in video 1 (SPSS – Alscal).



## Perceptual map of variables in video 2

### Derived Stimulus Configuration

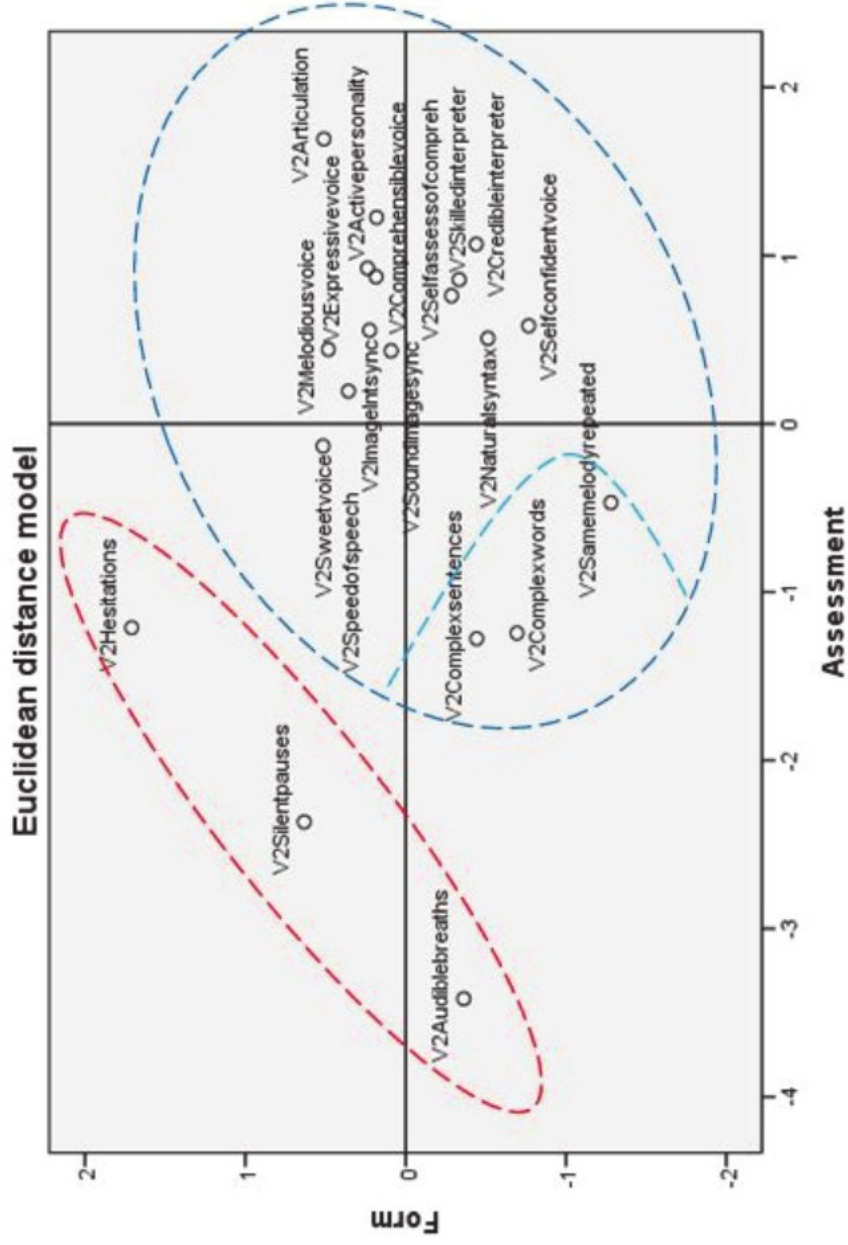


Figure 4.25. Multidimensional scaling of variables in video 2 (SPSS – Alscal).



Perceptual map of variables in video 3  
 Derived Stimulus Configuration

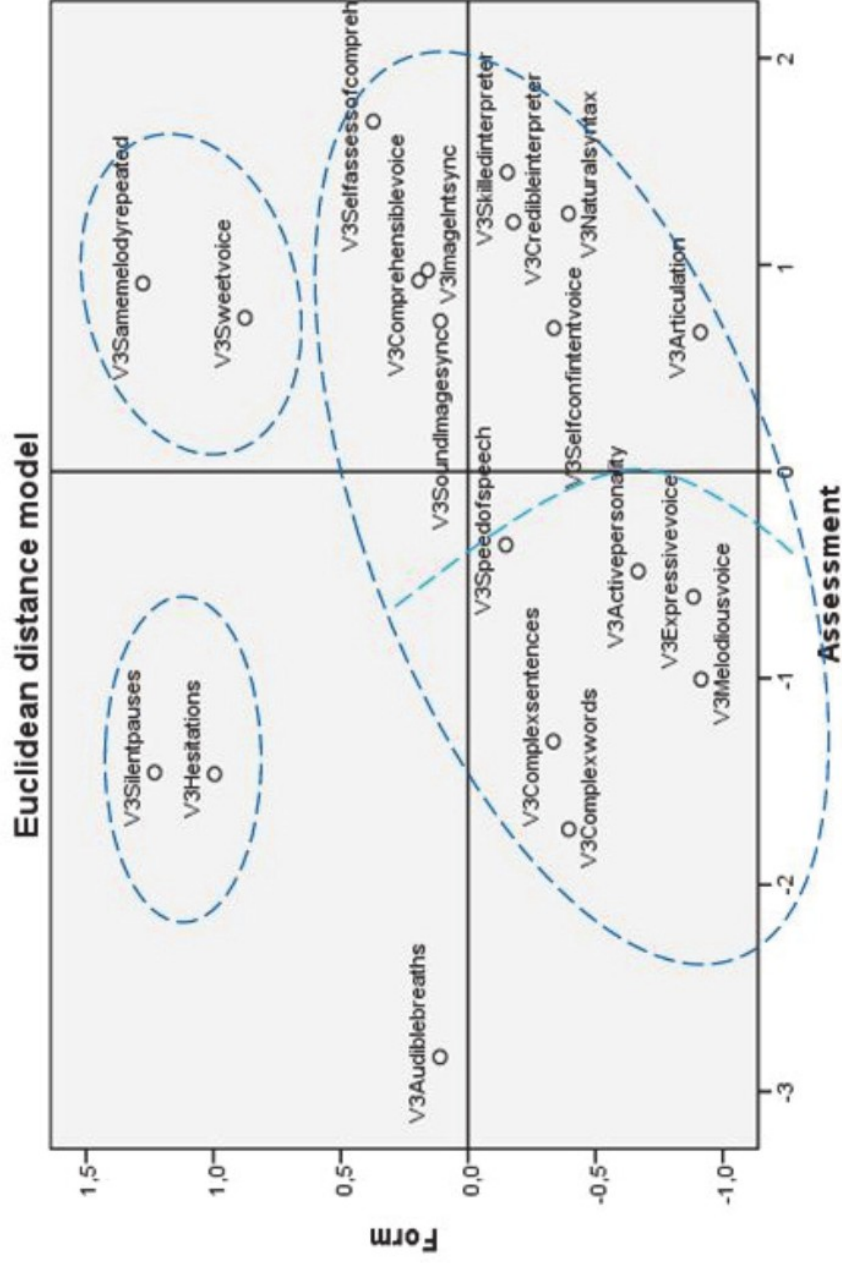


Figure 4.26. Multidimensional scaling of variables in video 3 (SPSS – Alscal).



Considering that the main aspect to take into account in the interpretation of these perceptual maps is the distance among the elements represented, in this case, the variables, the researcher has consequently interpreted the graphs by indicating the limits of the groups of variables (clusters), according to the distance/proximity among them.

In video 1, two main clusters were identified; one grouping the following variables, very close together: *image-interpretation synchrony*, *skilled interpreter*, *active personality*, *articulation*, *comprehensible voice*, *expressive voice*, *credible interpreter*, *comprehensible interpretation* (“self-assessment of comprehension”), *self-confident voice*, *sound-image synchrony*, *natural syntax*, *melodious voice*, *complex words*, *complex sentences* and *melodious voice*. A second cluster of variables, distant both from the elements of the first cluster and from one another, was identified, and it contains the following variables: *hesitations*, *speed of speech*, *same melody repeated*, *silent pauses* and *audible breaths*.

Also in video 2, two main clusters were identified, the first one is quite similar to the first cluster of video one, containing the majority of variables, positioned very close together, even more than in the similar group of video 1: *image-interpretation synchrony*, *skilled interpreter*, *active personality*, *articulation*, *sweet voice*, *comprehensible voice*, *expressive voice*, *credible interpreter*, *comprehensible interpretation* (“self-assessment of comprehension”), *self-confident voice*, *sound-image synchrony*, *natural syntax*, *melodious voice*, *complex words*, *complex sentences* and *melodious voice*, *same melody repeated*. The second cluster identified contains only three variables: *audible breaths*, *silent pauses* and *hesitations*.

In video 3, three clusters were identified; the first one is similar to the first one of the videos 1 and 2; however, the display of its elements shows a longer distance from one another; the variables grouped are the following: *comprehensible interpretation* (“self-assessment of comprehension”), *comprehensible voice*, *sound-image synchrony*, *image-interpretation synchrony*, *speed of speech*, *skilled interpreter*, *credible interpreter*, *self-confident voice*, *natural syntax*, *articulation*, *active personality*, *expressive voice*, *melodious voice*, *complex sentences*, *complex words*. The second cluster identified is constituted by two variables: *same melody repeated* and *sweet voice*. The third cluster identified contains two variables too: *silent pauses* and *hesitations*. The variable *audible breaths* remained isolated, out of any cluster.

The above described clusters were identified, for each plot, through a first glance at the display of the elements; therefore, the bigger distances among variables on the graph were considered. However, a more attentive

glance could lead to consider few elements of each of the bigger cluster in each video (referred to as “first” in the description above) as variables that could form either a sub-cluster or a different cluster; these variables are: *complex words and sentences* and *melodious voice* for video 1; *complex sentences and words* and *same melody repeated*, for video 2; *complex sentences and words*, *speed of speech*, *active personality*, *expressive voice* and *melodious voice* for video 3.

To conclude, by comparing and contrasting the three perceptual maps, it is evident that video 1 and 2 are similar; the main differences concern the distances among the variables, which is higher in video 3, and the position of some variables. In detail, in video 1, the *speed of speech* is close to *same melody repeated*; and *complex sentences and words* are close to *melodious voice*; in video 3, *same melody repeated* is close to *sweet voice* (and not to *speed of speech*), and *complex words and sentences* are close to *speed of speech*, *active personality*, *expressive voice* and *melodious voice*. In video 2, the one of the experimental variable, with a non-authentic simultaneous interpretation, the variables are more close together. Considering the comparison between videos 1 and 3, it can be hypothesized that *speed of speech* played an important role in the perception of the form of the interpretation speech.

#### 4.7.3.4.3 Hierarchical clusters of variables

In the statistical study of the interrelations among variables, a further step forward with respect to multidimensional scaling is represented by the analysis of clusters, because it is a more accurate approach to the definition of dissimilarities/distances among variables. In this case, three dendograms, one for each video, were plotted (through SPSS), in order to identify groups of variables that were assessed in a similar way. In this case, the identification of groups of variables is less subjective than that in the perceptual maps, because the clusters are created by the software. The definition of the number of clusters to be considered depends on the distance between a cluster and its subordinate/superordinate one, when this distance begins to increase with respect to the subordinate clusters, then the superordinate cluster should not be considered. In a dendogram plot, the distance among clusters is represented (rescaled to fall in the range of 0 to 25) on the top horizontal line of the plot, and the point of reference for the distance has to be the same for all the clusters in a plot; in our case we have three plots, then the same distance was considered for all of them (see below – figures 4.27 - 4.29).



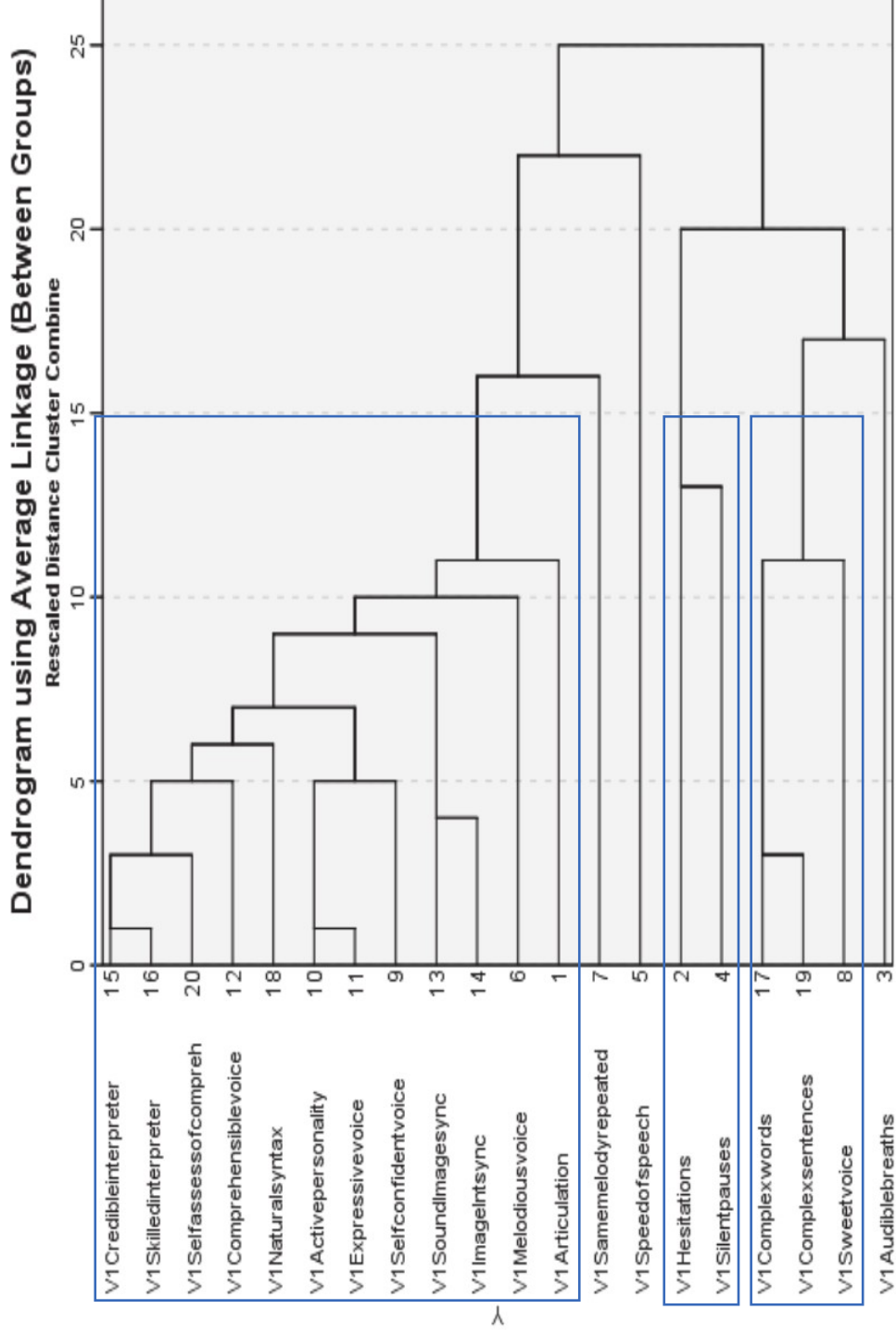
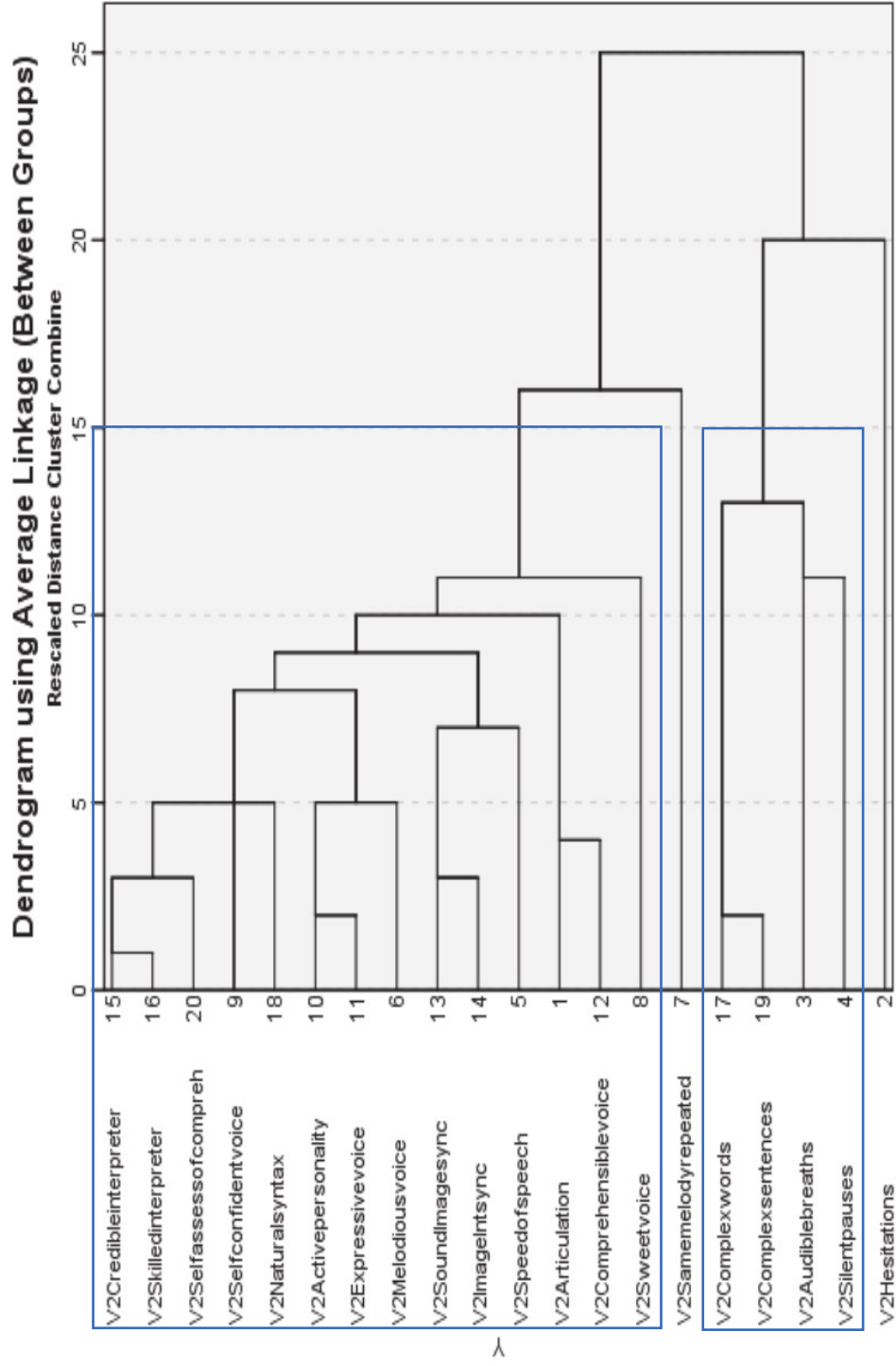


Figure 4.27. Dendrogram of video 1 (SPSS – hierarchical cluster analysis).





**Figure 4.28.** Dendrogram of video 2 (SPSS – hierarchical cluster analysis).



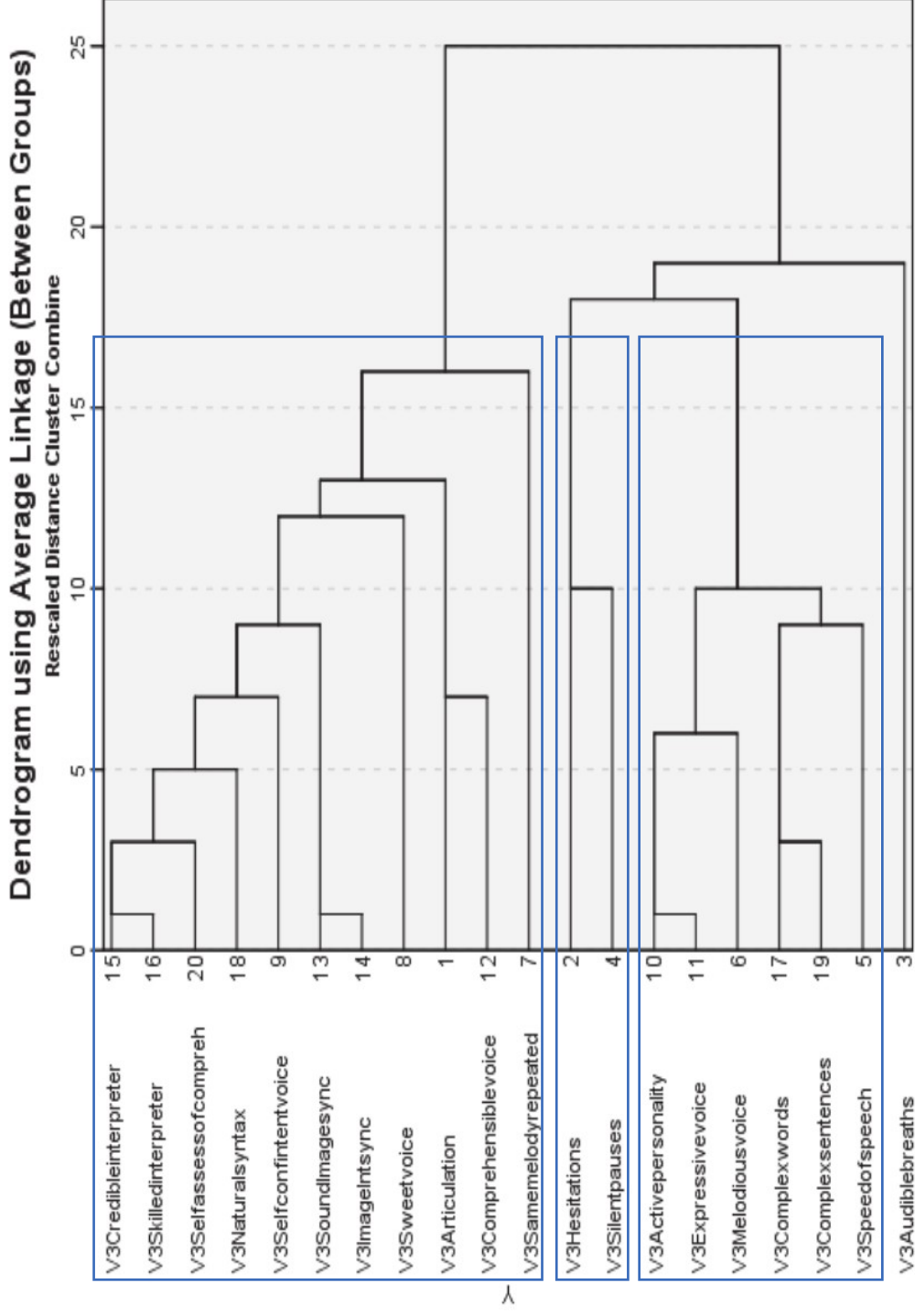


Figure 4.29. Dendrogram of video 3 (SPSS – hierarchical cluster analysis).



From a first glance, the clusters identified in the dendograms are similar to those identified in perceptual maps.

<b>Video 1</b>	Cluster 1	Credible interpreter, skilled interpreter, comprehensible interpretation (“self-assessment of comprehension”), comprehensible voice, natural syntax, active personality, expressive voice, self-confident voice, sound-image synchrony, image-interpretation synchrony, melodious voice, articulation, same melody repeated.
	Cluster 2	Hesitations, silent pauses.
	Cluster 3	Complex words, complex sentences, sweet voice.
	No cluster	Speed of speech, audible breaths.
<b>Video 2</b>	Cluster 1	Credible interpreter, skilled interpreter, comprehensible interpretation (“self-assessment of comprehension”), self-confident voice, natural syntax, active personality, expressive voice, melodious voice, sound-image synchrony, image-interpretation synchrony, speed of speech, articulation, comprehensible voice, sweet voice.
	Cluster 2	Same melody repeated complex words, complex sentences, audible breaths, silent pauses.
	No cluster	Hesitations.
<b>Video 3</b>	Cluster 1	Credible interpreter, skilled interpreter, comprehensible interpretation (“self-assessment of comprehension”), natural syntax, self-confident voice, sound-image synchrony, image-interpretation synchrony, sweet voice, articulation, comprehensible voice, same melody repeated.
	Cluster 2	Hesitations, silent pause.
	Cluster 3	Active personality, expressive voice, melodious voice, complex words, complex sentences, speed of speech.
	No cluster	Audible breaths.

**Table 4.7.** Composition of the clusters identified through the dendograms of the hierarchical cluster analysis.

The composition of the groups of variables identified through the dendograms of the hierarchical clusters shows the presence of the variables *complex words* and *sentences* in a single cluster in the three videos; in video 1 the two variables occur with *sweet voice*, in video 2 with *same melody repeated*, *audible breaths* and *silent pauses*, and in video 3 with *active personality*, *expressive* and *melodious voice*, and *speed of speech*.

The dendograms show – more clearly than the perceptual map – the pairs of variables that were assessed more or less in the same way in the three videos, since they were clustered together in a short distance; these pairs of variables are: *credible interpreter-skilled interpreter*, *active personality-expressive voice*; *sound-image synchrony-image-interpretation synchrony*; and *complex words-complex sentences*. This may suggest that in a future version of the questionnaire for the definitive survey, one element of each pair could be eliminated. Moreover, the relationship between voice and personality appears to be confirmed, as well as

the one between credible and skilled interpreter. The relationship between words and sentences, similar in perception, according to Bühler's proposal (1934/1983), also seems to be confirmed.

One way to use the dendograms to observe how one variable interrelates to other variables, in the three videos, is to consider at what level (distance) one variable is combined with other variables and see which other variables are clustered at the same level, and which other variables have been grouped in the subordinate clusters; then compare this information with that from other videos.

#### 4.7.3.4.4 Validation of questionnaire (analysis of principal component)

A statistical method used to reduce the questions of a questionnaire, i.e. to eliminate the variables/parameters that are being studied, is the analysis of principal component. It extracts the factors underlying a set of variables; the factors extracted can be used to interpret what the variables that were grouped in the same component have in common. The analysis of principal component was executed, through SPSS (Eigenvalue > 1 – varimax rotation), for each questionnaire (video) (see below – tables 4.8-4.10). In each table, corresponding to each video, the values that saturate each item/variable, i.e. that are more or less equivalent to the relative sum of the other values in other components, were highlighted.



Rotated Component Matrix <sup>a</sup>						
Video 1	Component					
	1	2	3	4	5	6
V1 Articulation	.746	.255	.057	-.176	.012	.022
V1 Hesitations	-.503	-.192	.527	.094	-.020	-.111
V1 Audible breaths	-.262	-.062	.525	.215	.428	-.276
V1 Silent pauses	-.145	.055	.856	-.070	-.142	.051
V1 Speed of speech	-.155	-.237	-.277	.717	-.259	-.059
V1 Melodious voice	.766	-.047	.167	-.187	.165	.154
V1 Same melody repeated	.139	.241	.349	.689	.166	.068
V1 Sweet voice	.344	-.153	-.047	-.230	.603	.190
V1 Self-confident voice	.810	.196	-.251	.111	-.039	.008
V1 Active personality	.832	.134	-.017	-.078	-.160	-.035
V1 Expressive voice	.861	.184	.071	-.077	.008	.080
V1 Comprehensible voice	.805	.073	-.189	-.119	.001	-.045
V1 Sound-image sync	.689	-.081	-.087	.274	.093	.365
V1 Image-Int sync	.487	.104	.038	.371	.090	.522
V1 Credible interpreter	.814	.306	-.261	.123	.055	.023
V1 Skilled interpreter	.838	.145	-.224	.165	.004	-.091
V1 Complex words	.158	.842	-.024	-.037	.044	.146
V1 Natural syntax	.759	.139	-.257	.123	.247	.091
V1 Complex sentences	.192	.783	.043	.093	.099	-.004
V1 Self-assess of compreh	.738	.023	-.297	.114	.312	-.139
V1 Effective compreh 1	-.013	-.128	.055	-.028	-.685	.031
V1 Effective compreh 2	-.069	.032	-.049	-.064	-.001	.870
V1 Effective compreh 3	.196	.363	-.098	-.088	-.231	-.195

Extraction Method: Principal Component Analysis.  
Rotation Method: Varimax with Kaiser Normalization.  
a. Rotation converged in 15 iterations.

**Table 4.8.** Table of coefficients extracted from video 1 (SPSS – principal component analysis).

Rotated Component Matrix <sup>a</sup>							
Video 2	Component						
	1	2	3	4	5	6	7
V2 Articulation	.521	.543	.062	.142	.178	-.072	-.039
V2 Hesitations	-.718	-.174	.306	.285	-.059	-.096	.034
V2 Audible breaths	-.490	-.543	-.271	.091	.031	.115	-.162
V2 Silent pauses	-.427	-.373	-.121	.605	.007	-.058	-.221
V2 Speed of speech	.089	-.145	.791	-.258	.123	-.029	-.095
V2 Melodious voice	.376	.171	.583	.537	.060	-.026	.015
V2 Same melody repeated	.119	-.120	-.104	.054	.098	.165	.844
V2 Sweet voice	.191	.229	-.139	.799	.016	.070	.108
V2 Self-confident voice	.816	.105	.195	-.028	.129	.223	.053
V2 Active personality	.526	.057	.537	.220	.196	.072	-.428
V2 Expressive voice	.482	.209	.361	.410	.304	.076	-.312
V2 Comprehensible voice	.556	.540	-.096	.132	-.024	.026	.198
V2 Sound-image sync	.319	.123	.081	.154	.746	.147	.013
V2 Image-Int sync	.298	.068	.102	-.069	.850	-.037	.037
V2 Credible interpreter	.827	.030	.249	.037	.262	.101	.012
V2 Skilled interpreter	.846	.091	.180	-.031	.170	.120	.124
V2 Complex words	.240	-.003	-.159	.046	.167	.827	-.061
V2 Natural syntax	.798	.016	.040	.221	.240	-.023	-.053
V2 Complex sentences	-.004	-.117	.124	-.003	-.081	.853	.253
V2 Self-assess of compreh	.733	.194	.004	.251	.098	.001	-.038
V2 Effective compreh 1	-.302	.407	.295	-.160	.330	-.079	.233
V2 Effective compreh 2	.005	.646	-.010	.103	.156	-.165	-.350
V2 Effective comreh 3	.124	.759	-.124	.064	.035	.050	-.097
Extraction Method: Principal Component Analysis.							
Rotation Method: Varimax with Kaiser Normalization.							
a. Rotation converged in 17 iterations.							

**Table 4.9.** Table of coefficients extracted from video 2 (SPSS – principal component analysis).

Rotated Component Matrix <sup>a</sup>								
Video 3	Component							
	1	2	3	4	5	6	7	8
V3 Articulation	.364	.291	.582	-.091	-.074	-.176	.031	-.357
V3 Hesitations	-.202	-.457	-.224	-.305	.328	-.194	.185	-.092
V3 Audible breaths	.095	-.722	-.284	.145	.110	-.015	-.149	.032
V3 Silent pauses	.002	-.786	.175	-.184	.002	-.099	.053	.081
V3 Speed of speech	.182	.506	-.486	.137	.094	-.152	-.225	.006
V3 Melodious voice	.664	-.080	-.018	.026	-.033	-.098	.428	-.177
V3 Same melody repeated	-.099	.019	.580	.252	-.097	-.184	-.073	.249
V3 Sweet voice	.007	.031	.144	.124	-.080	.034	.910	.052
V3 Self-confident voice	.394	.637	.250	.137	.254	.021	.069	.183
V3 Active personality	.802	.003	.094	.074	.129	.096	-.067	-.031
V3 Expressive voice	.889	.042	.019	.133	.139	-.025	-.048	-.077
V3 Comprehensible voice	.419	.014	.647	-.029	-.115	-.013	.315	-.163
V3 Sound-image sync	.170	.214	.113	.829	.139	-.010	.041	-.149
V3 Image-Int sync	.155	-.013	.084	.867	-.093	-.005	.102	.066
V3 Credible interpreter	.612	.222	.338	.432	.030	.081	-.042	.052
V3 Skilled interpreter	.572	.403	.476	.351	.077	.008	-.143	-.078
V3 Complex words	-.017	-.023	-.015	.048	.838	-.064	-.021	.055
V3 Natural syntax	.575	.409	.235	.249	-.189	-.110	.116	.033
V3 Complex sentences	.206	.012	-.075	-.021	.817	.063	-.092	-.106
V3 Self-assess of compreh	.283	.114	.689	.390	.074	.044	.121	-.001
V3 Effective compreh 1	.253	-.041	-.139	-.145	.106	.733	.075	.194
V3 Effective compreh 2	-.113	.012	-.007	-.052	-.047	.007	.021	.912
V3 Effective comreh 3	-.248	.151	.018	.148	-.135	.780	-.052	-.170

Extraction Method: Principal Component Analysis.  
 Rotation Method: Varimax with Kaiser Normalization.  
 a. Rotation converged in 12 iterations.

Table 4.10. Table of coefficients extracted from video 3 (SPSS – principal component analysis).

When there is no value being more or less equivalent to the relative sum of the others, then that item/question/variable could be omitted (from the questionnaire). This could be the case of: *audible breaths* and *image-interpretation synchrony*, in video 1; *audible breaths*, *melodious voice*, *active personality*, *expressive voice* and *comprehensible voice*, in video 2; *hesitations*, *credible* and *skilled interpreter* and *natural syntax*, in video 3. It is evident that each video shows a different number of components: if we do not consider the variables *effective comprehension 1*, *2* and *3*, for the reason explained above (section 6.3.1.2.3.1); then video 1 has 5 components, video 2 has 6, and video 3 has 7. It is also evident that the same variables may belong to different components in the three videos. For example, the variable *speed of speech*: in video 1, is in the component 4, together with *same melody repeated*; in video 2, it is in the component 3, alone; and in video 3, it is in the component 3, together with *silent pauses*, *audible breaths* and *self-confident voice*. It may be noticed that the variables *complex words* and *complex sentences*, in the three videos, are grouped together, without any other variable, but in different components, if considered each single video. If we consider all the videos (see below – table 4.11: correspondence between saturated variables, components and videos), there is no variable that has no saturation. Does this mean that all the items/variable should be confirmed in view of a new version of the questionnaire for a future definitive survey? If one looks at the variables saturated in the three videos, these are: *silent pauses*, *speed of speech*, *self-confident voice*, *complex words* and *complex sentences*, *same melody repeated*, *sound-image synchrony*.

Considering the results of other analysis, it may be supposed that *credible* and *skilled interpreter* almost coincide, that *audible breaths* are perceived as *silent pauses*, unless the volume or the audio definition are high.

Principal component analysis								
Correspondence item-component-saturation values for the three videos								
Videos 1 / 2 / 3	Component Video 3							
	Component Video 2							
	Component Video 1							
	1	2	3	4	5	6	7	8
Articulation	V1		V3					
Hesitations	V2		V1					
Audible breaths		V3						
Silent pauses		V3	V1	V2				
Speed of speech		V3	V2	V1				
Melodious voice	V1 / V2							
Same melody repeated			V3	V1			V2/V3	
Sweet voice				V2	V1			
Self-confident voice	V1/V2	V3						
Active personality	V1/V3							
Expressive voice	V1/V3							
Comprehensible voice	V1		V3					
Sound-image sync	V1			V3	V2			
Image-Int sync				V3	V2			
Credible interpreter	V1/V2							
Skilled interpreter	V1/V2							
Complex words		V1			V3	V2		
Natural syntax	V1/V2							
Complex sentences		V1			V3	V2		
Self-assess of compreh	V1/V2		V3					
Effective compreh 1					V1	V3		
Effective compreh 2		V2				V1		V3
Effective compreh 3		V2						

**Table 4.11** Comparison among variables present in each component, for the three videos – from the principal component analysis through SPSS (tables 4.8 – 4.10).

A similar table, also processed from the results of the principal component analysis, was created, to display the series of variables occurring in each component in each video (see below – table 4.12).



Variables in each component for the three videos								
Component								
	1	2	3	4	5	6	7	8
<b>Video 1</b>	Articulation Melodious voice Self-confident voice Active personality Expressive voice Comprehensible voice Sound-image sync Credible interpreter Skilled interpreter Natural syntax Comprehensible interpretation	Complex words Complex sentences	Hesitations Silent pauses	Speed of speech Same melody repeated	Sweet voice Effective comprehension 1	Effective comprehension 2		
<b>Video 2</b>	Hesitations Self-confident voice Credible interpreter Skilled interpreter Natural syntax Comprehensible interpretation	Effective comprehension 2 Effective comprehension 3	Speed of speech	Silent pauses Sweet voice	Sound-image sync Image-interpretation sync	Complex words Complex sentences	Same melody repeated	
<b>Video 3</b>	Active personality Expressive voice	Audible breaths Silent pauses Speed of speech Self-confident voice	Articulation Same melody repeated Comprehensible voice Comprehensible interpretation	Sound-image sync Image-interpretation sync	Complex words Complex sentences	Effective comprehension 1	Sweet voice	Effective comprehension 2

**Table 4.12.** Variables in each component (from the principal component analysis – SPSS – tables 4.8, 4.9, 4.10) in each video.





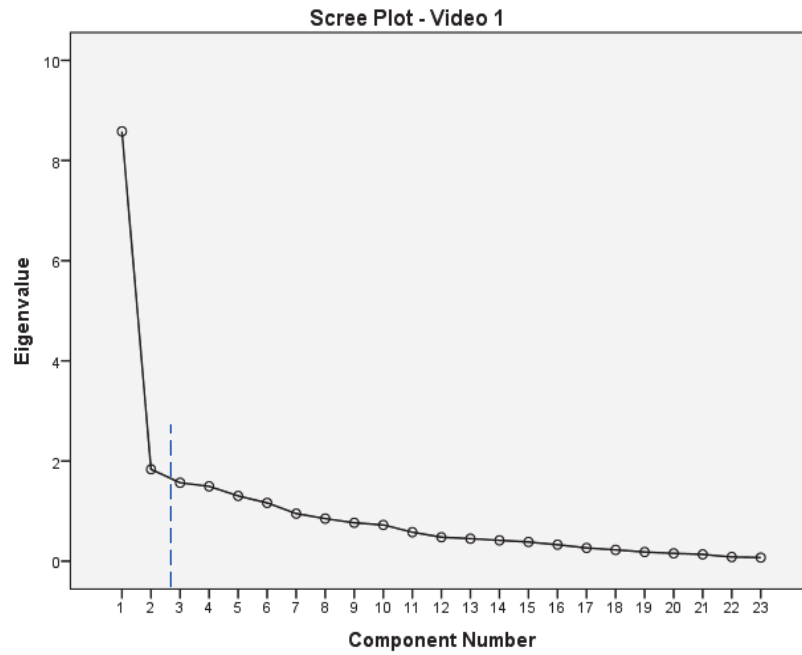
The analysis of principal component requires that a series of variables occurring in the same component should be synthesized in a new superordinate factor; however, in our case, given the gestaltic approach of the present study, it is not viable. Video 3 presents the most balanced distribution of variables in the 7 components; it may be hypothesized that the aspects of the interpreted speech were perceived in a more distinct way, especially considering the relatively low speed of speech. On the contrary, video 1 presents the most imbalanced distribution of variables among 5 components, the majority of them occurring in the component 1; this may be due to its relatively high *speed of speech* that could have induced the listeners to perceive a narrow form of speech. However, this form of speech hypothesized in the perception of video 1, may have rendered difficult the *comprehension* (in component 5), but led the respondents to consider the *interpreter* as *skilled* and *credible*. The analysis of principal component confirms the similarity of perception related to both *complex words* and *complex sentences*, since in the three videos the two variables occur together in the same component, even though these components have a different distance (component 2) from the component 1; in particular, they appear to be closer to other variables in video 1 while the same variables appear in component 6 in video 2 and in component 5 in video 3. The variables *sound-image synchrony* and *image interpretation synchrony* were included in component 1 in video 1, in component in video 2, and in component 4 in video 3; this results may induce to hypothesize that the *skilled* and *credible interpreter* of video 1 led the respondents to appreciate his ability not to lag behind (*décalage*), even if this had a negative impact on *effective comprehension* (considering also the *complex sentences* and *words*). On the contrary, in videos 2 and 3, *sound-image synchrony* and *image interpretation synchrony* occur in more distanced components, this could mean that the *sound-image synchrony* was appreciated for its function to facilitate the effective comprehension, even if a better comprehension might have been facilitated by the relatively low *speed of speech*. The variable *same melody repeated* occurs together with *speed of speech* in video 1, alone in video 2 and together with *articulation*, *comprehensible voice* and *comprehensible interpretation* in video 3. It may be hypothesized that *same melody repeated* had a negative perception in video 1, because the tonal range might have been reduced by high *speed of speech*; the same could be said for video 2, where the high tonal range of a non-authentic interpretation might have had an alienation effect on listeners; on the contrary, it might have had a positive perception in video 3, since it facilitated

*comprehension*. The alienating effect hypothesized for the experimental video deserves more attention; as a matter of fact, even if respondents did not recognize the non-authentic simultaneous interpretation, they knew what it was, because they had experience of simultaneous interpreting speech; therefore, their expectations might have been deceived by the dubbing actor's performance, and their attention might have been altered by the aesthetic effect of the highly marked (telegenic) narrating style of the speaker, as it was demonstrated in perception of music (cf. Imberty 1986:127-129):

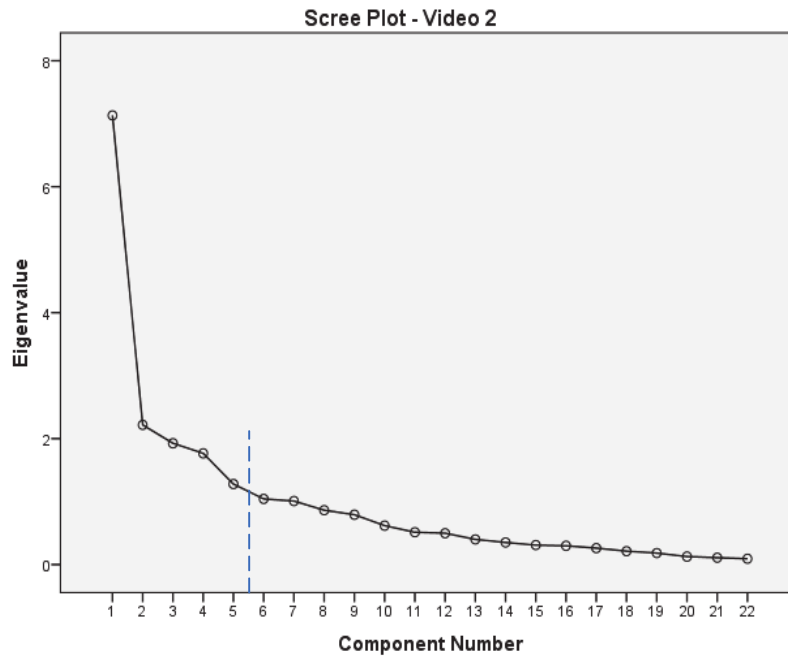
In breve, il potere che la musica ha di impressionare dipenderebbe dal suo livello di strutturazione e dalla conformità di tale strutturazione a taluni modelli immagazzinati nella memoria del soggetto. Per rifarci alla teoria dell'informazione, una debole *entropia* dei patterns melodici, ritmici o armonici e una forte *ridondanza* interna di questi stessi patterns non avrebbero sullo stato emotivo del soggetto le medesime conseguenze di una forte entropia e di una bassa ridondanza interna. Infatti Werbik (1969), in una ricerca originale, dimostra che il potere impressivo è una funzione di secondo grado dell'entropia (o grado di indeterminazione) dei patterns: quando si chiede a dei soggetti di valutare il livello impressivo di sequenze melodiche o ritmiche sulla base di scale bipolari di aggettivi antonimi, ne risulta che le sequenze dall'entropia molto elevata (indeterminazione massima) o dall'entropia molto debole (indeterminazione minima) hanno un potere impressivo più basso (valori su scala prossimi allo zero) rispetto alle sequenze a entropia media (valori su scala molto distanti dallo zero). Il fatto è che, in sequenze stocastiche appositamente costruite, un eccesso di ordine rende la sequenza troppo noiosa in seguito a concatenamenti troppo prevedibili; un eccesso di disordine rende impossibile cogliere la sequenza come unità a causa dell'alto grado di incertezza dei concatenamenti (Imberty 1986: 128-129).

This hypothesis should be tested with different typologies of respondents, for example, those identified for the definitive survey, i.e. common TV viewers, TV experts, actors and musicians.

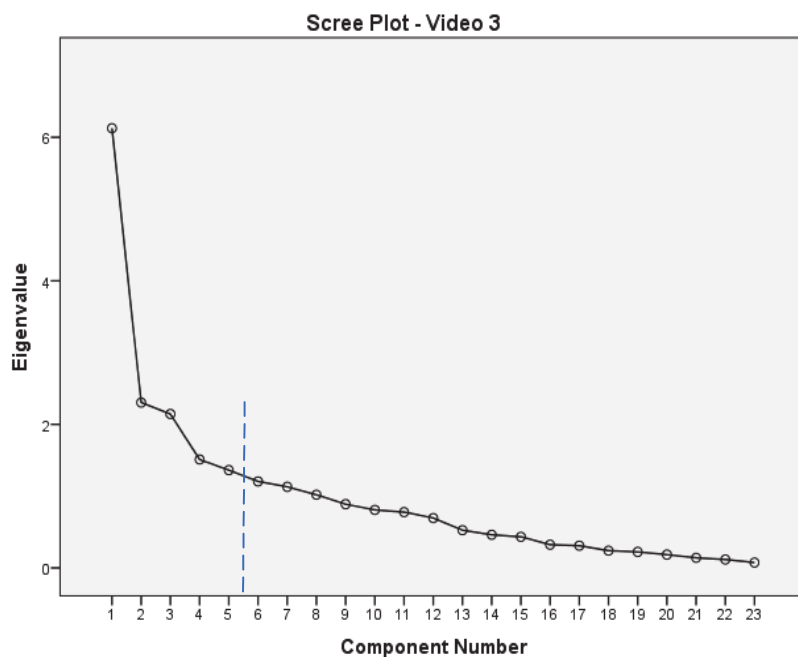
A more complete exploitation of the principal component analysis requires the observation of the scree plots, in order to decide how many components to consider for each video (see below – figures 4.30-4.32). According to the scree plot criterion, the maximum number of components that should be extracted is just prior to where the plot first begins to straighten out into a horizontal line; the line is nearly horizontal because the components all explain approximately the same amount of variance, which is not much.



**Figure 4.30.** Scree plot of video 1. The plot straightens out horizontally starting at component 3; therefore two components should be extracted for video 1.



**Figure 4.31.** Scree plot of video 2. The plot straightens out horizontally starting at component 6; therefore five components should be extracted for video 2.



**Figure 4.32.** Scree plot of video 3. The plot straightens out horizontally starting at component 6; therefore five components should be extracted for video 3.

According to the scree plots, the number of the main components to consider is: 2 for video 1, 5 for video 2, and 5 video 3; at this point, a table was created, through SPSS (varimax rotation – selected numbers of components: two for video 1, five for video 2 and 5 for video 3), with the variables arranged in the new components (see below – table 4.13; the original tables from the software output are reported in Appendix 13).

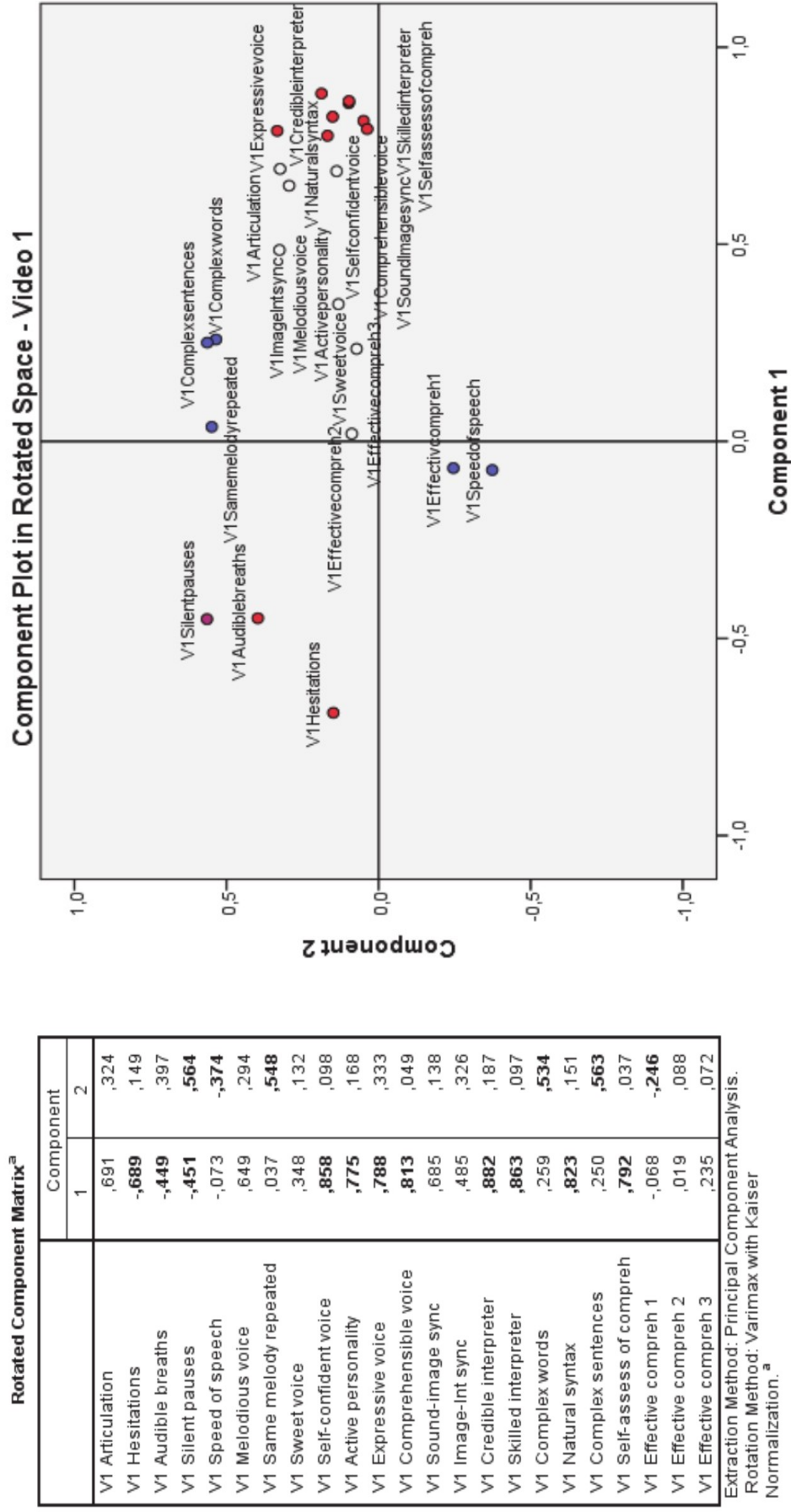
This analysis is more accurate than the first one, and it shows the similarities among videos 2 and 3, and the differences among these videos and video 1, since this video shows the variables distributed in a lower number (2) of components than those (5) in videos 2 and 3. The number of components for each video seems to suggest different forms of perception of the speeches related to the three videos; in particular, video 1 seems to have the narrowest form, while videos 2 and 3 seem to have the broadest form. The occurrences of the variables in the different components also seem to indicate the elements that make the different forms. In videos 2 and 3, the variables *skilled interpreter* and *self-confident voice* are not present in any component; conversely, in video 3, the variable *natural syntax* is not present, while in videos 1 and 2 is present in the same component (1). In video 2, the variables *hesitations*, *audible breaths* and *silent pauses* occur in component 2, isolated from any other variable; in video 3, the same components occur, without any other variable, in a

different component (3), while in video 1 they occur in component 1, together with other variables related to voice, the interpreter’s personality, sound-image synchronization, and natural syntax. The variables *complex words* and *complex sentences* occur together in the three videos, but in different components; in video 1, they occur with *speed of speech*, *same melody repeated* and *effective comprehension 1*, while in video 2 they occur with *same melody repeated*, and in video 3 with *sweet voice*. In video 3, the variable *speed of speech* is absent, while in video 2 it occurs with *sweet voice*, and in video 1 with *complex words* and *sentences* and *same melody repeated*. The occurrence of *same melody repeated*, *sound-image synchrony* and *image-interpretation synchrony* in the same component in video 3 could suggest the best realization – among the three videos – of the “synchresis” (i.e. the synchronized perception of image and sound – Chion 1990/1994: 63; see above – section 3.5.1). This could explain the occurrence, in the same component, in the same video, of *complex words* and *sentences* and *sweet voice*; it may also be supposed that these aspects be related to the absence, in this video, of *speed of speech*, where it actually showed the lowest rate in the three videos (see above – section 4.5.2.3.2).

Variables in each component (SPSS – selected components – varimax rotation)					
	Component				
	1	2	3	4	5
<b>Video 1</b>	Articulation Hesitations Audible breaths Melodious voice Self-confident voice Active personality Expressive voice Comprehensible voice Sound-image sync Image-interpretations sync Credible interpreter Skilled interpreter Natural syntax Comprehensible interpretation Effective comprehension 1	Speed of speech Same melody repeated Complex words Complex sentences Effective comprehension 1			
<b>Video 2</b>	Melodious voice Active personality Expressive voice Natural syntax	Hesitations Audible breaths Silent pauses	Same melody repeated Complex words Complex sentences	Speed of speech Sweet voice Effective comprehension3	Effective comprehension 1
<b>Video 3</b>	Articulation Melodious voice Active personality Expressive voice Comprehensible voice Effective comprehension 2	Same melody repeated Sound-image synchrony Image-interpretations synchrony	Hesitations Audible breaths Silent pauses	Sweet voice Complex words Complex sentences	Effective comprehension 1-3

**Table 4.13.** Synthesis from a principal component analysis of the three videos with selected number of components (SPSS – varimax rotation).

A component plot with only two components (and after a varimax orthogonal rotation) was created, through SPSS (see below – figures 4.33-4.35), in order to see how the variables were displayed along the two components, represented by the x and y axes; the orthogonal rotation means that the components (common factors) are not interrelated. The variables with the highest values in each component were highlighted, firstly on the table (bold) and secondly on the plot (red for component 1, blue for component 2 and violet for the components relevant to both components), and were considered relevant to the interpretation of the plots.



**Figure 4.33.** Principal component plot of video 1 (SPSS – two components – varimax rotation). The relevant variables, selected through the table of coefficients, are indicated in the plot (in red for the opposed variables along the component 1, blue for the variables of component 2 and violet for the variable relevant to the two components).





Rotated Component Matrix<sup>a</sup>

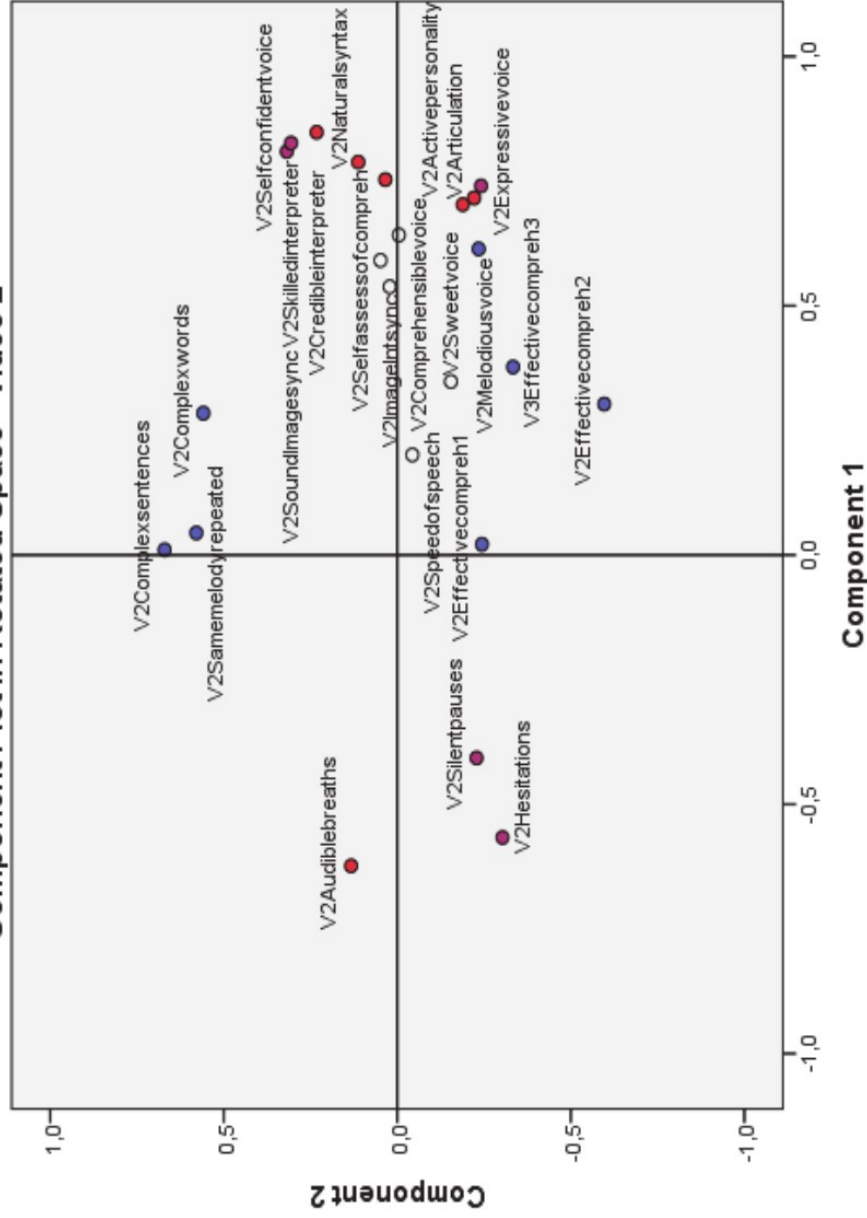
	Component	
	1	2
V2 Articulation	<b>,716</b>	-,221
V2 Hesitations	-,567	-,303
V2 Audible breaths	-,624	,133
V2 Silent pauses	-,407	-,229
V2 Speed of speech	,201	-,043
V2 Melodious voice	,614	-,234
V2 Same melody repeated	,044	<b>,579</b>
V2 Sweet voice	,348	-,155
V2 Self-confident voice	<b>,809</b>	<b>,318</b>
V2 Active personality	<b>,703</b>	-,189
V2 Expressive voice	<b>,740</b>	-,242
V2 Comprehensible voice	,642	-,004
V2 Sound-image sync	,591	,049
V2 Image-int sync	,538	,022
V2 Credible interpreter	<b>,848</b>	,232
V2 Skilled interpreter	<b>,826</b>	<b>,306</b>
V2 Complex words	,284	<b>,559</b>
V2 Natural syntax	<b>,788</b>	,112
V2 Complex sentences	,010	<b>,670</b>
V2 Self-assess of comreh	<b>,753</b>	,034
V2 Effective comreh 1	,021	-,244
V2 Effective comreh 2	,303	-,596
V2 Effective comreh 3	,377	-,333

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.<sup>a</sup>

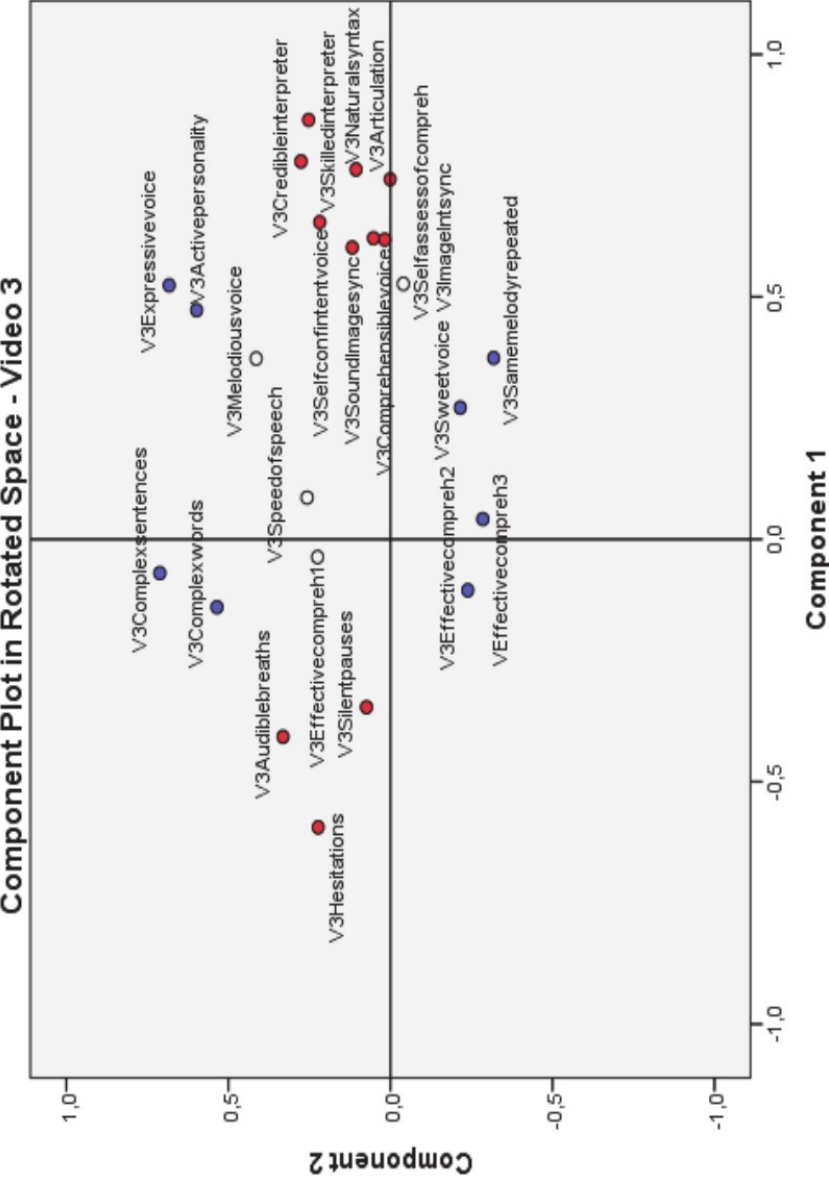
a. Rotation converged in 3 iterations.

Component Plot in Rotated Space - Video 2



**Figure 4.34.** Principal component plot of video 2 (SPSS – two components – varimax rotation). The relevant variables, selected through the table of coefficients, are indicated in the plot (in red for the opposed variables along the component 1, blue for the variables of component 2 and violet for the variable relevant to the two components).





**Rotated Component Matrix<sup>a</sup>**

	Component 1	Component 2
V3 Articulational	<b>,621</b>	,052
V3 Hesitations	<b>-,594</b>	,223
V3 Audible breaths	<b>-,407</b>	,331
V3 Silent pauses	<b>-,346</b>	,074
V3 Speed of speech	,086	,257
V3 Melodious voice	,373	,414
V3 Same melody repeated	,374	<b>-,319</b>
V3 Sweet voice	,271	<b>-,216</b>
V3 Self-confident voice	<b>,654</b>	,218
V3 Active personality	,473	<b>,598</b>
V3 Expressive voice	,524	<b>,683</b>
V3 Comprehensible voice	<b>,618</b>	,017
V3 Sound-image sync	<b>,602</b>	,118
V3 Image-int sync	,527	-,040
V3 Credible interpreter	<b>,779</b>	,276
V3 Skilled interpreter	<b>,865</b>	,253
V3 Complex words	-,140	<b>,535</b>
V3 Natural syntax	<b>,763</b>	,106
V3 Complex sentences	-,070	<b>,712</b>
V3 Self-assess of comprehend	<b>,743</b>	,001
V3 Effective comprehend 1	-,037	,225
V3 Effective comprehend 2	-,105	<b>-,239</b>
V3 Effective comprehend 3	,042	<b>-,285</b>

Extraction Method: Principal Component Analysis.  
 Rotation Method: Varimax with Kaiser Normalization.<sup>a</sup>

a. Rotation converged in 3 iterations.

**Figure 4.35.** Principal component plot of video 1 (SPSS – two components – varimax rotation). The relevant variables, selected through the table of coefficients, are indicated in the plot (in red for the opposed variables along the component 1, blue for the variables of component 2 and violet for the variable relevant to the two components).

As it can be seen from the plots, videos 1 and 3 show similar opposed variables along the two axes (components), while video 2 shows a slight difference. The peculiarity of video 2 is not only represented by the high number of variables present in its two components, with respect to other videos, but also by the way the variables of component 2 are displayed (opposed) along the correspondent axis. In detail, the group of variables made of *complex words*, *complex sentences*, *skilled interpreter*, *self-confident voice* and *same melody repeated* is opposed to the group of variables made of *hesitations*, *silent pauses*, *melodious voice*, *expressive voice* and *effective comprehension* 1, 2 and 3. In addition, in video 2, the variables *melodious voice* and *expressive voice* have a negative sign, while in videos 1 and 3 they have a positive sign and have not the same relevance they have in video 2. Looking at the differences among the three videos, it can be noticed that only in video 1, the variable *speed of speech* is relevant in the two components, while it is not relevant in any component in both video 2 and 3. The variable *sweet voice* is present only in the component 2 of video 3. The three video share the following opposed variables in the relative components:

Series of opposed relevant variables common to the three component plots		
	Negative	Positive
Component 1	Hesitations Audible breaths Silent pauses	Self-confident voice Skilled interpreter Natural syntax
Component 2	Same melody repeated Complex words Complex sentences Comprehensible interpretation	Effective comprehension

**Table 4.14.** Synthesis of the interpretation of the component plots – opposed variables common to the three videos.

The interpretation of the components plot would require to reach a synthesis of each component, by finding what its positive and negative variables have in common; however, in this case, it seems impossible, and considering the gestaltic approach of the study, it is not necessary. It may suffice to note that the above mentioned variables reflect the components of the sound aspect of words (Albano Leoni 2009; see above – section 3.4), namely: voice, syllable, prosody, sense, context, (linguistic) knowledge of the world. The variables also reflect, clearly, the main components of the model of speech perception proposed by Bühler (1934/1990: 199-200; see above –

section 3.1.7), i.e. the system voice-melody-words-sentences. In both models, voice is considered the mean of expression of the personality of the speaker and its interaction with the context. The relationship voice-personality (Fónagy 1983; see above – section 3.6) is also reflected in the variables *self-confident voice* and *skilled interpreter*. However, the transsensorial perception of sound-image, or “synchresis” (Chion 1990/1994: 63; see above – section 3.5.1), is not clearly reflected in these variables, given the absence of the variables *sound-image synchrony* or *image-interpretation synchrony*.

## 4.8 Conclusion - synthesis of the pilot survey

### 4.8.1 Form of speech and gestaltic perception

Considering the analysis of principal components, the variables validated by this measurement would be enough to support the theoretical approach, since in these variables can be found all the components of “*Il volto fonico delle parole*”, “*La vive voix*”, “the transsensorial perception of audio-vision”. Also Bühler’s threefold function of language appear confirmed, since the three speeches show three different forms of perception, given three different speakers, who talk about different topics, in different contexts, if considered the experimental variable introduced. From the results of the analysis of frequencies, correlations, perceptual maps and cluster analysis, it is evident that each (interpreted and televised) speech has different form in perception. There are some elements that are highly correlated, it is the case of *complex words*, *complex sentences*; or any variable referred to voice, her/his possessor and the voice-image relationship. The relationship among *complex words*, *complex sentences*, *sweet voice*, *sound-image synchrony*, *image-interpretation synchrony*, *same melody repeated* and *speed of speech*, concerning video 3 (see above – section 4.7.3.4.4), i.e. the one with the easiest topic, with the lowest numbers of words and syllables, and the lowest speed of speech; then it may be supposed that the synchresis, i.e. the gestaltic perception of sound-image, in a TV broadcast interpretation requires a slow speed of speech, with a good diction and voice control, as if the speech (sound) was viewed through the screen, at the same pace (slow) of the sequence of frame, and, above all, the same (high) definition of the screen, a sort of ‘pixel of the speech’.

As to the experimental variable (video 2), it was not recognized by respondent that it was not an authentic interpretation. The initial hypothesis of the dubbing actor’s voice as being more suitable to the comprehension of TV broadcast simultaneous interpretation was in part rejected by the analyses

of data, probably because, with reference to Bühler's threefold function of language, the component of expression was too marked (affectation) with respect to the representative component. This highly marked expression may have had an impact also in the perception of hesitations (mainly vocalizations and false starts), which also resulted marked; this aspect, together with the pauses artificially created, especially one long central pause, may have negatively influenced the general understanding of the interpretation. To this respect, it has to be remembered that, if expression is marked, not only hesitations are marked, but also the complex words and sentences are; and the marked expression may create problems in the receiver's interpretation of the message, in detriment of the representative function. In fact, in the gestaltic approach, not only hypoarticulation, but also hyperarticulation may create problems of perception (Albano Leoni 2009), having a negative impact on comprehension, as the results of comprehension of video 2 (experimental) compared to those of video 1 and 3 seem to confirm. In this case, the aesthetic effect of the marked expression, and, above all, of a non-authentic simultaneous interpretation, may have had a negative impact on the respondents, who were interpreting and translation students; maybe with different subjects the result would have been different.

Considering what has been said so far, it may suffice to note that sound material matters in speech perception; that this material is not only heard but also viewed. If it is true that words and sentences stay together, it is also true that these are made of sound too; therefore, findings support the proposal of 'the acoustic face of words', based on the gestaltic and physiognomic perception, described by Bühler (1931: 112-114) with the metaphor of the coin. A coin is recognizable at first sight thanks to the impress given to it by the coining stamp, and only in case of doubt, the coin is observed more carefully and, if necessary, refused. Similarly, like in trade, a communicative convention is "connected to the phonematic impress on the sound image of a word", a convention that fixes the symbolic value of the words. To this respect, it is worth to notice that, as it can be observed in the perceptual map of the experimental video (2), here the variables are more close together, distances are shorter than in the other two videos; moreover, the hierarchical cluster analysis show that the variable *same melody repeated* is superordinate to all the other variables, except *complex words and sentences*, *audible breaths*, *silent pauses* and *hesitations*, that form another cluster. From both the perceptual maps and the cluster analysis, it can be noticed that video 1 is more similar to video 3 in

the general form of speech; however, as it can be noticed from the perceptual maps and the analysis of principal components, videos 2 and 3 are similar in the higher distinction of the elements (variables) that constitute the form. The video 3 shows longer distances among variables; the dendogram shows that the variables *complex words* and *sentences* belong to the same cluster as *hesitations*, *silent pauses*, *active personality*, *expressive* and *melodious voice* and *speed of speech*. The interpretation of video 3 was better comprehend, and the speech presented the lowest speech rate, the lowest number of words, the most regular distribution of silent pauses, and the easiest topic.

To conclude, it could be relevant to mention a study (a MA thesis in interpretation) where a group of interpreting students were asked to interpret two different texts, very different in “structures and features”, for example, “new referents introduced, complex sentences and information structures, or the varied ways in which information progresses from one sentence to the next”, in order to analyse the effects in interpretation. Results showed, among other things, that: i) new referents introduced, complex sentences and information structures became “troublesome when some of them co-occur[red] in the same text segment”; ii) the “background knowledge” played a “paramount role” in “comprehension, anticipation and interpretation”; and iii) “students felt they had been confronted with a fast ST, but this feeling seem[ed] to be due to the number and complexity of the mental activities carried out during their task rather than to the speaker's actual speed of delivery” (Zani 2002: abstract).

#### **4.8.2 Suggestions for a further development of the questionnaire and the survey**

1. Considering the responses to the open-ended questions of the questionnaire, an item related to the assessment of cohesion should be introduced, as suggested by respondents, in an appropriate wording of the question, which should not include the linguistic term ‘cohesion’.

2. The duration of video excerpts could be extended, as suggested by some respondents, but it could not go beyond the 3.30 min; however, in this case, either the number of the questions could be reduced, or the video excerpts could be only two.

3. The questions could be reduced, at its minimum, to the items/variables that the analysis of principal component confirmed for the three videos, i.e. silent pauses, speed of speech, self-confident voice, complex

words and complex sentences, same melody repeated, sound-image synchrony. Nevertheless, the question on image-interpretation synchrony should be taken into account if the video excerpt should include a passage of an interpretation where there was a turn shift, a shot change, or something happening among the interlocutors that would require the intervention of the interpreter to explain what is being viewed, for example an overlapping intervention by an interlocutor off-screen triggering a reaction (even only in gestures) in the speaker onscreen.

4. If a similar experimental variable should be considered for a future definitive survey, to be administered to common TV viewers, TV experts, actors and musicians; then the acting dubber should be instructed to produce a less expressive speech, and more attention should be paid to the creation/deletion of silent pauses, especially by checking the rhythm of the original speech, to avoid altering it too much.



## CONCLUSIONS

### 1. Sound material and form

With reference to a question posed at the end of the introduction, it may appear that, if considered the last analysis of principal component (section 4.7.3.4.4.), the questionnaire can be reduced to a smaller number of items (see table 4.14). These are, in component 1: *hesitations, audible breaths* and *silent pauses* (positive) opposed to *self-confident voice, skilled interpreter* and *natural syntax* (negative); in component 2: *same melody repeated, complex words, complex sentences* and *comprehensible interpretation* (positive) opposed to *effective comprehension* (negative). The researcher finds it interesting to note that the aspect *same melody repeated* occurs in the same component as *complex words, complex sentences* and *comprehensible interpretation* (opposite to *effective comprehension*); similarly, in the same component as *self-confident voice* and *skilled interpreter*, there is the variable *natural syntax* (opposed to *hesitations, audible breaths* and *silent pauses*). Taking into account these results, the strict relationship between sound material and form proposed by Bühler (1934/1983) and reconsidered by Albano Leoni (2009) appears to be confirmed: the variable *natural syntax* was associated to voice, or rather to the vocal personality (see table 4.13), and the variable *same melody repeated* was associated to *words, sentences* and *comprehensible interpretation* (which could be referred to sense).

Again, with respect to the relationship between sound material and form, in the second analysis of principal component (table 4.13), video 1 shows a

lower number of components and a higher number of items in component 1. This datum appears to be related to the highest *speed of speech* in video 1, with respect to videos 2 and 3, which present a similar *speed of speech* (table 4.4) and the same number of components; the conclusion suggested by the researcher is that the more condensed speech in video 1 (see speed, highest number of words, even though not so different given the type/token ratio, highest articulation ratio, a more difficult topic with respect to other two videos) led the respondents to assess the variables in a similar way; the perceptual map of video 1 (figure 4.24) appears to confirm this hypothesis. Should this hypothesis be considered valid, then it would be another evidence of the strict relationship between sound material and form.

Other examples of this hypothesis could be done in this sense – the reader has all the data to find other examples to confirm or reject the researcher’s hypotheses.

## 2. Synchresis

As for synchresis, i.e. the gestaltic perception of the audio-video synchronization, none of the two items related to it (*sound-image* and *interpretation-image synchronization*) appears in the last analysis of principal component (table 4.14). As stated in section 4.7.3.1, the duration of video excerpts (1 min) did not contain any sequence of camera takes, but a single shot, a medium close-up, of the speaker; this could be the reason why the two variables in the end resulted irrelevant. Nevertheless, some signs of its impact in perception may be found in the second analysis of principal component (table 4.13), where it can be read that the two variables related to synchresis in video 1 occur in the first component, in video 2 do not occur in any component, while in video 3 occur in the second component. It may be hypothesized that, due to the good *décalage* of the interpreter in video 1, it resulted relevant according to the respondents’ assessment; in video 2 this did not happen, maybe due to the artificial *décalage* created for the experimental video, while in video 3 it resulted relevant, but not as much as in video 1. Indeed, the slow *speed of speech* of interpretation in video 3 (based on the low speed of speech of the speaker), the easy topic dealt with by the speaker, and consequently by the interpreter (cf. table 4.3), the lowest number of words, the high number of silent pauses (table 4.4), may be at the basis of a good *décalage* by the interpreter, which was not perceived (appreciated?) as much as in video 1.

### 3. Expressivity and comprehensibility – the experimental variable

The variable same melody repeated was created to elicit melodicity, i.e. a regular distribution of the fundamental frequencies among syllables, detected by Fónagy (1983: 310 – see section 3.6.5) in artistic voice or poet's readings, which was defined by the author as a third dimension of the melodic movement, being the other two time (duration) and height (tone). Melodicity represents an added value to the ordinary vocal expression, for this reason, it was detected by studying the melodic curves of artists' and poets' performances. Such melodic curves present a recurrent structure, a reiterated pattern, a frequent form. Considering the melodic curves of the interpretations in the video excerpts (figures 4.10-4.16), it could be assumed that they have a recurrent melodic structure (compare with figure 3.8), especially as a consequence of the expression of the cognitive activity of interpretation.

Now, looking at the perceptual maps (4.24 - 4.26) and at the second analysis of principal component (table 4.13), it can be noticed that, in the maps it is placed more or less in the same position in both videos 1 and 3, and in a completely different position in video 2; while in the components it occurs in the second component both in videos 1 and 2, and in the third component in video 2. In the components 1 and 2, it occurs with *complex words* and *complex sentences* (and *speed of speech*), while in video 3, it occurs with the variables *sound-image* and *image-interpretation synchronization*. It may be supposed that the variable same melody repeated: i) in video 1 depended also on the high *speed of speech*; ii) in video 2 only highlighted *complex words* and *complex sentences*; iii) and in video 3 was appreciated for the synchronization. This could be considered a case of synchresis, that could explain why the video 3 was the most comprehended one; the reasons may be detected by reading the linguistic units and indices of analysis (table 4.4): few words, slow speed of speech, easy topic, many short silent pauses. The video 2 (experimental), was less comprehended than video 1, despite the highly expressive voice of the dubbing actor. It may be assumed that too much expression may have worsened the perception of *complex words* and *sentences*, but also of filled pauses and of the long central pause created by the researcher to adjust the artificial *décalage*. Too much expressivity may reduce comprehensibility (Fónagy 1983: 310; see section 3.6.5). Another reason may lie on the fact that most of the subjects (interpretation students) knew what is an interpretation; therefore,

even if they did not recognize that in video 2 the interpretation was not authentic, they may have been disturbed by a melody they were not used to, different from that of a real interpretation (cf. Imberty 1986:127-129 – see section 4.7.3.4.4).

#### **4. Effects on TV interpreting profession and didactics**

Considering the hypotheses formulated in the previous section, it may be hypothesized that the perception of a TV broadcast simultaneous interpretation is favoured by a regular distribution of short silent pauses, no audible breath, a slow speed of speech, simple words and simple sentences, simple topics: all aspects that favour a good *décalage*, consequently, a good synchresis. In addition, the expression has to be moderate, avoiding affectation, in order not to impede the perception of words and sentences, and then the comprehension – not only the hypoarticulation, but also the hyperarticulation does not guarantee the comprehension (Albano Leoni 2009).

A television interpreter should be aware that her/his speech is not only listened to but also viewed on the screen; therefore, it cannot be too fast, it can be slow, so that it is observed in detail; it should have a natural melody, just like that used in a brilliant conversation. The TV interpreting speech should have the same definition of the screen, a sort of ‘pixel of the speech’.

## REFERENCES

- Ahrens, Barbara (2005). "Prosodic phenomena in simultaneous interpreting: A conceptual approach and its practical application". In *Interpreting* 7:1. Pag. 51-76.
- Albano Leoni, Federico (2009). *Dei suoni e dei sensi. Il volto fonico delle parole*. Bologna, Il Mulino.
- Albano Leoni, Federico; Clemente, Giuliana (2005). "Numeri, fonemi e foni". In *Parole e numeri. Analisi quantitative dei fatti di lingua*. Edited by Tullio De Mauro, Isabella Chiari, Roma, Aracne. Pag. 27-44.
- Albano Leoni, Federico; Cutugno, Francesco (1999). "The role of context in spontaneous speech recognition". In Proceedings of XIV ICPHS99, San Francisco. Pag. 861-864.
- Alfano, Iolanda; Marciano, Elio; Riccardi, Pasquale. (2007). *Sistemi per la valutazione della qualità della voce patologica*. On-line: <<http://www.parlaritaliano.it/index.php/it/dati/658-sistemi-per-la-valutazione-della-qualita-della-voce-patologica>> (last access: 01.2015).
- Babetto, Alice (2013). *Dibattiti presidenziali USA 2012: le strategie di cortesia nelle interpretazioni simultanee in italiano e in spagnolo*. Unpublished MA dissertation, SSLMIT, Università di Trieste.
- Bailey, Kenneth D. (1991). *Methods of Social Research*. New York, The Free Press [Italian translation by Maurizio Rossi (1995). *Metodi della ricerca sociale*. Bologna, Il Mulino].

- Baker, Mona (1993). “Corpus linguistics and translation studies: implications and applications.”. In *Text and Technology: In Honour of John Sinclair*. Edited by Mona Baker, Gill Francis, Elena Tognini-Bonelli, Amsterdam-Philadelphia, John Benjamins. Pag. 17-45.
- Baker, Mona (1995). “Corpora in Translation Studies: An overview and some suggestions for future research”. In *Target* (7), 2. Pag. 223-243.
- Baker, Mona (1996). “Corpus-based translation studies: The challenges that lie ahead”. In *Terminology, LSP and translation* Edited by Harold Somers, Amsterdam-Philadelphia, John Benjamins. Pag. 175-186.
- Baker, Mona (1998). “Réexplorer la langue de la traduction: une approche par corpus”. In *Meta* (43), 4. Pag. 480-485.
- Baker, Mona (1999). “The role of corpora in investigating the linguistic behaviour of professional translators”. In *International Journal of Corpus Linguistics* (4). Pag. 281-298.
- Bally, Charles (1950). *Linguistique générale et linguistique française*, Bern, Francke; [Italian translation by Giovanni Caravaggi (1963). *Linguistica generale e linguistica francese*, Milano, Il Saggiatore].
- Bally, Gérard, Aubergé, Veronique, Morlec, Yann (1997). “Des représentations cognitives aux représentations phonétiques de l'intonation”. In *Polyphonie pour Iván Fónagy. Mélanges offerts à Iván Fónagy par un group de disciples, collègues et admirateurs*. Edited by Jean Perrot, Paris, L'Harmattan. Pag. 19-28.
- Balzani, Maurizio (1988). *Il contatto visivo in interpretazione simultanea. Risultati di un esperimento*. Unpublished MA dissertation, SSLMIT, University of Trieste.
- Barbera, Manuel; Onesti, Cristina; Corino, Elisa (2007). “Cosa è un corpus?” Per una definizione più rigorosa di corpus, token, mark-up”. In *Corpora e linguistica in rete*. Edited by Manuel Barbera, Cristina Onesti, Elisa Corino,. Torino, Guerra Edizioni. Pag. 25-88.
- Barthes, Roland (1973). *Le plaisir du texte*, Paris, Seuil [Italian translation by Lidia Lonzi (1975). *Il piacere del testo*, Torino, Einaudi].
- Baudouin de Courtenay, Jan Ignacy Niecislaw (1895). *Versuch einer Theorie phonetischer Alternationen. Ein Kapitel aus der Psychophonetik*, Strassburg, Trubner.
- Bazzanella, Carla (1995). “I segnali discorsivi”. In *Grande grammatica italiana di consultazione. III Tipi di frasi, deissi, formazione delle parole*. Edited by Lorenzo Renzi; Gampaolo Salvi; Anna Cardinaletti. Bologna, Il Mulino. Pag. 225-257.

- Bendazzoli, Claudio (2010). *Testi e contesti dell'interpretazione di conferenza. Uno studio etnografico*. Bologna, Asterisco.
- Benveniste, Émile (1966). *Problèmes de linguistique générale*. Paris, Gallimard.
- Benveniste, Émile (1974). *Problèmes de linguistique générale II*, Paris, Gallimard.
- Blanche-Benveniste, Claire; Jeanjean, Colette (1987). *Le français parlé. Transcription et édition*. Paris, INALF, Didier Erudition.
- Bloch, Bernard (1948). "A set of postulates for phonemic analysis". In *Language*, 24. Pag. 3-46.
- Bloomfield, Leonard (1933). *Language*, London, Allen and Unwin.
- Bolinger, Dwight (1986). *Intonation and Its Parts. Melody in Spoken English*, Stanford, Stanford University Press.
- Bolinger, Dwight (1986). *Intonation and Its Uses. Melody in Grammar and Discourse*, London, Arnold.
- Bologna, Corrado (1992). *Flatus vocis. Metafisica e antropologia della voce*. Bologna, Il Mulino.
- Bonitzer, Pascal (1975). "Les silences de la voix". *Cahiers du cinema*, n. 256 (février-mars). Paris, Les Editions de l'Etoile. Pag. 22-33.
- Brancato, G., Macchia, S., Murgia, M., Signore, G., Simeoni, G., Blanke, G., Körner, T., Nimmergut, A., Lima, P., Paulino, R., Hoffmeyer-Zlotnik, J.H.P. (2006). *Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System*. European Commission Grant Agreement 200410300002. On-line: <[http://epp.eurostat.ec.europa.eu/portal/page/portal/research\\_methodology/documents/Handbook\\_questionnaire\\_development\\_2006.pdf](http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/documents/Handbook_questionnaire_development_2006.pdf)> (last access: 05.2014).
- Bregman, Albert .S. (1990). *Auditory Scene Analysis. The Perceptual Organization of Sound*, Cambridge-London, The Mit Press.
- Browman, Cathy. P.; Goldstein, Louis (1990) "Tiers in articulatory phonology, with some implication in casual speech". In *Papers in Laboratory Phonology, vol. I: Between the Grammar and Physics of Speech*. Edited by John Kingston, Mary .E. Beckman, Cambridge, Cambridge University Press. Pag. 341-376.
- Brown, Gillian (1974). "Practical Phonetics and Phonology". In *Techniques in Applied Linguistics*. Edited by John Patrick B. Allen and Stephen Pit Corder. Oxford, Oxford University Press. Pag. 24-58.
- Brown, Penelope; Levinson; Stephen C. (1987). *Politeness: some universals in language usage*. Cambridge, Cambridge University Press.

- Bühler, Hildegund (1986). "Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters". In *Multilingua* 5:4. Pag. 231-235.
- Bühler, Karl (1933/1978) *Ausdrucks theorie. Das System an der Geschichte aufgezeigt*, Stuttgart, Fischer [Italian translation by Lucio Pusci (1978). *Teoria dell'espressione . Il sistema alla luce della storia*, Roma, Armando].
- Bühler, Karl (1934/1983). *Sprachtheorie. Die Darstellungsfunktion der Sprache*, Jena, Fischer [Italian translation by Serena Cattaruzza Derossi (1983). *Teoria del linguaggio. La funzione rappresentativa del linguaggio*, Roma, Armando].
- Bühler, Karl (1934/1990). *Sprachtheorie. Die Darstellungsfunktion der Sprache*, Jena, Fischer [English translation by Donald Fraser Goodwin and Achim Eschbach (1990). *Theory of Language. The representational function of language*, Amsterdam-Philadelphia, John Benjamins].
- Calamai, Silvia (2002a). "Vocali d'Italia. Una prima rassegna". In *Voce, canto, parlato. Scritti in onore di Franco Ferrero*. Edited by Piero Cosi, Emanuela Magno Caldognetto, Alberto Zamboni, Padova, Unipress. Pag. 49-58.
- Calamai, Silvia (2002b). "Vocali atone e toniche a Pisa". In *Atti delle XIII Giornate di studio del Gruppo di Fonetica Sperimentale-ALA, (Macerata, 13-15 dicembre 2001)*. Edited by Agostino Regnicoli, Roma, Il Calamo. Pag. 39-46.
- Cartellieri, Claus (1983). "The inescapable dilemma: Quality and/or quantity". In *Babel* 29:4. Pag. 209-213.
- Casetti, Francesco (2002). *Communicative Negotiation in Cinema and Television*. Milano, Vita e Pensiero.
- Casetti, Francesco; Di Chio, Federico (1994). *Analisi del film*. Milano, Bompiani.
- Catana, Cinzia (2005). *Le qualità ben pronunciate: la dizione dell'interprete e la percezione dell'utente in simultanea*. Unpublished MA dissertation, SSLMIT, Università di Bologna.
- Cerrato, Loredana; Falcone, Mauro; Albano Leoni, Federico (1998) "Is it possible to evaluate the contribution of visual information to the process of speech comprehension?". In *Proceedings of the AVSP 98, International Conference on Auditory-Visual Speech Processing*. Edited by Denis Burnham, Jordi Robert-Ribes, Eric Vatikiotis-Bateson, Terrigal-Sydney. Pag. 141-145.
- Chiaro, Delia; Nocella, Giuseppe (2004). "Interpreters' Perception of Linguistic and Non- Linguistic Factors Affecting Quality: A Survey through the World Wide Web". In *Meta* 49:2. Pag. 278-293.



- Chion, Michel (1990). *L'Audio-Vision*. Paris, Editions Nathan [English edition and translation by Claudia Gorbman (1994). *Audio-vision: sound on screen*. New York, Columbia University Press].
- Chomsky, Noam; Halle, Morris (1968). *The Sound Pattern of English*, New York-London, Harper & Row.
- Christodoulides, George (2013). "Prosodic features of simultaneous interpreting." In *Proceedings of the Prosody-Discourse Interface Conference 2013 (IDP-2013)*, Leuven, 11-13 September. Pag. 33-37.
- Christodoulides, George; Lenglet, Cédric (2014). "Prosodic correlates of perceived quality and fluency in simultaneous interpreting". In *Proceedings of the Speech Prosody 7 Conference*, Dublin, 20-23 May. Pag. 1002-1006.
- Coblentz, Horst; Muhar, Franz (1999). *Atem und Stimme*. Wien, öbv et htp [Italian translation by Sabine Bruggisser and Gabriella Sacchetti (2004). *Respiro e voce. Istruzioni per parlare bene*, Milano, Ubulibri].
- Collados Aís, Ángela (1998). *La evaluación de la calidad en interpretación simultánea. La importancia de la comunicación no verbal*. Granada, Comares.
- Collados Aís, Ángela; García Becerra, Olalla (2015). "Quality". In *The Routledge Handbook of Interpreting*. Edited by Holly Mikkelsen, and Renée Jourdenais, London and New York, Routledge. Pag. 368-383.
- Collados Aís, Ángela; Iglesias Fernández, Emilia; Pradas Macías, Esperanza Macarena; Stévaux, Elisabeth, eds. (2011). *Qualitätsparameter beim Simultandolmetschen: Interdisziplinäre Perspektiven*. Tübingen: Verlag.
- Collados Aís, Ángela; Pradas Macías, Esperanza Macarena; Stévaux, Elisabeth; García Becerra, Olalla, eds. (2007). *La evaluación de la calidad en interpretación simultánea: parámetros de incidencia*. Granada, Comares.
- Combarieu, Jules (1982). *La musica e la magia*. Milano, Arnoldo Mondadori Editore.
- Crystal, David (1969). *Prosodic Systems and Intonation in English*, Cambridge, Cambridge University Press.
- Cutler, Anne (1987). "Speaking for listening". In *Language Perception and Production*. Edited by Alan Allport, Donald G. Mackay, Wolfgang Prinz, and Eckart Scheerer, London, Academic Press. Pag. 23-40.
- Dal Fovo, Eugenia (2011). "Through the CorIT looking-glass and what MA students found there". In *The Interpreters' Newsletter*. 16. Pag. 1-20.
- Dal Fovo, Eugenia (2014). *Simultaneous interpretation on television: the case of the question/answer group in the 2004 U.S. Presidential Debathon*. Ph.D. Dissertation, Università di Trieste.

- De Gregoris, Gregorio (2014) .“The limits of expectations vs. assessment questionnaire-based surveys on simultaneous interpreting quality: the need for a holistic model of perception”. In *RITT (Rivista internazionale di tecnica della traduzione)*, 16:2014. Pag. 57-87.
- De Gregoris, Gregorio (forthcoming). “A proposal for an assessment of voice quality in TV broadcast simultaneous interpreting through a gestaltic approach: theoretical paradigm for a new questionnaire”. In *MonTI*, 3:2016.
- De Mauro, Tullio (2007). "Saussure in cammino". In *La lezione di Saussure. Saggi di epistemologia linguistica*. Edited by Annibale Elia, Marina De Palo, Roma, Carocci. Pag. 19-32.
- De Meo, Anna (2012). “How credible is a non-native speaker? Prosody and surroundings”. In *Methodological Perspectives on Second Language Prosody. Papers from ML2P 2012*. Edited by Maria Grazia Busà, Antonio Stella. Padova, CLEUP. Pag. 3-9. On-line: <<http://www.maldura.unipd.it/LCL/ML2P/proceedings.html>> (last access: 02.2016).
- De Meo, Anna; Pettorino, Massimo; Vitale, Marilisa (2012) “Non ti credo: i correlati acustici della credibilità in italiano L2”. In *Proceedings of the 11th International Congress of AItLA. Competenze e formazione linguistiche. In memoria di Monica Berretta*. Edited by Giuliano Bernini, Cristina Lavinio, Ada Valentini, Miriam Voghera. Perugia, Guerra Edizioni. Pag. 229-248.
- Desbordes, Françoise (1988). "La prétendue confusion de l' écrit et de l'oral dans les théories de l'antiquité". In *Pour une théorie de la langue écrite. Actes de la Table Ronde internationale CNRS-HESO (Paris, 23-24 octobre 1986)*. Edited by Nina Catach, Paris, Editions du Centre National de la Recherche Scientifique. Pag. 27-33.
- Desbordes, Françoise (1990). *Idées romaines sur l'écriture*, Lille, Presses Universitaires de Lille.
- Doerflinger, Óscar (2003). "Gestión de la calidad total". In *La evaluación de la calidad en interpretación: investigación*. Edited by Ángela Collados Aís, María Manuela Fernández Sánchez, Daniel Gile, Granada, Comares. Pag. 171-176.
- Donegan, Patricia J.; Stampe, David (1979) "The study of natural phonology". In *Current Approaches to Phonological Theory*. Edited by Daniel .A. Dinnsen, Bloomington, Indiana University Press. Pag. 126-173.
- Ducard, Dominique (2002). *La voix et le miroir. Une étude sémiologique de l'imaginaire et de la formation de la parole*, Paris, L'Harmattan.

- Durand, Jacques; Laks, Bernard (2002). "Phonology, phonetics and cognition". In *Phonetics, Phonology and Cognition*. Edited by Jacques Durand, Bernard Laks, Oxford, Oxford University Press.
- Falbo, Caterina (2001). "Un corpus orale per l'interpretazione". In *Oralità nella parola e nella scrittura / Oralité dans la parole et dans l'écriture*. Edited by Mariagrazia Margarito, Enrica Galazzi, Monique Lebhar Politi. Torino, Edizioni Libreria Cortina. Pag. 319-335.
- Falbo, Caterina (2005). "La transcription: une tâche paradoxale". In *The Interpreters' Newsletter* 13. Pag. 25-38.
- Falbo, Caterina (2012). "CorIT – Television Interpreting Corpus". In *Breaking ground in corpus-based interpreting studies*. Edited by Francesco Straniero Sergio, Caterina Falbo. Bern, Peter Lang. Pag. 155-186.
- Falbo, Caterina (2012). "CorIT (Italian Television Interpreting Corpus): classification criteria". In *Breaking Ground in Corpus-based Interpreting Studies*. Edited by Francesco Straniero Sergio, Caterina Falbo. Bern, Peter Lang. Pag. 157-185.
- Florin Sider (1993). "Realia in translation". In *Translation as a social action. Russian and Bulgarian perspectives*. Edited by Palma Zlateva, London, Routledge. Pag. 122-128.
- Fónagy, Ivan (1983). *La vive voix. Essais de psycho-phonétique*. Paris, Payot.
- Fónagy, Ivan (1993). "Il significato dello stile vocale". In *Phoné Semantiké, Il Verri*, IX, 1-2 (maggio-giugno). Pag. 7-29.
- Fortuna, Sara (2005). *Il laboratorio del simbolico. Fisiognomica, percezione, linguaggio da Kant a Steinthal*, Perugia, Guerra.
- Fox, Barbara A.; Maschler, Yael; Uhmman, Susanne (2009). "Morpho-syntactic resources for the organization of the same-turn self-repair: Cross-linguistic variation in English, German and Hebrew". In *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion*, 10. Pag. 245-291. On-line: <<http://www.gespraechsforschung-ozs.de/heft2009/ga-fox.pdf>> (last access: 04.2013).
- García Becerra, Olalla (2013). *La incidencia de las primeras impresiones en la evaluación de la calidad de la interpretación simultánea: un estudio empírico*. Ph.D. dissertation, Universidad de Granada.
- Garroni, Emilio; Pautasso, Sergio. (a cura di) (1979) *Tesi pubblicate sul primo numero dei «Travaux du Cercle Linguistique de Prague» del 1929*, Napoli, Guida.
- Garzone, Giuliana (2003). "Reliability of quality criteria evaluation in survey research". In *La evaluación de la calidad en interpretación: investigación*. Edited

- by Ángela Collados Aís, María Manuela Fernández Sánchez, Daniel Gile. Granada, Comares. Pag. 23-30.
- Genette, Gérard (1972) *Figures III*. Paris, Editions du Seuil [Italian translation by Lina Zicchi (1976). *Figure III*. Torino, Einaudi].
- Gile, Daniel (1990). “L'évaluation de la qualité de l'interprétation par les délégués: une étude de cas”. In *The Interpreter's Newsletter* 3. Pag. 66-71.
- Gile, Daniel (1995). *Basic Concepts and Models for Conference Interpretation Training*. Amsterdam/Philadelphia, John Benjamins.
- Gile, Daniel. (1988) “Le partage de l'attention et le ‘modèle d'effort’ en interprétation simultanée”. In *Meta* 28, 3. Pag. 236-243.
- Gile, Daniel. (1991) “A communication-oriented analysis to quality” In *Translation: Theory and Practice. ATA Scholarly Monographs Series*. Edited by M.L. Larson, Binghamton, NY, Suny. Pag. 188-200.
- Gili Fivela, Barbara (2011). "Prosodia". In *Enciclopedia dell'italiano*. Roma, Istituto della Enciclopedia Italiana G. Treccani. On-line: <[http://www.treccani.it/enciclopedia/prosodia\\_\(Enciclopedia\\_dell'Italiano\)/>](http://www.treccani.it/enciclopedia/prosodia_(Enciclopedia_dell'Italiano)/>) (last access 05.2014).
- Gillham, Bill (2000). *Developing a Questionnaire*. London, Continuum.
- Goffman, Erving (1967). *On face-work: an analysis of ritual elements in social interaction*. Harmondsworth, Penguin Books.
- Goldman-Eisler, Frieda (1968). *Psycholinguistics. Experiments in spontaneous speech*. London/New York, Academic Press.
- Goldsmith, John A. (1990). *Autosegmental and Metrical Phonology*, Oxford-Cambridge, Blackwell.
- González Conde, María Julia (2001). *Comunicación radiofónica*. Madrid, Universitas.
- Greimas, Algirdas Julien (1966) *Sémantique Structurale: recherche de methode*. Paris, Larousse [Italian translation by I. Sordi, (1969) *La semantica strutturale: ricerca di metodo*, Milano, Rizzoli].
- Guardini, Paola (1995). *La traduzione simultanea del film: produzione e percezione. Un'analisi della prestazione dell'interprete e della risposta del pubblico*, Unpublished MA dissertation, SSLMIT, Università degli Studi di Trieste.
- Hair, Joseph F.; Anderson, Rolf E.; Thatham, Ronald L.; Black William C. (1995). *Multivariate Data Analysis with Readings*. New Jersey, Prentice Hall.

- Halliday, M.A.K. (1967). *Intonation and Grammar in British English*, Den Haag-Paris, Mouton.
- Hawkins, Sara (1999) "Auditory capacities and phonological development: Animal, baby and foreign listeners". In *The Acoustics of Speech Communication*. Edited by James M. Pickett, Needham Heights, MA, Allyn and Bacon. Pag. 183-198.
- Hawkins, Sara; Smith, Rachel (2001): "Polysp: A polysystemic, phonetically-rich. approach to speech understanding". In *Italian Journal of Linguistics/Rivista di linguistica*, 13, 1. Pag. 99-188.
- Hockett, Charles Francis (1955). *A Manual of Phonology*, Baltimore, Waverly Press.
- Hockett, Charles Francis (1963). *A Course in Modern Linguistics*, New York, Macmillan.
- Holub, Elisabeth (2010). "Does intonation matter? The impact of monotony on listener comprehension". In *The Interpreters' Newsletter* 15. Pag. 117-126.
- Hurch, Bernhard; Rhodes, Richard A. (eds.) (1996). *Natural Phonology: The State of the Art*, Berlin-New York, Mouton/de Gruyter.
- Iglesias Fernández, Emilia (2011). "Stimme". In *Qualitätsparameter beim Simultandolmetschen: Interdisziplinäre Perspektiven*. Edited by Ángela Collados Aís, Emilia Fernández Iglesias, Esperanza Macarena Pradas Macías, Elisabeth Stévaux. Tübingen, Verlag. Pag. 33-60.
- Iglesias Fernández, Emilia (2013). "Unpacking Delivery Criteria in Interpreting Quality Assessment". In *Assessment Issues in Language Translation and Interpreting*. Edited by Roelof van Deemter and Dina Tsagari. Frankfurt, Peter Lang. Pag. 51-66.
- Imberty, Michel (1970). "Polysémie et cohérence du langage musical: I La polysémie dans la réponse verbales associées à la musique et la construction d'une échelle circulaire des expressivités musicales". In *Sciences de l'Art-Scientific Aesthetics* VII, 1-2. Pag. 77-93.
- Imberty, Michel (1971). "Polysémie et cohérence du langage musical: II La structure des connotations verbales de la musique et la cohérence des contenus sémantiques. Vers une étude différentielle de la perception des styles". In *Sciences de l'Art-Scientific Aesthetics* VIII, 2. Pag. 65-81.
- Imberty, Michel (1978). *Sémantique psychologique de la musique*. Thèse de doctorat en Lettres et Sciences Humaines, Université de Paris X-Nanterre.
- Imberty, Michel (1979). *Entendre la musique*. Paris, Dunod.

- Imberty, Michel (1986). *Suoni, emozioni, significati. Per una semantica psicologica della musica*. [Italian edition and translation by Johannella Tafuri and Laura Callegari. Bologna, CLUEB].
- Jakobson, Roman (1960). "Closing statements: Linguistics and poetics". In *Style in Language*. Edited by Thomas A. Sebeok, Cambridge, Mass., The MIT Press. Pag. 350-377.
- Jakobson, Roman; Halle, Morris (1956). "Phonology and phonetics". In *Fundamentals of Language*. Edited by Roman Jakobson, Morris Halle, Den Haag, Mouton. Pag. 1-51.
- Kerbrat-Orecchioni, Catherine (1992). *Les interactions verbales – Variations culturelles et échanges rituels*, Paris, Armand Colin.
- Kerbrat-Orecchioni, Catherine (2005). *Les discours en interaction*, Paris, Armand Colin.
- Kerbrat-Orecchioni, Catherine (2011). "Politesse, impolitesse, "non-politesse", "polirudesse": aperçus théoriques et application aux débats politiques télévisuels". In *Cortesia – Politesse – Cortesía*. Edited by Held Gudrun, Uta Helfrich, Frankfurt am Main, Peter Lang. Pag. 93-116.
- Köhler, Wolfgang (1947). *Gestalt Psychology*, New York, Liveright Publishing Corporation.
- Kopczyński, Andrzej (1994). "Quality in conference interpreting: some pragmatic problems". In *Translation Studies: An Interdiscipline*. Edited by Mary Snell-Hornby, Franz Pöchhacker and Klaus Kaindl, Amsterdam/Philadelphia, John Benjamins. Pag. 189-198.
- Kravchenko, Alexander (2008). *Biology of Cognition and Linguistic Analysis. From Non-Realist Linguistics to a Realistic Language Science*, Frankfurt am Main, Peter Lang.
- Kurz, Ingrid (1993). "Conference interpretation: expectations of different user groups". In *The Interpreter's Newsletter* 5. Pag. 13-21.
- Kurz, Ingrid (1993-2002). "Conference interpretation: expectations of different user groups". In *The Interpreting Studies Reader*. Edited by Franz Pöchhacker and Miriam Shlesinger. London/New York, Routledge. Pag. 313-324.
- Kurz, Ingrid; Pöchhacker, Franz (1995). "Quality in TV Interpreting". In *Traslato-Nouvelles de la FIT- FIT Newsletter* 14, 3-4. Pag. 350-358.
- Lakoff, George (1987). *Women, Pire and Dangerous Things*, Chicago-London, Chicago University Press.

- Lamberger-Felber, Heike (1998). *Der Einfluss kontextueller Faktoren auf das Simultandolmetschen. Eine Fallstudie am Beispiel gelesener Reden*, unpublished doctoral dissertation, Karl-Franzens-Universität Graz.
- Lamberger-Felber, Heike (2001.) "Text oriented research into interpreting. Examples from a case-study". In *Hermes* 26. Pag. 39-64.
- Landi, Rossella; Savy, Renata (1996). "Durata vocalica, struttura sillabica e velocità di eloquio nel parlato connesso". In *Atti del XXIV Convegno nazionale dell'Associazione italiana di acustica*. Edited by Alessandro Peretti, Paolo Simonetti, Roma, Esagrafica. Pag. 65-70.
- Langacker, Ronald W. (1987). *Foundations of Cognitive Grammar, vol. I: Theoretical Prerequisites*, Stanford, Stanford University Press.
- Laspia, Patrizia (1997). *L'articolazione linguistica. Origini biologiche di una metafora*, Roma, Nuova Italia Scientifica.
- Laver, John (1980). *The Phonetic Description of Voice Quality*, Cambridge, Cambridge University Press.
- Laver, John (1994). *Principles of Phonetics*, Cambridge, Cambridge University Press.
- Laviosa, Sara (2002). *Corpus-based translation studies: theory, findings, applications*, Amsterdam-New York, Rodopi.
- Levelt, Willem J.M. (1989). *Speaking. From Intention to Articulation*, Cambridge-London, The MIT Press.
- Mack, Gabriele (1999). "L'interpretazione in TV: vecchie e nuove ipotesi di ricerca". Proceedings of the *Congresso di fondazione dell'Associazione italiana di linguistica applicata AITLA*, Pisa 22-23.10.1999. On-line: <<http://home.sslmit.unibo.it/~aitla/pisa/mack.htm>> (last access 03.2012).
- Mack, Gabriele; Cattaruzza, Lorella (1995). "User surveys in SI: a means of learning about quality and/or raising some reasonable doubts". In *Topics in Interpreting Research*. Edited by Jorma Tommola. Turku, Centre for Translation and Interpreting, University of Turku. Pag. 37-49.
- Magno Caldognetto, Emanuela; Cavicchio, Federica; Cosi, Piero, eds. (2008). *Comunicazione parlata e manifestazione delle emozioni*, Atti del I convegno del Gruppo di Studio sulla Comunicazione Parlata (Padova, 29 novembre-1 dicembre 2004), Napoli, Liguori (e-book).
- Magno Caldognetto, Emanuela; Zmarich, Claudio; Ferrero, Franco (1998). "Indici acustici macroprosodici dello stato emotivo del parlante". In *Atti del XXVI Convegno nazionale dell'Associazione italiana di acustica* (Torino, 27-29 maggio 1998). Torino, IEN. Pag. 263-266.

- Marchese, Angelo (1978). *Dizionario di retorica e di stilistica*. Milano, Mondadori.
- Marotta, Giovanna (2001) "Non solo spiranti. La «gorgia toscana» nel parlato di Pisa". In *L'Italia dialettale*, n.s. XXXIX, vol. LXII. Pag. 27-60.
- Marotta, Giovanna (2005). "Il consonantismo romano. Processi fonologici e aspetti acustici". In *Italiano parlato. Analisi di un dialogo*. Edited by Federico Albano Leoni, Rosa Giordano, Napoli, Liguori. Pag. 1-24.
- Marradi, Alberto (2007). *Metodologia delle scienze sociali*. Edited by Rita Pavsic and Maria Concetta Pitrone. Bologna, il Mulino.
- Marrone, Stefano (1993). "Quality: a shared objective". In *The Interpreter's Newsletter* 5. Pag. 35-41.
- Mason, Ian (2001). "Translator behaviour and language use: some constraints on contrastive studies". In *Hermes* (26). Pag. 65-80.
- Maturana, Humberto R.; Varela, Francisco J. (1980). *Autopoiesis and Cognition. The Realization of Living*, Dordrecht, Reidel [Italian translation by Alessandra Stragapede (2001). *Autopoiesi e cognizione. La realizzazione del vivente*, Venezia, Marsilio].
- McCarthy, John J. (2002). *A Thematic Guide to Optimality Theory*, Cambridge, Cambridge University Press.
- Meak, Lidia (1990). "Interprétation simultanée et congrès medical: attentes et commentaires". In *The Interpreter's Newsletter*, 3. Pag. 8-13.
- Mehler, Jacques; Dupoux, Emmanuel (1990). *Naître humain*, Paris, Odile Jacob [Italian translation by Elena Mohlo (1992). *Appena nato. Che cosa vede, sente, capisce un bambino sin dai primi giorni di vita*, Milano, Mondadori].
- Meschonnic, Henri (1982). *Critique du rythme. Anthropologie historique du langage*. Verdier, Lagrasse.
- Meyer, Bernd (1998). "What transcriptions of authentic discourse can reveal about interpreting". In *Interpreting*, 3. Pag. 65-83.
- Mill, John Stuart (1877). *System der deduktiven und induktiven Logik. Eine Darlegung der Prinzipien wissenschaftlicher Forschung, insbesondere der Naturforschung*. [Translated into German from the 8<sup>th</sup> edition by J. Schiel. 4<sup>th</sup> German edition].
- Mill, John Stuart (1974). *A system of logic ratiocinative and inductive. Being a connected view of the principles of evidence and the methods of scientific investigation*. In *Collected works of John Stuart Mill*, Vol. VII, Toronto, University of Toronto Press.
- Moser, Peter (1995). "Survey on expectations of users of conference interpretation. Final report commissioned by AIIC". On-line:



- <<http://aiic.net/page/736/survey-on-expectations-of-users-of-conference-interpretation/lang/1>> (last access 10.2014).
- Moser, Peter (1996). "Expectations of users of conference interpretation". In *Interpreting* 1:2. Pag. 145-178.
- Moser-Mercer, Barbara; Künzly, Alexander; Korac, Marina (1998). "Prolonged turns in interpreting: Effects on quality, physiological and psychological stress". In *Interpreting* 3:1. Pag. 47-64.
- Munro, Murray J.; Derwing, Tracey M. (1995). "Foreign accent, comprehensibility, and intelligibility in the speech of second language learners". In *Language Learning* 45. Pag. 73-97.
- Munro, Murray J.; Derwing, Tracey M. (1999). "Foreign accent, comprehensibility, and intelligibility in the speech of second language learners". In *Language Learning* 49. Pag. 285-310.
- Munro, Murray J.; Derwing, Tracey M. (2001). "Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate". In *Studies of Second Language Acquisition* 23. Pag. 451-468.
- Myers, Scott (2000). "Boundary disputes: The distinction between phonetic and phonological sound patterns". In *Phonological Knowledge*. Edited by Noel Burton-Roberts, Philip Carr, Gerard Docherty, Oxford, Oxford University Press. Pag. 245-272.
- Nathan, Geoffrey S. (1996). "Steps toward a cognitive phonology". In *Natural Phonology: The State of the Art*. Edited by Bernhard Hurch, Richard A. Rhodes, Berlin-New York, Mouton/de Gruyter. Pag. 107-120.
- Ng, Bee Chin (1992). "End-Users' Subjective Reaction to the Performance of Students Interpreters". In *The Interpreter's Newsletter*, Special Issue No. 1. Pag. 35-41.
- Noseda, Paolo (2008). *Blog Raitre*. On-line: <<http://chetempocheffa.blog.rai.it/2008/11/16/da-paolo-maria-noseda-interprete/>> (last access 06.2012).
- O'Connor, Joseph Desmond; Arnold, Gordon Frederick (1961). *Intonation of Colloquial English. A Practical Handbook*, London, Longman.
- Oliverio, Alberto (1999) *Biologia e filosofia della mente*, Roma-Bari, Laterza.
- Oppenheimer, Daniel M. (2008). "The secret life of fluency". In *Trends in Cognitive Science* 12. Pag. 237-241.
- Osgood, Charles E. (1952). "The nature and measurement of meaning". In *Psychological Bulletin* XLIX, 3 (May). Pag. 197-237.

- Palazzini Finetti, Mariacarla (2000). *Interpretazione simultanea per un festival cinematografico: analisi linguistica e pragmatica dei risultati*, Unpublished MA dissertation, SSLMIT, Università di Trieste.
- Papa, Carlo (2010), *Il fenomeno del self-repair in interpretazione simultanea*, Unpublished MA Dissertation, SSLMIT, Università di Trieste.
- Paul, H. (1920). *Prinzipien der Sprachgeschichte*. Halle-Saale., Niemeyer.
- Pedersen J. (2007). "How is culture rendered in subtitles?". On-line: <[http://euroconferences.info/proceedings/2005\\_Proceedings/2005\\_Pedersen\\_Jan.pdf](http://euroconferences.info/proceedings/2005_Proceedings/2005_Pedersen_Jan.pdf)> (last access 11.2013).
- Penso, Francesca (2014). *Dibattiti presidenziali USA 2012: le autocorrezioni nelle interpretazioni simultanee*. Unpublished MA dissertation, SSLMIT, Università di Trieste.
- Perrot, Jean, ed. (1997). *Polyphonie pour Iván Fónagy. Mélanges offerts à Iván Fónagy par un group de disciples, collègues et admirateurs*, Paris, L'Harmattan.
- Persyn-Vialard, Sandrine (2005). *La linguistique de Karl Bühler. Examen critique de la Sprachtheorie et de sa filiation*, Rennes, Presses Universitaires de Rennes.
- Péterfalvi, Jean-Michel (1978). *Recherches expérimentales sur le symbolisme phonétique*, Paris, Editions du CNRS (Monographies françaises de psychologie, XIX).
- Piaget, Jean (1968/1994). *Le structuralisme*, Paris, PUF [Italian translation by Andrea Bonomi (1994). *Lo strutturalismo*, Milano, Il Saggiatore].
- Pickett, James M., ed. (1999). *The Acoustics of Speech Communication. Fundamentals, Speech perception Theory, and Technology*, Needham Heights, Allwyn and Bacon.
- Pierrehumbert, Janet; Beckman, Mary E.; Ladd, D.Robert. (2000). "Conceptual foundations of phonology as a laboratory science". In *Phonological Knowledge*. Edited by Noel Burton-Roberts, Philip Carr, Gerard Docherty, Oxford, Oxford University Press. Pag. 273-303.
- Pinker, Steven; Jackendoff, Ray (2005). "The Faculty of language: What's special about it?". In *Cognition*, 95, 2. Pag. 201-236. On-line: <<https://ase.tufts.edu/cogstud/jackendoff/papers/FacultyofLanguage.pdf>> (last access 12.2015).
- Pisoni, David B. (1986). "Contextual variability and the problem of acoustic-phonetic invariance in speech". In *Invariance and variability in speech processes*. Edited by Joseph S. Perkell, Dennis H. Klatt, New York-London, Psychology Press. Pag. 154-161.

- Pöchhacker, Franz (2011). “Researching TV interpreting: selected studies on US Presidential material”. In *The Interpreters’ Newsletter* 16. Trieste, EUT. Pag. 21-36.
- Pöchhacker, Franz; Zwischenberger, Cornelia (2010). “Survey on quality and role: conference interpreters’ expectations and self-perceptions”. On-line: <http://aiic.net/page/3405/survey-on-quality-and-role-conference-interpreters-expectations-and-self-perceptions/lang/1> (last access 10.2014).
- Pradas Macías, Esperanza Macarena (2003). *Repercusión del intraparámetro pausas silenciosas en la fluidez: Influencias en las expectativas y en la evaluación de la calidad en interpretación simultánea*. Unpublished Ph.D. dissertation, Universidad de Granada.
- Pradas Macías, Esperanza Macarena (2006). “Probing quality criteria in simultaneous interpreting. The role of silent pauses in fluency”. In *Interpreting* 8:1. Pag. 25-43.
- Prado Pico, Emili et al. (1997). *El Modelo Acústico de Credibilidad (MAC) de la voz en el marco de la gramática de la expresión fonostésica. Memoria de investigación PB93-0850 de la DGICYT, Departament de Comunicació Audiovisual i Publicitat de la UAB*. Vilanova i la Geltrú, Universitat Autònoma de Barcelona.
- Prince, Alan; Smolensky, Paul (1993) *Optimality Theory: Constraint Interaction in Generative Grammar*, New Brunswick, Rutgers University Center for Cognitive Science.
- Reber, Rolf; Schwarz, Norbert (1999). “Effects of perceptual fluency on judgments of truth”. In *Consciousness and Cognition* 8. Pag. 338-342.
- Reber, Rolf; Unkelbach, Christian (2010). “The epistemic status of processing fluency as source for judgments of truth”. In *Review of Philosophy and Psychology* 1:4. Pag. 563-581.
- Reis, Carlos; Lopes, Ana Cristina Macário (1990). *Dicionário de Narratologia*. Coimbra, Edições Almedina [Spanish translation by Ángel Marcos de Dios (2002). *Diccionario de Narratología*. Salamanca, Almar].
- Reithofer, Karin (2013). “Comparing modes of communication. The effect of English as a lingua franca vs. interpreting”. In *Interpreting* 15:1. Pag. 48-73.
- Rennert, Sylvi (2010). “The impact of fluency on the subjective assessment in interpreting quality”. In *The Interpreters’ Newsletter* 15. Pag. 101-115.
- Riccardi, Alessandra (2003). “The relevance of interpreting strategies for defining quality in simultaneous interpreting”. In *La evaluación de la*

- calidad en interpretación: investigación*. Edited by Ángela Collados Aís, María Manuela Fernández Sánchez, Daniel Gile, Granada, Granada, Comares. Pag. 257-265.
- Roca, Iggy; Johnson, Wyn (1999). *A Course in Phonology*, Oxford, Blackwell.
- Rodríguez Bravo, M. Ángel (1989) *La construcción de una voz radiofónica*. Unpublished Ph.D. dissertation, Universidad Autónoma de Barcelona.
- Rossi, Mario (1999). *L'intonation. Le système du français: description et modélisation*, Paris, Ophrys.
- Russo, Mariachiara (2005). "Simultaneous film interpreting and users' feedback". In *Interpreting*, 7:1. Pag. 1-26.
- Sapir, Edward (1925). "Sound patterns in language". In *Language*, 1. Pag. 37-51.
- Schiffman, Susan; Reynolds, M. Lance; Forrest W. Young (1981). *Introduction to multidimensional scaling: Theory, methods and applications*, New York, Academic Press.
- Schleifer, Ronald (1987) *A.J. Greimas and the nature of meaning: linguistics, semiotics and discourse theory*. London and Sydney, Croom Helm.
- Schneider, Marius (1970). *Il significato della musica*. Milano, Rusconi.
- Schwarthorer, Micaela (2010). *Mediendolmetschen aus Zuschauer- und Zuhörersicht am Beispiel der Amtsantrittsrede von Präsident Obama*. Unpublished MA dissertation, University of Vienna.
- Shlesinger, Miriam (1989). "Extending the theory of translation to interpretation: Norm as a case in point". In *Target* 1. Pag. 111-115.
- Shlesinger, Miriam (1994). "Intonation in the Production and Perception of Simultaneous Interpretation". In *Bridging the Gap: Empirical Research in Simultaneous Interpretation*. Edited by Sylvie Lambert and Barbara Moser-Mercer. Amsterdam/Philadelphia, John Benjamins. Pag. 225-236.
- Shlesinger, Miriam (1998). "Corpus-based Interpreting Studies as an off-shoot of corpus-based Translation Studies". In *Meta* 43:4. Pag. 486-493.
- Soler Caamaño, Emma (2006). *La calidad en formación especializada en interpretación: Análisis de los criterios de evaluación de un jurado en un posgrado de interpretación de conferencia médica*. Ph.D. dissertation, Universitat Pompeu Fabra.
- Sorianello, Patrizia; Calamai, Silvia (2005). "Il sistema vocalico romano". In *Italiano parlato. Analisi di un dialogo*. Edited by Federico Albano Leoni, Rosa Giordano, Napoli, Liguori. Pag. 25-69.

- Soto Sanfiel, María Teresa (2000). *Influencia de la percepción visual del rostro del hablante en la credibilidad de su voz*. Ph.D. dissertation, Universitat Autònoma de Barcelona.
- Stampe, David (1979). *A Dissertation on Natural Phonology*, Bloomington, Indiana University Linguistics Club.
- Straniero Sergio, Francesco (2003). "Norms and quality in media interpreting: The case of formula one press conferences", *The Interpreters' Newsletter* 12. Pag. 135-174.
- Straniero Sergio, Francesco (2007). *Talkshow Interpreting. La mediazione linguistica nella conversazione spettacolo*. Trieste, Edizioni Università di Trieste.
- Tissi, Benedetta (2000). "Silent pauses and disfluencies in simultaneous interpretation: A descriptive analysis". In *The Interpreter's Newsletter* 10. Pag. 103-127.
- Tognini-Bonelli, Elena (2000) "'Unità funzionali complete' in inglese e in italiano: verso un approccio corpus-driven". In *I Corpora nella didattica della traduzione*. Edited by Silvia Bernardini, Federico Zanettin, Bologna, CLUEB. Pag. 153-175.
- Tohyama, Hitomi; Matsubara, Shigeki (2006). "Influence of pause length on listeners' impressions in simultaneous interpretation". In *Proceedings of the Ninth International Conference on Spoken Language Processing (Interspeech 2006 – ICSLP)*, Pittsburgh, Pennsylvania, 17-21 September. Pag. 893-896.
- Tomasetig, Arianna (2013). *Dibattiti presidenziali USA 2012: gli elementi culturali nelle interpretazioni simultanee in italiano e in spagnolo*. Unpublished MA dissertation, SSLMIT, Università di Trieste.
- Trubeckoj, S. Nikolaj (1939/1971). *Grundzüge der Phonologie*, Prague, Travaux du Cercle Linguistique de Prague, VII [Italian translation by Giulia Mazzuoli Porro (1971). *Fondamenti di fonologia*, Torino, Einaudi].
- Tymoczko, Maria (1998). "Computerized Corpora and the Future of Translation Studies". In *Meta* (43), 4. Pag. 652-660.
- Valentini, Cristina (2009). *Creazione e sviluppo di corpora multimediali. Nuove metodologie di ricerca nella traduzione audiovisiva*. Ph.D. Dissertation, Università di Bologna.
- Vallone, Marianna; Caniparoli, Valentina; Savy, Renata (2002). "Una prima indagine sui fenomeni di riduzione consonantica nella varietà napoletana del corpus Avip". In *Atti delle XIII Giornate di studio del Gruppo di Fonetica Sperimentale-AIA*, (Macerata, 13-15 dicembre 2001). Edited by Agostino Regnicoli, Roma, Il Calamo. Pag. 83-90.

- Veneziano, Corrado (2001). *Manuale di dizione, voce e respirazione*. Lecce, Besa.
- Venuti, Lawrence (1995). *The translator's invisibility*. London-New York, Routledge.
- Vermeer, Hans J. (1996). *A Skopos Theory of Translation (Some Arguments For and Against)*. Heidelberg, TEXTconTEXT.
- Viezzi, Maurizio (1996). *Aspetti della qualità in interpretazione*. Trieste, Università degli studi di Trieste - Scuola Superiore di Lingue Moderne per Interpreti e Traduttori.
- Viezzi, Maurizio (2004), *Denominazioni proprie e traduzione*, LED, Milano.
- Vilches, Lorenzo (1989). *Manipulación de la información televisiva*. Barcelona, Paidós.
- Vuorikoski, Anna Riitta (1993). “Simultaneous interpretation- user experience and expectations”. In *Translation- The Vital Link. XIII FIT World Congress, Proceedings*. Edited by Catriona Picken. Brighton, Institute of Translation and Interpreting. Pag. 317-327.
- Werbik, Hans (1969). “L’indétermination et les qualités impressives des modèles stimulants mélodiques”. In *Sciences de l’Art-Scientific Aesthetics* VI. Pag. 25-37.
- Zani, Marzia (2002). *Organizzazione del discorso e perdita di informazioni nell’interpretazione simultanea dallo spagnolo in italiano*. Unpublished MA dissertation, SSLMIT, Università di Bologna.
- Zumthor, Paul (1983). *Introduction à la poésie orale*, Paris, Seuil [Italian translation by Costanzo Di Girolamo (2001). *La presenza della voce. Introduzione alla poesia orale*, Bologna, Il Mulino].
- Zwischenberger, Cornelia; Behr, Martina (2013). *Interpreting quality: A look around and ahead*. Berlin, Frank&Timme.

# VOICE QUALITY AND TV INTERPRETING: A PROPOSAL FOR A GESTALTIC EVALUATION

---

## APPENDICES

APPENDIX 1:	I
Chronological and contrastive table of evaluation criteria used in surveys on quality expectations (ideal preferences) and quality assessment (real judgment) of simultaneous interpreting.	I
APPENDIX 2:	V
Questionnaire – draft 1	V
APPENDIX 3:	VIII
Questionnaire – draft 2	VIII
APPENDIX 4:	XII
Questionnaire – draft 3 (pre-test)	XII
APPENDIX 5:	XVI
Questionnaire – draft 4	XVI
BREVE COMMENTO ALLA BOZZA 4:	XXII
Questionnaire – draft 5	XXIII
Questionnaire – draft 6	XXVII
Questionnaire – draft 7	XXXI
APPENDIX 6:	XXXV
Questionnaire – Last version (video sequence: 1-2-3)	XXXV
APPENDIX 7:	L
Results from pre-test (questionnaire - draft 3)	L

<b>APPENDIX 8:</b>	LVI
Grid for the selection of video excerpts	LVI
<b>APPENDIX 9:</b>	LX
WinPitch measurements of silent pauses and audible breaths (video excerpts for pilot survey).	LX
APPENDIX 10:	LXIV
Frequencies of ratings (pilot survey)	LXIV
APPENDIX 11:	LXXII
Bivariate correlation of variables (pilot survey)	LXXII
APPENDIX 12:	LXXIX
Perceptual maps of variables by multidimensional scaling (pilot survey)	LXXIX
APPENDIX 13:	LXXXIII
Analysis of principal components – following scree plots figures 4.30 – 4.32)	LXXXIII
APPENDIX 14:	LXXXVI
Open-ended questions (pilot survey)	LXXXVI



## **APPENDIX 1:**

**Chronological and contrastive table of evaluation criteria used in surveys on quality expectations (ideal preferences) and quality assessment (real judgment) of simultaneous interpreting.**



# 1. Chronological and contrastive table of evaluation criteria used in surveys on

STUDY / SURVEY	Bühler 1986	Kurz 1989	Gile 1990	Meak 1990	Chin Ng 1992	Kurz 1993	Marrone 1993	Vuorikoski 1993	Kopczynski 1994	Moser 1995	Mack & Cattaruzi	
INTERPRETING MODE / TYPE OF SURVEY	SI - Expectations	SI - Expectations	SI - Assessment	SI - Expectations	SI - Assessment	SI - Expectations	CI - Expectations + Assessment	SI - Assessment	SI - Expectations	SI - Expectations	SI - Expectations and Assessment	
USERS	Interpreters	Users: medical doctors	Users: medical doctors	Users: medical doctors	Users	Users: medical doctors, engineers, Council of Europe delegates	Users: Constitutional Law experts	Users: seminaries participants	Users	Conference participants (listeners + speakers)	Users	
TOPIC				medicine				Foster care, advertising and marketing, economics crimes, European agricultural policy		different topics (specialised subject matters)		
EVALUATION CRITERIA / PARAMETERS	Native accent	Native accent				Native accent				5. speak in a lively and animated way		
	Pleasant voice	Pleasant voice				Pleasant voice			1b. Voice qualities 1b. Diction		4. Ritmo 5. Score	
	Fluency of delivery	Fluency of delivery	Voix, rythme et intonation / Quality of voice and delivery	7. La rapidité constitue-t-elle un facteur de dérangement, nuit-elle à la compréhension ou ne joue-t-elle aucun rôle?	(Stage 2) 1. As a [Japanese] native speaker, what do you think of the inappropriate use of speech levels in conference interpreting? Is it important? 2. Do you think it interferes with the content delivery? 3. Do you find it offensive? 4. Does it make you feel uncomfortable?	Fluency of delivery	8. Rhythm of discourse was pleasant 8. Discourse was fluent	8. Interpretation was coherent and easy to follow	1b. Fluency of delivery 2b. Lack of fluency 2b. Poor diction 2b. Monotonous intonation 2b. Monotonous tempo 2b. Speeding up and slowing down dynamics of speech - tempo, intensity of voice, gestures?	2. Clarity of expression		
	Logical cohesion of utterance	Logical cohesion of utterance				Logical cohesion of utterance						
	Sense consistency with original message	Sense consistency with original message	Fidelity	2. Dans une interprétation simultanée qu'est-ce qui vous dérange le plus: a) l'impropriété (mais pas l'erreur) des termes? b) le caractère incomplet des exemples ou leur omission? c) le caractère incomplet, l'omission ou l'erreur de chiffres ou de données? Dites pourquoi et dans quelle mesure 4. Dans un tableau quelles sont les données qu'il est indispensable d'indiquer? 5. Lorsqu'on commente un film (par exemple une intervention chirurgicale) on une série de diapositives, qu'est-il indispensable de traduire? 6. Y a-t-il des branches de la médecine qui nécessitent une précision particulière? 8. Les conclusions sont-elles facilement prévisibles ou sont-elles importantes? 9. Les abréviations sont-elles compréhensibles ou exigent-elles absolument une traduction?		Sense consistency with original message			3. Should the interpreter empathize with the speaker's intention? 5. Should the interpreter remain in the background or be visible? 6. Should the interpreter correct a speaker's mistakes?	4. Faithfulness of the meaning of the original 4. literal reproduction of what is said	3. precision	
	Completeness of interpretation	Completeness of interpretation				Completeness of interpretation			1b. Rendering the general content of T1 1b. Rendering the general content of T1 2b. Incomplete sentences 2b. Too general rendition of content 2b. Too detailed rendition of content 7. Is the interpreter allowed to summarize the speech? 8. Is the interpreter allowed to add his or her own explanations to clarify what the speaker has said?	2. Completeness of rendition 3. focus on essentials 3. give a complete rendition 5. Speak in complete sentences 5. interpret titles, name of functions, positions, office held 5. interpret subtitles in graphs and tables on transparencies or on slides 5. interpret abbreviations		
	Correct grammatical usage	Correct grammatical usage				Correct grammatical usage		8. Interpretation was correct	1b. Grammatical correctness of utterances 2b. Ungrammatical sentences			6. Terminological adequacy
	Use of correct terminology	Use of correct terminology	Terminological usage			Use of correct terminology			Terminology used was correct	1b. Terminological precision 2b. Faulty terminology	2. Correct terminology	
	Use of appropriate style									1b. Style 2b. Stylistic mistakes		
	Thorough preparation of conference documents				3. Est-il important de connaître la fonction, le ou les titres de l'orateur, le centre hospitalier où il travaille ou il a travaillé, l'amitié qui le lie au modérateur, les réunions auxquelles il a déjà participé, l'institution à laquelle il appartient? (Indiquer pourquoi et dans quelle mesure ou dans quelle circonstance).				8. Interpreters were well informed			1. Denotative consciousness
	endurance											
	poise											
	Pleasant appearance											
	reliability								Interpreters' professionalism is satisfactory			
Ability to work in a team												
Positive feedback from delegates												
Other criteria (please specify):			Other							2. Other		
			Linguistic output quality		(Stage 1) 2. General impression of the Japanese used by each interpreter?				10. Did you find the presentations interesting? 11. I think I understood what speakers wanted to communicate and points they wanted to highlight.			
			General quality of the interpretation		(Stage 1) 1. Can you follow the interpretation? Do you get the interpreter's message?						2. Fidelity	
			Main weaknesses of interpretation	1. L'interprétation simultanée vous permet-elle de suivre un congrès médical dont vous ne connaissez pas les langues de travail?	(Stage 1) 3. best/worst interpretation by which interpreter (ranking). Brief explanation of ranking.							



**APPENDIX 2:****Questionnaire – draft 1****BOZZA DEL QUESTIONARIO PER TELESPETTATORI**Sesso: M  F 

Età: \_\_\_\_\_

Professione: \_\_\_\_\_

Istruzione: \_\_\_\_\_

**D1 GRADEVOLEZZA DELLA VOCE**

Ritiene che la voce dell'interprete sia gradevole?

 per niente       poco  abbastanza  molto       moltissimo**D2 PRONUNCIA**

Ritiene che l'interprete abbia pronunciato le parole in maniera chiara e netta?

 per niente       poco  abbastanza  molto       moltissimo**D3 VELOCITÀ DI ELOQUIO**

Quanto velocemente ritiene che l'interprete abbia parlato per favorire la comprensione dell'ascoltatore?

 per niente       poco  abbastanza  molto       moltissimo

#### D4 INTONAZIONE

Quanto bene ritiene che l'interprete abbia usato la voce per far capire meglio quello che ha detto?

per niente                       poco  abbastanza    molto                       moltissimo

#### D5 INTONAZIONE

Quanto bene ritiene che l'interprete abbia usato il tono della voce per trasmettere anche il comportamento dell'oratore che si vede nel video?

per niente                       poco  abbastanza    molto                       moltissimo

#### D6 REGISTRO

In generale, ritiene che l'interprete abbia usato le parole appropriate per un discorso del genere?

per niente                       poco  abbastanza    molto                       moltissimo

#### D7 COESIONE

Ritiene che l'interprete abbia costruito le frasi in modo corretto?

per niente                       poco  abbastanza    molto                       moltissimo

#### D8 RITMO

Ora, volendo unire tutti gli aspetti giudicati nelle precedenti domande in un unico aspetto, che chiamiamo ritmo, quanto buono ritiene che sia?

per niente                       poco  abbastanza    molto                       moltissimo

#### D9 ESPRESSIONE

Quanto bene ritiene che il ritmo abbia favorito la chiarezza d'espressione dell'interprete?

per niente                       poco  abbastanza    molto                       moltissimo

D10 COMPrensIONE

Quanto bene ritiene che la chiarezza d'espressione abbia favorito la comprensione dell'interpretazione simultanea?

per niente                       poco  abbastanza    molto                       moltissimo

D11 COMPrensIONE

Quanto bene ritiene che il ritmo dell'interprete abbia favorito la comprensione dell'interpretazione simultanea?

per niente                       poco  abbastanza    molto                       moltissimo

D12-D13 COMPrensIONE EFFETTIVA DEL TESTO INTERPRETATO

Domande specifiche sulla comprensione del testo.

## APPENDIX 3:

### Questionnaire – draft 2

#### BOZZA DEL QUESTIONARIO PER telespettatori

##### D0 – DATI PERSONALI

Sesso:            M    F

Età: \_\_\_\_\_

Occupazione: \_\_\_\_\_

Grado di istruzione: \_\_\_\_\_

Indichi la sua scelta (può indicarne anche più di una) barrando con una “x” l’opzione proposta.

D1   LA VOCE DELL’INTERPRETE APPARTIENE:  
maschio      femmina

D2   LA VOCE DELL’INTERPRETE È:  
gradevole    amabile    dolce    chiara    noiosa    lagnosa    irritante  
monotona

D3   NEL COMPLESSO RISULTA:  
abbastanza chiara      appena comprensibile      indecifrabile

Le chiediamo di esprimere un giudizio in merito agli aspetti di seguito proposti, utilizzando una scala da 0 a 7 dove 0 indica disaccordo rispetto a quanto affermato.

Indichi la sua scelta barrando con una “x” il valore corrispondente.



D4 LA VOCE DELL'INTERPRETE È GLOBALMENTE INTELLEGIBILE.

0 1 2 3 4 5 6 7

D5 LA PRONUNCIA RISULTA CHIARA.

0 1 2 3 4 5 6 7

D4 L'INTERPRETE HA PARLATO VELOCEMENTE.

0 1 2 3 4 5 6 7

D5 L'INTERPRETE ACCENTUA ALCUNE PAROLE PER METTERLE IN RISALTO.

0 1 2 3 4 5 6 7

D6 L'INTERPRETE HA PARLATO "CON MELODIA".

0 1 2 3 4 5 6 7

D6 L'INTERPRETE HA PARLATO IN MODO SCORREVOLE.

0 1 2 3 4 5 6 7

D7 L'INTERPRETE HA EFFETTUATO DELLE PAUSE.

0 1 2 3 4 5 6 7

D8 L'INTERPRETE HA AVUTO DELLE ESITAZIONI.

0 1 2 3 4 5 6 7

D10 L'INTERPRETE ACCENTUA ALCUNE PAROLE PER METTERLE IN RISALTO.

0 1 2 3 4 5 6 7

D11 L'INTERPRETE HA UTILIZZATO PAROLE APPROPRIATE PER IL DISCORSO CHE HA INTERPRETATO.

0 1 2 3 4 5 6 7

D12 L'INTERPRETE HA COSTRUITO CORRETTAMENTE LE FRASI.

0 1 2 3 4 5 6 7

D12 L'INTERPRETE HA COSTRUITO CORRETTAMENTE LE FRASI.

0 1 2 3 4 5 6 7

D13 NEL COMPLESSIVO INSIEME DELLA SPETTACOLARITÀ NELLE PRECEDENTI DOMANDE (voce + pause + melodia + esitazioni + parole + frasi) È ACCETTABILE.

0 1 2 3 4 5 6 7

D14 NEL COMPLESSIVO ESPRESSIONE DELL'INTERPRETE CHIARA.

0 1 2 3 4 5 6 7

D15 NEL COMPLESSO, LA MIA COMPrensIONE È SUFFICIENTE.

0 1 2 3 4 5 6 7

*Le chiediamo di esprimere liberamente le sue osservazioni in merito al questionario o di indicare aspetti che secondo lei non sono contemplati nelle domande.*

---



---



---



---



---



---



---

## APPENDIX 4:

### Questionnaire – draft 3 (pre-test)

#### BOZZA 3 DEL QUESTIONARIO per interpreti

##### D0 – DATI PERSONALI

Sesso:       M     F

Età: \_\_\_\_\_

Professionista / Studente: \_\_\_\_\_

Lingue      (A,      B      /      A,      C,      C)

\_\_\_\_\_

Anni di esperienza anni di studi di interpretazione:

\_\_\_\_\_

Indichi la sua scelta (può indicarne anche più di una) barrando con una "x" l'opzione proposta.

D1    LA VOCE DELL'INTERPRETE APPARTIENE A:  
maschio      femmina

D2    LA VOCE DELL'INTERPRETE È:  
gradevole    amabile    dolce    chiara    noiosa    lagnosa    irritante  
monotona

Altro (specificare): \_\_\_\_\_

D3 NEL COMPLESSO RISULTA:

abbastanza chiara      appena comprensibile      indecifrabile

Le chiediamo di esprimere un giudizio in merito agli aspetti di seguito proposti, utilizzando una scala da 0 a 7 dove 0 indica disaccordo rispetto a quanto affermato.

Indichi la sua scelta barrando con una "x" il valore corrispondente.

D4 LA VOCE DELL'INTERPRETE È GLOBALMENTE INTELLEGIBILE.

0    1    2    3    4    5    6    7

D5 LA PRONUNCIA RISULTA CHIARA.

0    1    2    3    4    5    6    7

D6 LA VELOCITÀ DI ELOQUIO È TENDENZIALMENTE ALTA.

0    1    2    3    4    5    6    7

D7 L'INTERPRETE ACCENTUA ALCUNE PAROLE PER METTERLE IN RISALTO (accento di frase).

0    1    2    3    4    5    6    7

D8 LA CURVA MELODICA DEL PARLATO PRESENTA UNA GRADEVOLE ESCURSIONE TONALE.

0 1 2 3 4 5 6 7

D9 L'ELOQUIO DELL'INTERPRETE È SCORREVOLE E FLUIDO (velocità di eloquio media).

0 1 2 3 4 5 6 7

D10 L'INTERPRETE EFFETTUA DELLE PAUSE PIENE (allungamenti vocalici e consonantici, esitazioni vocalizzate).

0 1 2 3 4 5 6 7

D11 L'INTERPRETE AVUTA INTERRUZIONI e ritardamenti, e partenze, correzioni, ristrutturazioni).

0 1 2 3 4 5 6 7

D12 L'INTERPRETE HA UTILIZZATO UN LESSICO PERTINENTE AL DISCORSO.

0 1 2 3 4 5 6 7

D13 IL DISCORSO PRESENTA UN'ALTA DENSITÀ LESSICALE.

0 1 2 3 4 5 6 7

D14 L'INTERPRETE HA COSTRUITO CORRETTAMENTE LE FRASI.

0 1 2 3 4 5 6 7

D15 L'INTERPRETE HA COSTRUITO FRASI COMPLESSE.

0 1 2 3 4 5 6 7

D16 NEL COMPLESSO, INSIEME GLI ASPETTI MALUTATI NELLE PRECEDENTI DOMANDE (voce + pause + melodia + esitazioni + parole + frasi) È ACCETTABILE.

0 1 2 3 4 5 6 7

D17 NEL COMPLESSO, L'ESPRESSIONE DELL'INTERPRETE RISULTA CHIARA.

0 1 2 3 4 5 6 7

D18 NEL COMPLESSO, LA MIA COMPrensIONE È SUFFICIENTE.

0 1 2 3 4 5 6 7

D19 CONOSCEVO LE INFORMAZIONI DAL TESTO ASCOLTATO.

0 1 2 3 4 5 6 7

*Le chiediamo di esprimere liberamente le sue osservazioni in merito al questionario o di indicare aspetti che secondo lei non sono contemplati nelle domande.*

---



---



---



---



---



---

## APPENDIX 5:

### Questionnaire – draft 4

<b>BOZZA 4 DEL QUESTIONARIO PER interpreti</b>	
<b>D0 – DATI PERSONALI</b>	
Sesso:	<input type="checkbox"/> M <input type="checkbox"/> F
Età:	_____
Professionista	/ Studente:
_____	
Lingue	(A, B / A, C, C)
_____	
_____	
Anni di esperienza / anni di studio di interpretazione: _____	
<u><i>Indichi la sua scelta (può indicarne anche più di una) barrando con una "x" l'opzione proposta.</i></u>	
<b>D1</b>	<b>LA VOCE DELL'INTERPRETE APPARTIENE</b>
maschio	femmina



D2 LA VOCE DELL'INTERPRETE È:

gradevole amabile dolce chiara noiosa lagnosa  
irritante monotona

Altro (specificare): \_\_\_\_\_

D3 NEL COMPLESSO RISULTA:

abbastanza chiara appena comprensibile  
indecifrabile

Le chiediamo di esprimere un giudizio in merito agli aspetti di seguito proposti, utilizzando una scala da 0 a 7 dove 0 indica disaccordo rispetto a quanto affermato.

Indichi la sua scelta barrando con una "x" il valore corrispondente.

D4 LA VOCE DELL'INTERPRETE È GLOBALMENTE INTELLEGIBILE.

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Non sono affatto d'accordo sono  
assolutamente d'accordo

D5 LA PRONUNCIA ITALIANA RISULTA CHIARA.

Non sono  
affatto  
d'accordo

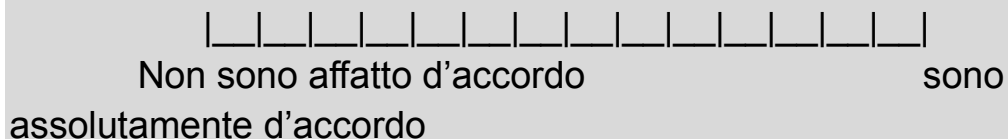
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Sono  
assolutamente  
d'accordo

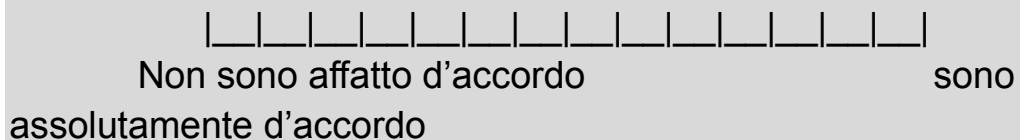
D6 LA VELOCITÀ DI ELOQUIO È TENDENZIALMENTE ALTA.



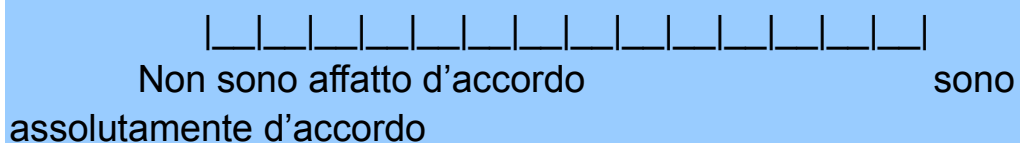
D7 L'INTERPRETE HA ACCENTUATO ALCUNE PAROLE PER METTERLE IN RISALTO (accento di frase).



D8 LA CURVA MELODICA DEL PARLATORE PRESENTA UNA GRADEVOLE ESCURSIONE TONALE.



D9 LA PARLATA DELL'INTERPRETE È IN TUTTA CANTILENA (Ai petiziforsetidi di dsi anamonia della voce)



D10 LA PARLATA DELL'INTERPRETE È CARATTERIZZATA DA UN FASTIDIOSO ACCENTO REGIONALE

Non sono affatto d'accordo sono  
 assolutamente d'accordo

D11 L'ELOQUIO DELL'INTERPRETE RISULTA SCORREVOLE / FLUIDO (velocità di eloquio media).

Non sono affatto d'accordo sono  
 assolutamente d'accordo

D12 L'INTERPRETE HA EFFETTUATO DELLE PAUSE PIENE (allungamenti vocalici e consonantici, esitazioni vocalizzate).

Non sono affatto d'accordo sono  
 assolutamente d'accordo

D13 L'INTERPRETE HA AVUTO INTERRUZIONI (ripetizioni, false partenze, correzioni, ristrutturazioni).

Non sono affatto d'accordo sono  
 assolutamente d'accordo

D14 L'INTERPRETE HA UTILIZZATO IL LESSICO PERTINENTE AL DISCORSO.

Non sono affatto d'accordo sono  
 assolutamente d'accordo

D15 IL DISCORSO PRESENTA UN'ALTA DENSITÀ LESSICALE.

Non sono affatto d'accordo sono  
 assolutamente d'accordo

D16 L'INTERPRETE HA COSTRUITO CORRETTAMENTE LE FRASI.

Non sono affatto d'accordo sono  
 assolutamente d'accordo

D17 L'INTERPRETE HA COSTRUITO FRASI COMPLESSE.

Non sono affatto d'accordo sono  
 assolutamente d'accordo

D18 NEL COMPLESSO, INSIEME AGLI ASPETTI VALUTATI NELLE PRECEDENTI DOMANDE (voce e pause e melodia e esitazioni e parole e frasi)

ACCETTABILE.



Non sono affatto d'accordo sono  
assolutamente d'accordo

D19 NEL COMPLESSO, L'ESPRESSIONE  
DELL'INTERPRETE RISULTA CHIARA.



Non sono affatto d'accordo sono  
assolutamente d'accordo

D20 NEL COMPLESSO, LA COMPRENSIONE  
È SUFFICIENTE.



Non sono affatto d'accordo sono  
assolutamente d'accordo

D21 CONOSCE VOIÀ L'ARGOMENTO DEL TESTO  
INTERPRETATO.



Non sono affatto d'accordo sono  
assolutamente d'accordo

*Le chiediamo di esprimere liberamente le sue osservazioni in merito al questionario o di indicare aspetti che secondo lei non sono contemplati nelle domande.*

---

---

---

---

---

---

**BREVE COMMENTO ALLA BOZZA 4:**

Ho modificato le domande secondo le indicazioni date; pertanto, ho aggiunto le domande 9 e 10 relative alla cantilena. Dato che con il termine “cantilena” si può fare riferimento sia alla fastidiosa ripetizione di una monotonia della voce, che alla ripetizione di gruppi consonantici non standard (per esempio, la sostituzione di ---nch- con -ngh-, -nt- con -nd-), oppure alla presenza di cadenza o accento regionale<sup>1</sup>, ho pensato di inserire una domanda specifica per l'accento regionale, che può comprendere la realizzazione fonetica di gruppi consonantici secondo le norme del dialetto e non dell'italiano standard.

---

<sup>1</sup> Veneziano, C. (2001). *Manuale di dizione, voce e respirazione*. Lecce: Besa.

Questionnaire – draft 5

**BOZZA 5 DEL QUESTIONARIO**

Indichi la sua scelta (può indicarne anche più di una) barrando con una "x" l'opzione proposta (o le opzioni) proposte.

D1 LA VOCE DELL'INTERPRETE APPARTIENE A: maschio  
femmina

D2 LA VOCE DELL'INTERPRETE È:  
gradevole amabile dolce chiara noiosa lagnosa irritante  
monotona  
Altro (specificare): \_\_\_\_\_

D3 NEL COMPLESSO RISULTA:  
molto chiara chiara appena comprensibile indecifrabile

Esprima un valore in merito agli aspetti di seguito proposti: indichi con una "X" il valore che vuole assegnare su ciascuna scala

D4 LA VOCE DELL'INTERPRETE  
Intellegibile | | | | | | | | | | | | | | | | | | | | non  
intellegibile

D5 LA PRONUNCIA È  
Intellegibile | | | | | | | | | | | | | | | | | | | | non  
intellegibile

D6 LA VELOCITÀ DI ELOQUIO È  
| | | | | | | | | | | | | | | | | | | | alta bassa





D14 DAL PUNTO DI VISTA INFORMATIVO IL DISCORSO È

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
denso non denso

D15 L'INTERPRETE HA COSTRUITO FRASI GRAMMATICAMENTE

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
corrette non corrette

D16 L'INTERPRETE HA COSTRUITO FRASI

complesse  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
complesse non

D17 NEL COMPLESSO, L'INSIEME DEGLI ASPETTI VALUTATI NELLE  
PRECEDENTI DOMANDE (voce + pause + melodia + esitazioni + parole  
+ frasi) È

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
accettabile non  
accettabile

D18 NEL COMPLESSO, L'INTERPRETAZIONE RISULTA VOCALMENTE

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
chiara non chiara

D19 NEL COMPLESSO, L'INTERPRETAZIONE È

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
comprensibile non  
comprensibile

D20 L'ARGOMENTO DEL TESTO MI ERA

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

noto	ignoto
<b>D21 – DATI PERSONALI</b>	
Genere: <input type="checkbox"/> M <input type="checkbox"/> F	
Età: _____	
Professione / Studente/ssa di _____	
Lingue straniere / lingue studiate (A, B / A, C, C) _____ _____	
Anni di esperienza / anni di studio di interpretazione: _____	
<i>Le chiediamo di esprimere liberamente le sue osservazioni in merito al questionario o di indicare aspetti che secondo lei non sono contemplati nelle domande.</i>	
_____ _____ _____ _____ _____ _____ _____ _____ _____ _____	
<i>Grazie per la disponibilità!</i>	

Questionnaire – draft 6

**BOZZA 6 DEL QUESTIONARIO<sup>2</sup>**

Dopo aver visionato il video, risponda alle seguenti domande sulla valutazione degli aspetti dell'interpretazione simultanea ascoltata

Indichi la sua scelta (può indicarne anche più di una) barrando con una "x" l'opzione proposta (o le opzioni) proposte.

D1 LA VOCE DELL'INTERPRETE APPARTIENE A: uomo donna

D2 LA VOCE DELL'INTERPRETE È:

gradevole amabile dolce chiara noiosa lagnosa irritante monotona

Altro (specificare): \_\_\_\_\_

D3 NEL COMPLESSO RISULTA:

molto chiara chiara appena comprensibile indecifrabile

Esprima un valore in merito agli aspetti di seguito proposti: indichi con una "X" il valore che vuole assegnare su ciascuna scala

D4 LA VOCE DELL'INTERPRETE È

poco intellegibile										molto intellegibile					

D5 L'INTERPRETE

Non scandisce per nulla le parole							scandisce molto bene le parole							

<sup>2</sup> Reviewed by Prof. Delli Zotti

D6 L'INTERPRETE PARLA

molto lentamente  molto velocemente

---

D7 L'INTONAZIONE RISULTA

Per nulla gradevole  assolutamente gradevole

---

D8 LA PARLATA DELL'INTERPRETE RISULTA

Per nulla cantilenante  molto cantilenante

---

D9 LA PARLATA DELL'INTERPRETE È CARATTERIZZATA DA UN ACCENTO REGIONALE

Per nulla marcato  molto marcato

---

D10 LA PARLATA DELL'INTERPRETE RISULTA

Per nulla scorrevole  molto scorrevole

---

D11 LA PARLATA DELL' INTERPRETE PRESENTA

nessuna esitazione  molte esitazioni

---

D11 LA PARLATA DELL' INTERPRETE PRESENTA

Nessuna pausa silenziosa  molte pause silenziose

---

D12 LA PARLATA DELL' INTERPRETE PRESENTA

nessuna interruzione  molte interruzioni



D20 L'ARGOMENTO DEL TESTO MI ERA

Assolutamente sconosciuto                      molto noto

D21 – DATI PERSONALI

Genere:  Uomo  Donna

Età: \_\_\_\_\_

Titolo di studio \_\_\_\_\_

Professione / Studente(ssa) di  
\_\_\_\_\_

Lingue straniere / lingue studiate (A, B / A, C, C)  
\_\_\_\_\_  
\_\_\_\_\_

Anni di esperienza / anni di studio  
di interpretazione: \_\_\_\_\_

*Le chiediamo di esprimere liberamente le sue osservazioni in merito al questionario o di indicare aspetti che secondo lei non sono contemplati nelle domande.*

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

*Grazie per la disponibilità!*

NOTE A MARGINE

Finora il presente questionario è stato revisionato da:

- il prof. Corrado Veneziano (collabora con l'Università di Roma La Sapienza);
- il prof. Delli Zotti (Università di Trieste – Dipartimento di Scienze Politiche e Sociali);

## Questionnaire – draft 7

**BOZZA 7 DEL QUESTIONARIO<sup>3</sup>**

Dopo aver visionato il video, risponda alle seguenti domande sulla valutazione degli aspetti dell'interpretazione simultanea ascoltata

Indichi la sua scelta (può indicarne anche più di una) barrando con una "x" l'opzione proposta (o le opzioni) proposte.

D1 LA VOCE DELL'INTERPRETE APPARTIENE A:      uomo      donna

D2 LA VOCE DELL'INTERPRETE È:

gradevole   amabile   dolce   chiara   noiosa   lagnosa   irritante  
monotona

Altro (specificare): \_\_\_\_\_

D3 NEL COMPLESSO RISULTA:

molto chiara   chiara   appena comprensibile   indecifrabile

Esprima un valore in merito agli aspetti di seguito proposti: indichi con una "X" il valore che vuole assegnare su ciascuna scala

D4 LA VOCE DELL'INTERPRETE È

poco intellegibile   | | | | | | | |   molto intellegibile

D5 L'INTERPRETE

Non scandisce per nulla   | | | | | | | |   scandisce molto bene  
le parole   le parole

<sup>3</sup> Reviewed by Olalla García Becerra

D6	L'INTERPRETE PARLA	_ _ _ _ _ _ _	molto lentamente	molto velocemente
D7	L'INTONAZIONE RISULTA	_ _ _ _ _ _ _	Per nulla gradevole	assolutamente gradevole
D8	LA PARLATA DELL'INTERPRETE RISULTA CANTILENANTE (ripetizione fastidiosa della stessa melodia vocale)	_ _ _ _ _ _ _	per nulla	molto
D9	LA PARLATA DELL'INTERPRETE È CARATTERIZZATA DA UN ACCENTO REGIONALE	_ _ _ _ _ _ _	Per nulla marcato	molto marcato
D10	LA PARLATA DELL'INTERPRETE RISULTA	_ _ _ _ _ _ _	Per nulla scorrevole	molto scorrevole
D11	LA PARLATA DELL' INTERPRETE PRESENTA	_ _ _ _ _ _ _	nessuna esitazione	molte esitazioni
D11	LA PARLATA DELL' INTERPRETE PRESENTA	_ _ _ _ _ _ _	Nessuna pausa silenziosa	molte pause silenziose



D12	LA PARLATA DELL' INTERPRETE PRESENTA	_ _ _ _ _ _ _	nessuna interruzione	molte interruzioni
D13	L'INTERPRETE HA UTILIZZATO UN LESSICO	_ _ _ _ _ _ _	Per nulla pertinente al discorso discorso	molto pertinente al discorso
D14	IL TESTO INTERPRETATO PRESENTA (ALL'ASCOLTATORE)	_ _ _ _ _ _ _	nessuna informazione nuova nuove	molte informazioni nuove
D15	L'INTERPRETE HA COSTRUITO FRASI GRAMMATICAMENTE	_ _ _ _ _ _ _	poco corrette	molto corrette
D16	L'INTERPRETE HA COSTRUITO FRASI	_ _ _ _ _ _ _	Poco complesse	molto complesse
D17	NEL COMPLESSO, L'INSIEME DEGLI ASPETTI VALUTATI NELLE PRECEDENTI DOMANDE (voce + pause + melodia + esitazioni + parole + frasi) RISULTA	_ _ _ _ _ _ _	Per nulla accettabile	molto accettabile
D18	NEL COMPLESSO, L'INTERPRETAZIONE RISULTA VOCALMENTE	_ _ _ _ _ _ _	Poco chiara	molto chiara

D19 NEL COMPLESSO, L'INTERPRETAZIONE È

\_\_\_\_\_

poco comprensibile \_\_\_\_\_ molto comprensibile

D20 L'ARGOMENTO DEL TESTO MI ERA

\_\_\_\_\_

Assolutamente sconosciuto \_\_\_\_\_ molto noto

D21 – DATI PERSONALI

Genere:  Uomo  Donna

Età: \_\_\_\_\_

Titolo di studio \_\_\_\_\_

Professione / Studente(ssa) di

\_\_\_\_\_

Lingue straniere / lingue studiate (A, B / A, C, C)

\_\_\_\_\_

\_\_\_\_\_

Anni di esperienza / anni di studio  
di interpretazione: \_\_\_\_\_

*Le chiediamo di esprimere liberamente le sue osservazioni in merito al  
questionario o di indicare aspetti che secondo lei non sono contemplati nelle  
domande.*

\_\_\_\_\_

*Grazie per la disponibilità!*

**APPENDIX 6:**

**Questionnaire – Last version (video sequence: 1-2-3)**







V1.D22 - QUAL È LA CRITICA FONDAMENTALE DI OBAMA ALLE PRECEDENTI POLITICHE IN QUESTIONE (I PARTE)?

- indebitamento e deficit eccessivi
- Posizione di McCain al riguardo
- Relazioni internazionali pericolose
- Preferisco non rispondere

V1.D23 - QUAL È LA PROPOSTA FONDAMENTALE DI OBAMA SULLE POLITICHE IN QUESTIONE (II PARTE)?

- Modificare le spese sull'energia
- Taglio delle spese e riforme sugli investimenti
- Rivedere le spese sulla sanità e sull'istruzione
- Preferisco non rispondere

V1.D24 – EVENTUALI COMMENTI SU ASPETTI DELL'INTERPRETAZIONE MOSTRATA NON CONTEMPLATI NELLE DOMANDE:























**DATI SOCIO-DEMOGRAFICI**

- UOMO  
 DONNA

Età: \_\_\_\_\_

- Studente/ssa di Interpretazione di Conferenza

Anno di corso:  I anno     Il anno

- Studente/ssa di Traduzione Specialistica

Anno di corso:  I anno     Il anno

- Studente/ssa di Comunicazione Interlinguistica applicata

Interprete

Musicista

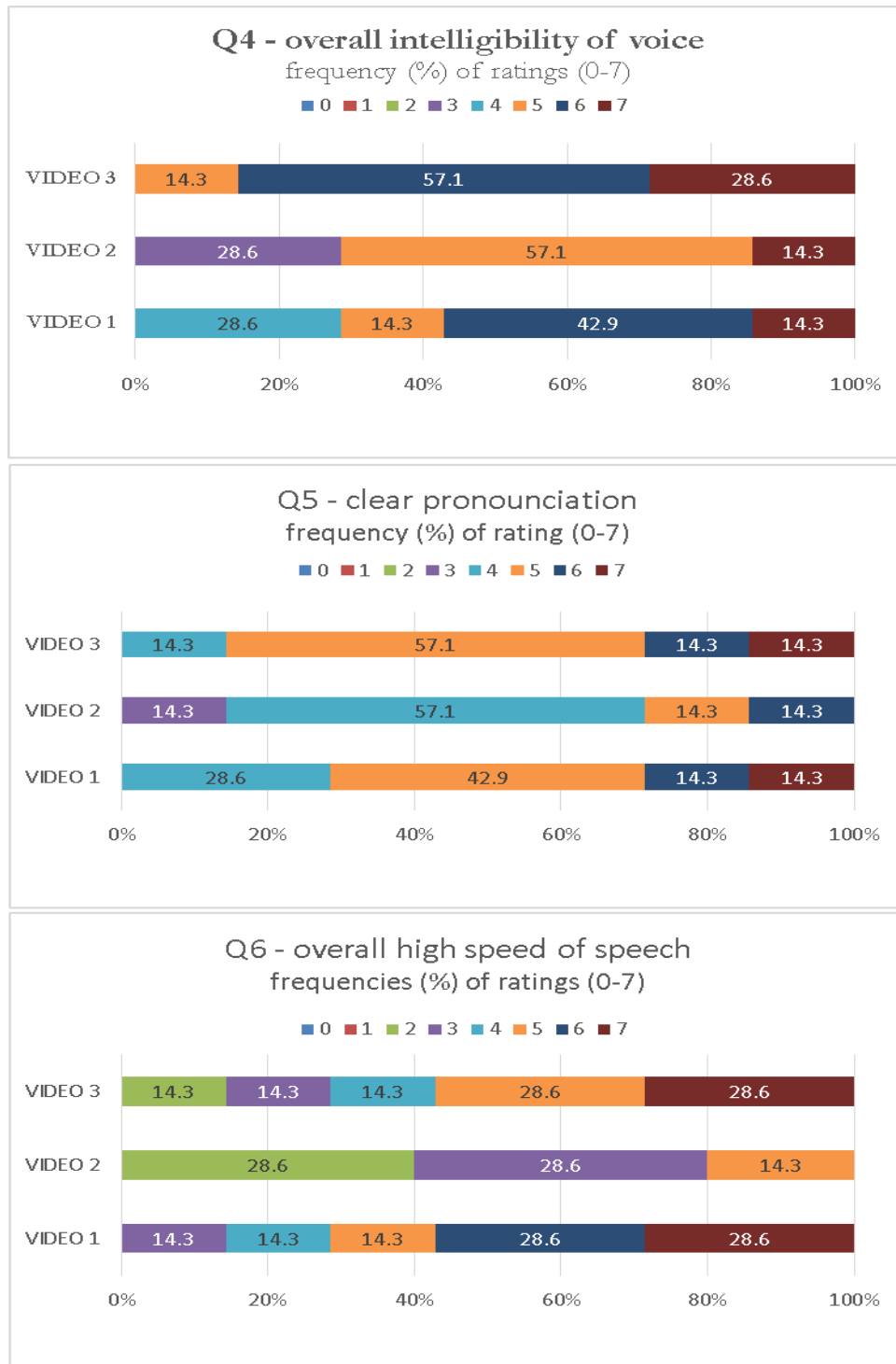
Altro

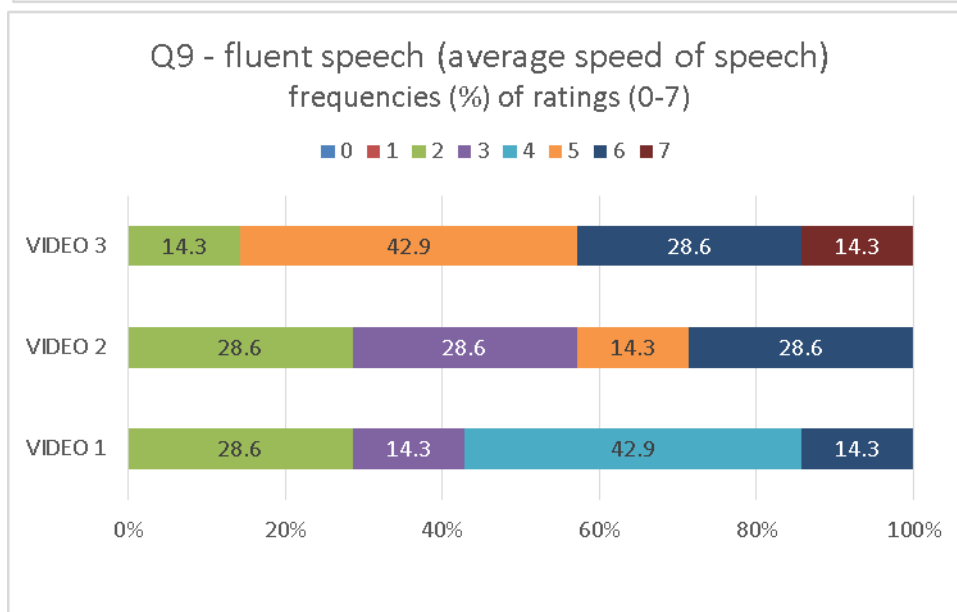
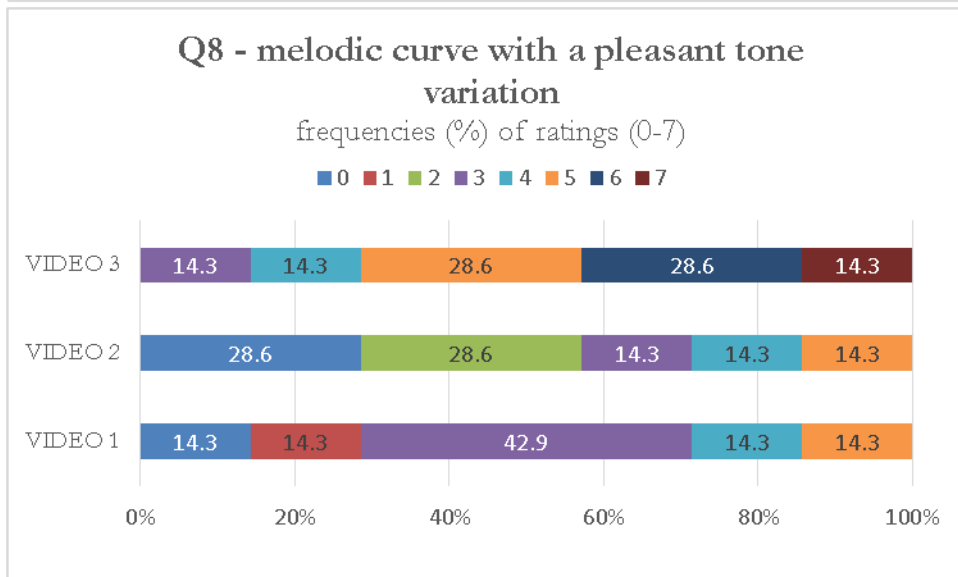
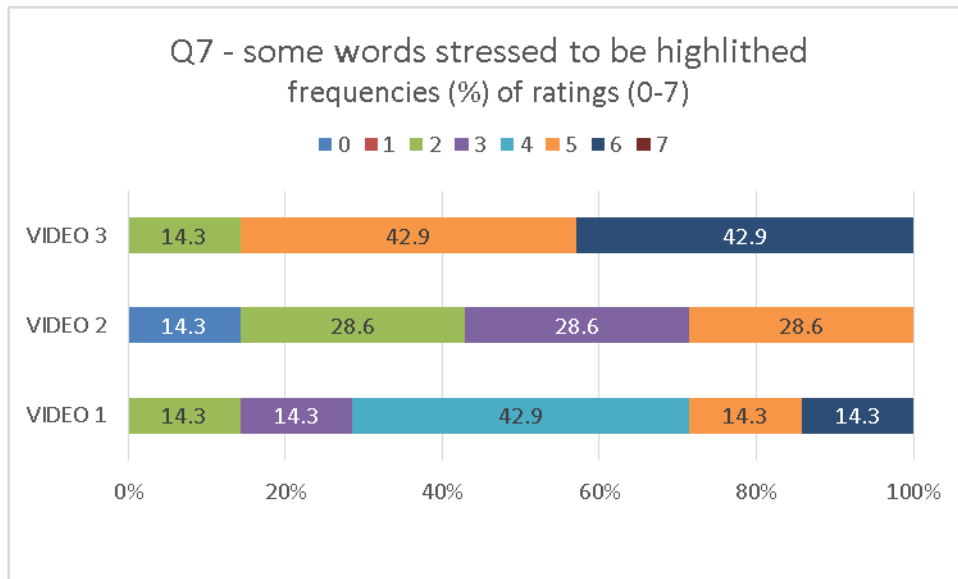
*Le chiediamo di esprimere liberamente le sue osservazioni in merito al questionario o di indicare aspetti che secondo lei non sono contemplati nelle domande.*

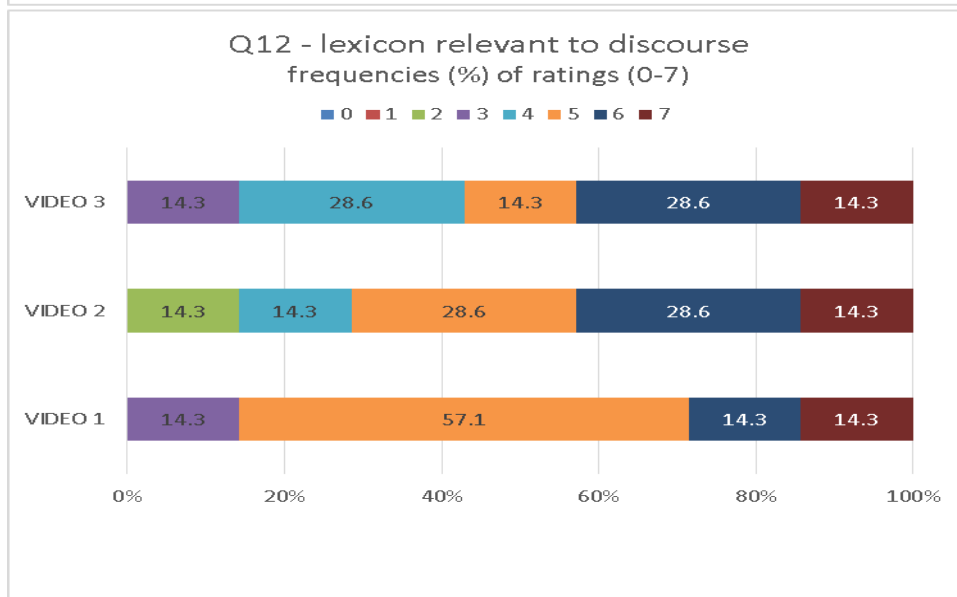
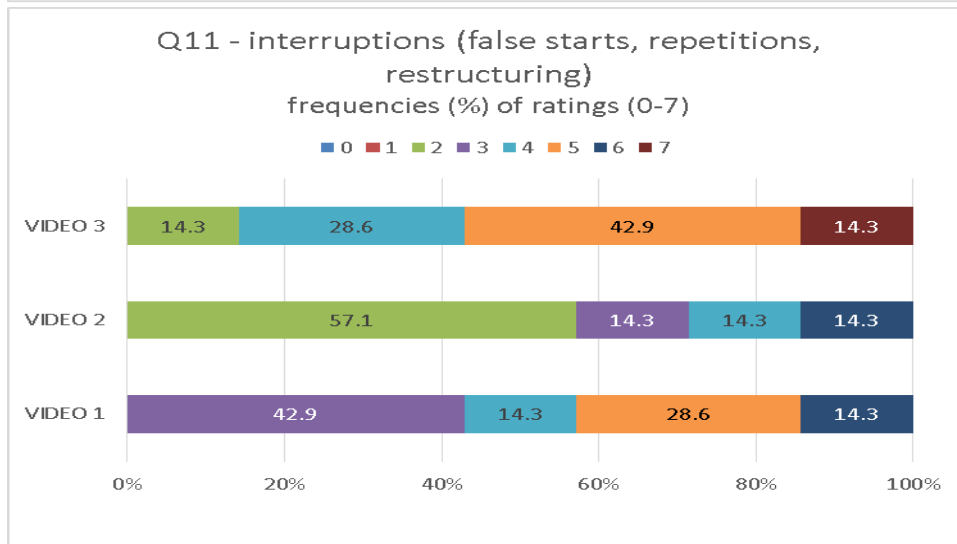
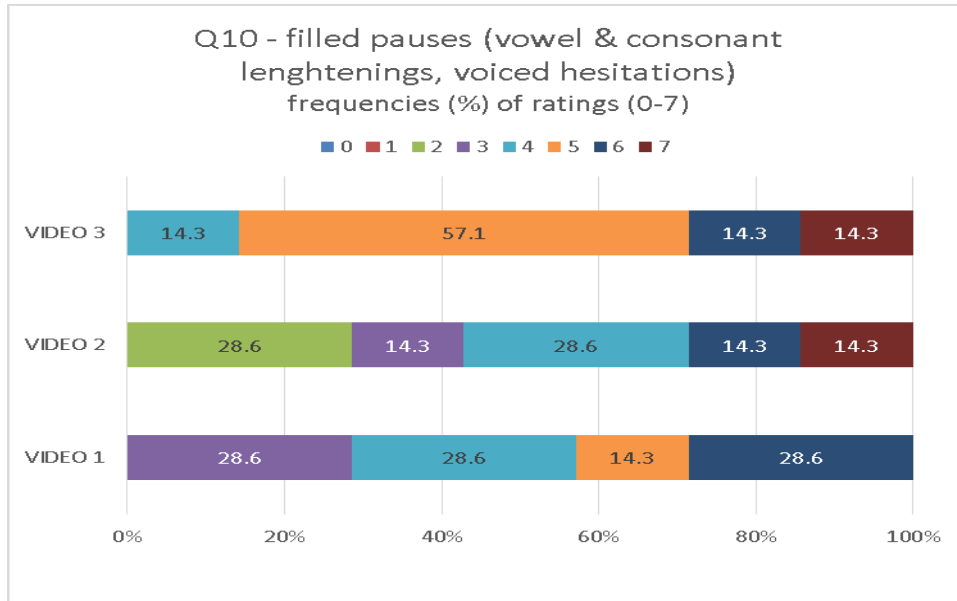
*Grazie per la collaborazione!*

## APPENDIX 7:

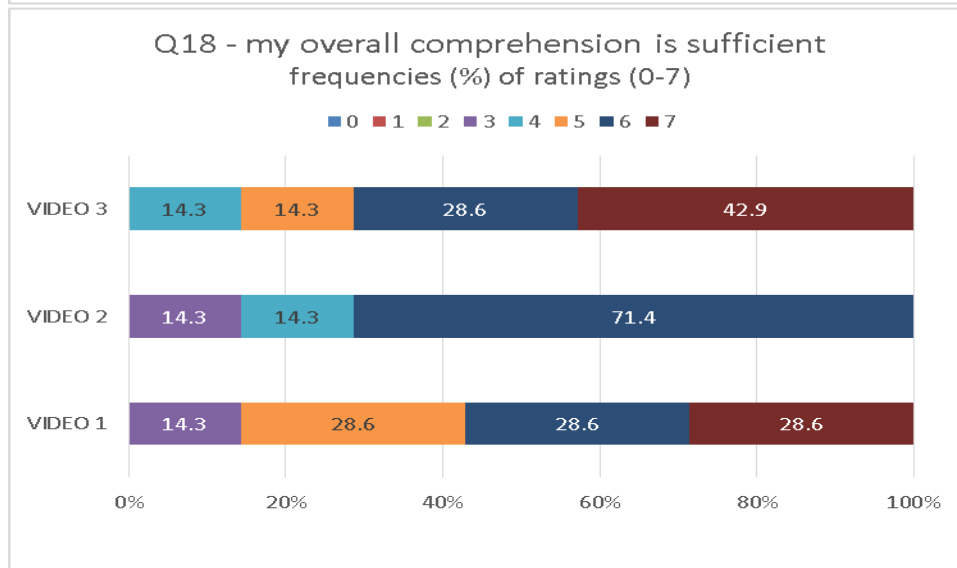
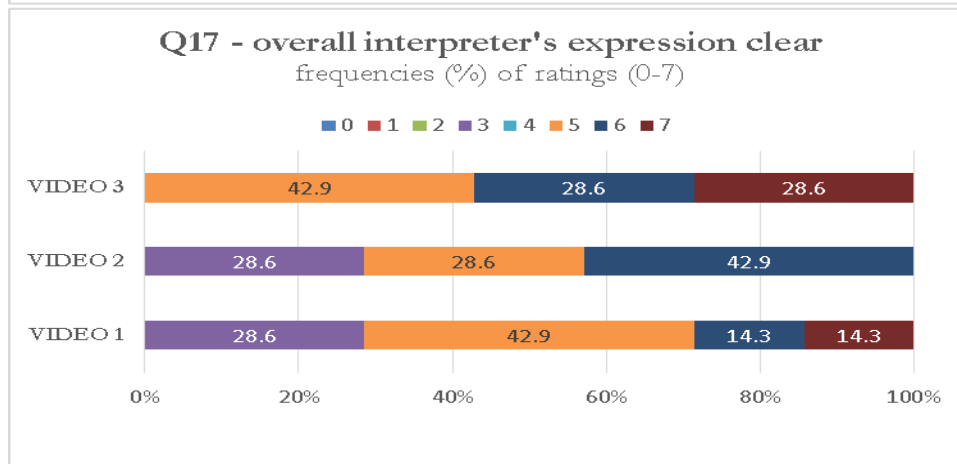
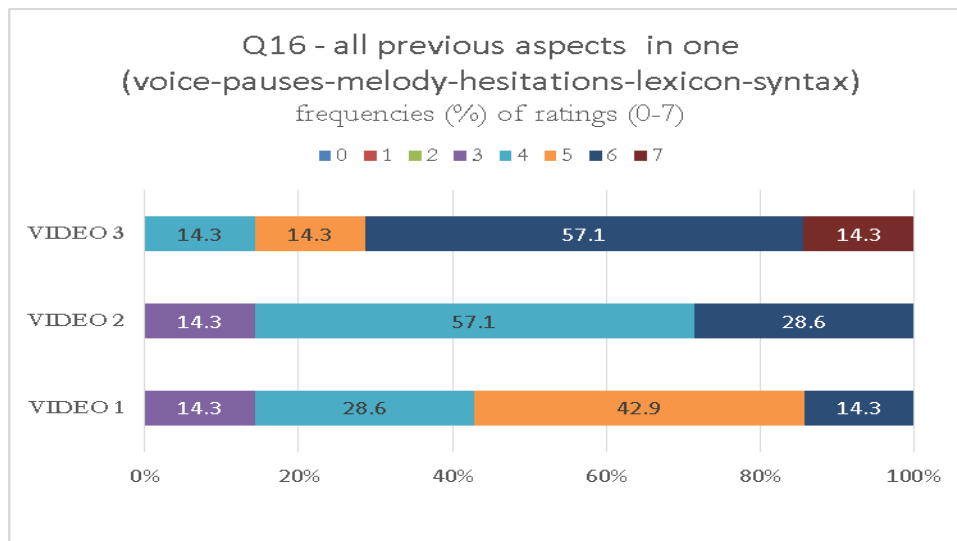
### Results from pre-test (questionnaire - draft 3)

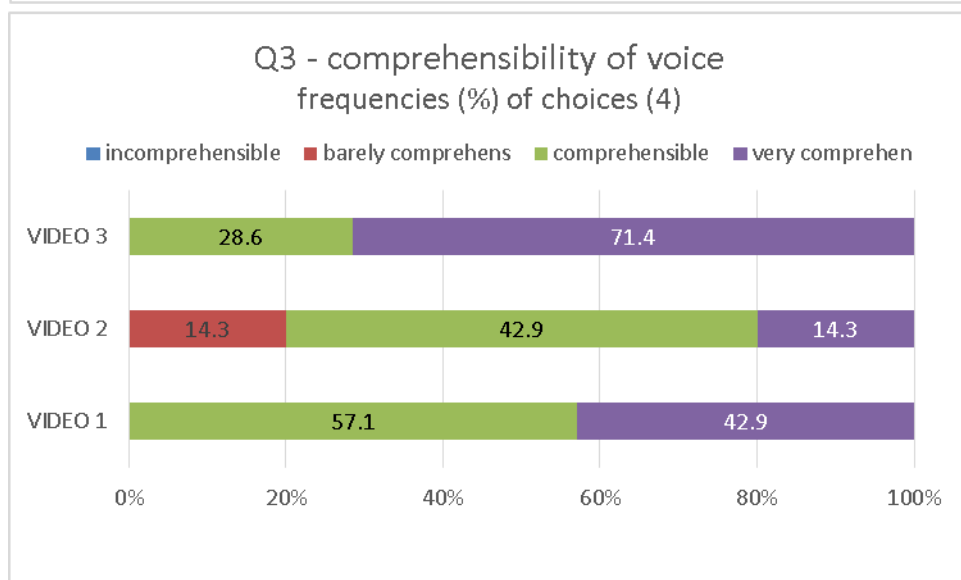
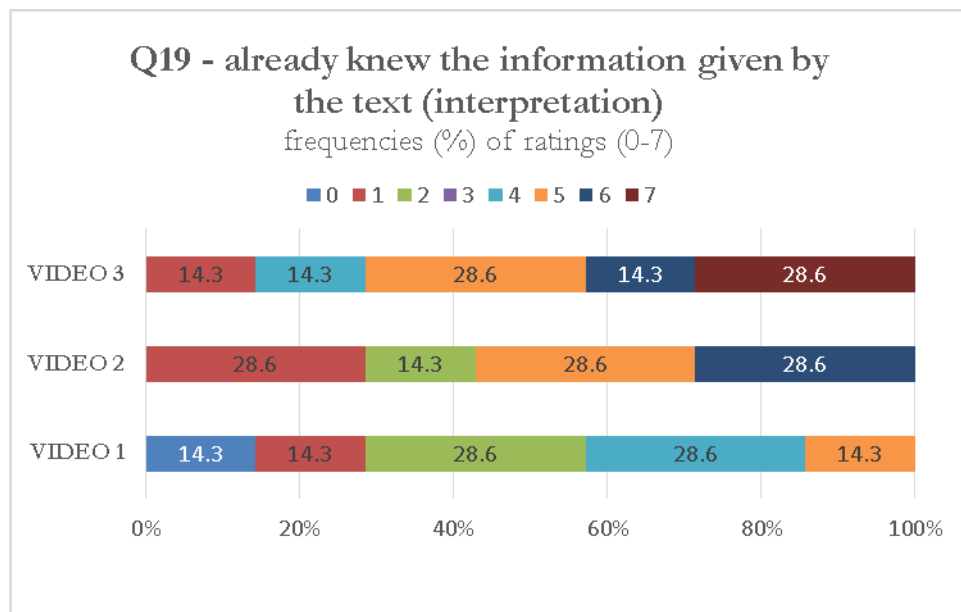












## APPENDIX 8:

### Grid for the selection of video excerpts

ESTRATTO	INTERPRET. + N° interpreti	LESSICO	SINTASSI	INTONAZIONE	DIS
OBAMA- McCain 3 RAINEWS					
Clip 4 (1° int.)	1 introd. Moder.	Denso+	complessa	Compatta	Poch
	2 McCain	d-	Semplice, ipotassi	Buona	Id
	3 McCain ?	d-	idem	buona	id
Clip 5	1 Mc	denso	buona	ridotta	Paus
	2 Obama	?	buona	buona	poch
Clip 6	1 McCain 2 Obama	Facile difficile	Semplice difficile	Buona buona	Asse Pres part



<b>ESTRATTO</b>	<b>INTERPRET. + N° interpreti</b>	<b>LESSICO</b>	<b>SINTASSI</b>	<b>INTONAZIONE</b>	<b>DIS</b>
					ripet
Clip 7	1 McCain 2 Obama 3 Moderatore	Facile facile	Facile Facile/diff	Media buona	Poch false Pres
Clip 8	Moderatore Obama	Diff/fac Diff	Facile complessa	Facile buona	Poch Pres part tensi
Clip 9	McCain Obama (1 min)	medio	complesso	buona	Poch
Clip 10	Obama McCain	Facile	facile	buona	poch
Clip 11	McCain Obama Moderatore	Medio	media	buona	
Clip 12	McCain Obama	facile	difficile	faticosa	
Clip 13	Obama McCain Obama	facile	difficile	buona	
Clip 14	Obama McCain Obama	facile	facile	buona	
Clip 15	Obama McCain Obama	facile	media		
Clip 16	Obama	difficile	media		

<b>ESTRATTO</b>	<b>INTERPRET. + N° interpreti</b>	<b>LESSICO</b>	<b>SINTASSI</b>	<b>INTONAZIONE</b>	<b>DIS</b>
	McCain Moderatore Obama				
Clip 18	McCain Obama McCain	Facile	Facile poi difficile		
Clip 19	Moderatore McCain	Difficile (energia)	difficile		
Clip 20	Obama (energia + Nafta) McCain	Difficile (denso)	medio		
Clip 21	McCain Obama McCain	medio	facile		
Clip 22	McCain Moderatore Obama McCain	medio	facile		
Clip 23	McCain Obama	Denso difficile	media		
Clip 24	Obama	Denso difficile	Media		
Clip 25	Obama Moderatore McCain	medio	facile		
Clip 26	McCain Obama Moderatore McCain	Medio-facile	facile		

<b>ESTRATTO</b>	<b>INTERPRET. + N° interpreti</b>	<b>LESSICO</b>	<b>SINTASSI</b>	<b>INTONAZIONE</b>	<b>DIS</b>
Clip 27 Aborto	Obama McCain Obama	Difficile (tecnico – omissione)	facile		
Clip 28 aborto	Obama Moderatore McCain Moderatore Obama	Medio medio	Difficile facile		
Clip 29 Istruzione	Obama McCain	Medio medio	Facile facile		
Clip 30 Istruzione	McCain Obama	difficile	Facile Difficile		

**APPENDIX 9:****WinPitch measurements of silent pauses and audible breaths  
(video excerpts for pilot survey).**

VIDEO 1							
Values extracted from WinPitch							Graphic syllables [number]
Rank	Speaker	Text Phonic chains [PAUSE] / (BREATH)	Start [s]	End [s]	Channel	Duration [s]	
1	L 1	RESPIRO	0.02	0.818	M	0.798	
2	L 1	è vero che nessuno è completamente innocente qui	0.818	3.129	M	2.311	18
3	L 1	RESPIRO	3.129	3.681	M	0.552	
4	L 1	negli ultimi otto anni abbiamo avuto il più g- il più grosso aumento di spesa:	3.681	7.998	M	4.317	28
5	L 1	RESPIRO	7.998	8.468	M	0.47	
6	L 1	di deficit e di debito nazionale dena nostra storia-e-il senatore McCain ha votato	8.468	12.968	M	4.5	31
7	L 1	RESPIRO	12.98	13.594	M	0.614	
8	L 1	per COLPO DI GLOTTIDE quattro su cinque di questi: bilanci di:	13.594	16.805	M	3.211	13
9	L 1	eh	16.815	17.061	M	0.246	1
10	L 1	[PAUSA]	17.061	17.327	M	0.266	
11	L 1	eh	17.329	17.461	M	0.132	1
12	L 1	(RESPIRO)	17.478	18.041	M	0.563	
13	L 1	eh	18.058	18.273	M	0.215	1
14	L 1	leggi finanziarie di Bush quindi io spenderò dei soldi sulle questioni chiare su cui dobbiamo lavorare	18.308	23.259	M	4.951	33
15	L 1	(RESPIRO)	23.259	23.924	M	1.979	
16	L 1	[PAUSA]	23.924	25.238	M	1.314	
17	L 1	eh	25.238	25.519	M	0.281	1
18	L 1	[PAUSA]	25.529	25.725	M	0.196	
19	L 1	lei ha visto un aumento dei premi sull'assistenza sanitaria dobbiamo riformare: l'assistenza sanitaria dobbiamo affrontare l'energia perché non possiamo continuare a prendere in prestito d-	25.745	33.666	M	7.921	64
20	L 1	(RESPIRO)	33.667	33.855	M	0.188	
21	L 1	dai cinesi dando: i soldi all'Arabia Saudita	33.868	36.713	M	2.845	16
22	L 1	(RESPIRO)	36.713	37.893	M	1.18	
23	L 1	[PAUSA]	37.893	38.146	M	0.253	
24	L 1	eh	38.146	38.968	M	0.822	1
25	L 1	ci vuole un piano energetico diverso dobbiamo investire /sulle/, su e- / sulla scuola:	38.978	42.645	M	3.667	17
26	L 1	(RESPIRO)	42.645	43.056	M	0.411	
27	L 1	dobbiamo fare degli investimenti ma dobbiamo fare anche dei tagli alla spesa e la mia proposta	43.065	47.006	M	3.941	33
28	L 1	(RESPIRO)	47.006	47.734	M	0.728	
29	L 1	(i)l senatore McCai(n) disse che queste: le mie pro(po)ste sono tutte nuove spese ma in realtà io sto tagliando più di quanto spendo	47.742	53.391	M	5.649	39
30	L 1	(RESPIRO)	53.404	53.856	M	0.452	
31	L 1	quindi ci sarà un taglio sulla spesa netta allora la chiave è se noi abbiamo delle priorità che funzionano per voi	53.869	59.758	M	5.889	38
32	L 1	(RESPIRO)	59.766	60.16	M	0.394	
33	L 1	rispetto	60.16	60.871	M	0.711	1
34	L 1	(RESPIRO)	60.871	60.283	M	0.412	
35	L 1	a coloro che hanno	61.283	62.166	M	0.883	7
36	L 1	[PAUSA]	62.166	62.544	M	0.378	
37	L 1	dettato le politiche di Washington negli ultimi otto anni cioè soprattutto le lobby e gli interessi speciali noi dobbiamo porre fine a tutto questo senatore McCain	62.544	68.93	M	6.386	57

VIDEO 2 - ORIGINAL							
Values extracted from WinPitch							
Rank	Speaker	Text	Start [s]	End [s]	Channel	Duration [s]	Graphic syllables [number]
		Phonic chains					
		[PAUSE] / (BREATH)					
1	PC/Br/	voglio dire un ultima cosa	2.844	4.141	M	1.297	10
2	PC/Br/	(RESPIRO)	4.141	5.278	M	1.136	
3	PC/Br/	perché: il senatore McCain ha parlato:	5.278	7.167	M	1.889	13
4	PC/Br/	[PAUSA]	7.167	7.668	M	0.501	
5	PC/Br/	del NAFTA e delle questioni commerciali una questione che voglio sottolineare	7.668	10.876	M	3.207	26
6	PC/Br/	(RESPIRO)	10.876	11.126	M	0.25	
7	PC/Br/	io credo nel libero scambio	11.148	12.622	M	1.474	9
8	PC/Br/	(RESPIRO)	12.622	12.822	M	0.199	
9	PC/Br/	ma credo anche	12.822	13.638	M	0.816	5
10	PC/Br/	(RESPIRO)	13.638	13.867	M	0.229	
11	PC/Br/	[PAUSA]	13.881	15.24	M	1.359	
12	PC/Br/	che per troppo tempo sicuramente nel corso dell'amministrazione Bush	15.24	18.917	M	3.677	22
13	PC/Br/	[PAUSA]	18.927	19.07	M	0.143	
14	PC/Br/	con l'appoggio del senatore McCain	19.084	20.715	M	1.631	11
15	PC/Br/	(RESPIRO)	20.729	21.344	M	0.615	
16	PC/Br/	l'a- l'atteggiamento è stato che qualsiasi:	21.344	23.606	M	2.262	8
17	PC/Br/	(RESPIRO)	23.606	23.707	M	0.1	
18	PC/Br/	accordo commerciale fosse giusto	23.721	25.695	M	1.974	11
19	PC/Br/	[PAUSA]	25.695	28.256	M	2.561	
20	PC/Br/	eh!- nel NAFTA non c'erano degli accordi commerciali	28.256	31.192	M	2.936	17
21	PC/Br/	[PAUSA]	31.192	32.022	M	0.829	
22	PC/Br/	e e di lavoro e: secondo me do(ve)vamo introdurli	32.022	34.869	M	2.847	18
23	PC/Br/	(RESPIRO)	34.878	35.522	M	0.643	
24	PC/Br/	così come avremmo dovuto applicare delle:	35.522	37.482	M	1.96	16
25	PC/Br/	(RESPIRO)	37.482	37.696	M	0.213	
26	PC/Br/	[PAUSA]	37.696	38.14	M	0.444	
27	PC/Br/	regole contro la Cina:	38.14	39.599	M	1.458	8
28	PC/Br/	(RESPIRO)	39.599	40.386	M	0.787	
29	PC/Br/	[PAUSA]	40.386	41.016	M	0.629	
30	PC/Br/	per rendere	41.027	41.556	M	0.528	2
31	PC/Br/	(RESPIRO)	41.556	41.957	M	0.401	
32	PC/Br/	[PAUSA]	41.971	42.672	M	0.701	
33	PC/Br/	per rendere le loro	42.672	43.688	M	1.016	7
34	PC/Br/	[PAUSA]	43.688	43.788	M	0.099	
35	PC/Br/	eh: importazioni	43.798	45.169	M	1.37	5
36	PC/Br/	(RESPIRO)	45.169	45.597	M	0.428	
37	PC/Br/	hm più economiche	45.597	46.705	M	1.107	7
38	PC/Br/	(RESPIRO)	46.705	47.672	M	0.966	
39	PC/Br/	[PAUSA]	47.672	48.399	M	0.727	
40	PC/Br/	adesso abbiamo	48.399	49.19	M	0.79	6
41	PC/Br/	[PAUSA]	49.19	49.669	M	0.478	
42	PC/Br/	adesso loro stanno inviando centinaia di migliaia di: macchine da Corea del Sud negli Stati Uniti da Corea del Sud questo va benissimo hm	49.682	55.919	M	6.236	47
43	PC/Br/	(RESPIRO)	55.925	56.102	M	0.176	
44	PC/Br/	ma noi invece riusciamo soltanto a mandarne quattromila cinquemila in Corea del Sud questo non è libero scambio noi dobbiamo avere un presidente	56.117	62.32	M	6.203	49
45	PC/Br/	(RESPIRO)	62.335	63.144	M	0.808	
46	PC/Br/	[PAUSA]	63.144	64.29	M	1.146	
47	PC/Br/	che lavori	64.295	65.05	M	0.754	4

VIDEO 2 - IMITATION							
Values extracted from WinPitch							Graphic syllables [number]
Rank	Speaker	Text Phonic chains Phonic chains [PAUSE] / (BREATH)	Start [s]	End [s]	Channel	Duration [s]	
1	PC/Br/	Voglio dire un ultima cosa:	0.498	2.089	M	1.591	10
2	PC/Br/	[PAUSA]	2.089	2.607	M	0.518	
3	PC/Br/	perché: il senatore McCain ha parlato del NAFTA e delle questioni commerciali	2.607	6.671	M	4.064	16
4	PC/Br/	(RESPIRO)	6.671	7.035	M	0.363	
5	PC/Br/	la questione che voglio sottolineare:	7.035	8.434	M	1.398	12
6	PC/Br/	(RESPIRO)	8.434	8.703	M	0.269	
7	PC/Br/	io credo: nel libero scambio ma:	8.719	11.532	M	2.813	10
8	PC/Br/	[PAUSA - COLPO DI GLOTTIDE]	11.532	11.735	M	0.202	
9	PC/Br/	credo . anche . che:	11.735	13.438	M	1.703	5
10	PC/Br/	[PAUSA]	13.438	13.695	M	0.256	
11	PC/Br/	per troppo tempo	13.695	15.01	M	1.314	5
12	PC/Br/	(RESPIRO)	15.018	15.467	M	0.448	
13	PC/Br/	sicuramente nel corso dell'amministrazione Bush con l'appoggio del senatore McCain l'atteggiamento è stato che qualsiasi accordo	15.478	21.758	M	6.279	42
14	PC/Br/	(RESPIRO)	21.758	22.042	M	0.284	
15	PC/Br/	commerciale fosse giusto (COLPO DI GLOTTIDE) il:	22.042	24.483	M	2.44	8
16	PC/Br/	(RESPIRO)	24.483	24.838	M	0.355	
17	PC/Br/	nel NAFTA non c'erano accordi commerciali e . e . e di lavoro e secondo me:	24.838	28.699	M	3.861	26
18	PC/Br/	(RESPIRO)	28.699	29.383	M	0.683	
19	PC/Br/	eh:	29.383	29.775	M	0.391	1
20	PC/Br/	[PAUSA]	29.781	30.3	M	0.519	
21	PC/Br/	dovevamo introdurli così come avremmo dovuto applicare delle regole	30.3	34.443	M	4.142	27
22	PC/Br/	(RESPIRO)	34.451	34.57	M	0.118	
23	PC/Br/	contro la Cina	34.579	35.499	M	0.92	5
24	PC/Br/	(RESPIRO)	35.506	35.916	M	0.409	
25	PC/Br/	per rendere:	35.916	37.017	M	1.101	4
26	PC/Br/	[PAUSA]	37.017	37.919	M	0.901	
27	PC/Br/	per rendere le loro:	37.919	39.394	M	1.475	7
28	PC/Br/	[PAUSA]	39.394	39.796	M	0.402	
29	PC/Br/	importazioni più economiche	39.796	41.362	M	1.566	11
30	PC/Br/	[PAUSA]	41.362	43.639	M	2.277	
31	PC/Br/	eh: adesso: abbiamo	43.639	45.246	M	1.606	7
32	PC/Br/	(RESPIRO)	45.246	45.51	M	0.263	
33	PC/Br/	adesso loro stanno:	45.515	46.398	M	0.883	7
34	PC/Br/	(RESPIRO)	46.398	46.626	M	0.227	
35	PC/Br/	[PAUSA]	46.626	47.09	M	0.464	
36	PC/Br/	inviando centinaia di migliaia di macchine: da Corea del Sud negli Stati Uniti	47.09	50.769	M	3.678	21
37	PC/Br/	(RESPIRO)	50.769	51.091	M	0.322	
38	PC/Br/	la Corea: del Sud questo va benissimo ma noi invece	51.091	53.877	M	2.786	17
39	PC/Br/	(RESPIRO)	53.884	54.112	M	0.228	
40	PC/Br/	riusciamo soltanto a mandarne quattromila cinquemila in Corea del Sud questo	54.112	57.536	M	3.423	25
41	PC/Br/	(RESPIRO)	57.544	57.781	M	0.237	
42	PC/Br/	non è libero scambio noi dobbiamo avere un presidente che lavori [ per gli Stati Uniti]	57.781	60.941	M	3.16	23

VIDEO 3							
Values extracted from WinPitch							Graphic syllables [number]
Rank	Speaker	Text	Start [s]	End [s]	Channel	Duration [s]	
		Phonic chains					
		[PAUSE] / (BREATH)					
1	L 1	Ma qual è il vantaggio	1.385	2.71	M	1.325	7
2	L 1	[PAUSA]	2.71	2.89	M	0.18	
3	L 1	in una: zona:	2.91	4.035	M	1.125	7
4	L 1	(RESPIRO)	4.035	4.235	M	0.2	
5	L 1	povera di mandare i lo- i figli	4.255	6.624	M	2.369	12
6	L 1	[PAUSA]	6.624	6.825	M	0.201	
7	L 1	in scuole che stanno fallendo	6.825	8.632	M	1.806	9
8	L 1	(RESPIRO)	8.632	9.595	M	0.963	
9	L 1	la concorrenza tra le scuole quindi	9.595	11.242	M	1.647	11
10	L 1	[PAUSA]	11.25	11.37	M	0.119	
11	L 1	è uno d- degli elementi chiave	11.391	13.159	M	1.768	11
12	L 1	[PAUSA]	13.159	13.179	M	0.2	
13	L 1	si è già visto . il successo . a New Orleans a: New York	13.179	16.931	M	3.752	16
14	L 1	(RESPIRO)	16.937	17.79	M	0.852	
15	L 1	in cui abbiamo	17.803	18.564	M	0.76	5
16	L 1	[PAUSA]	18.564	18.944	M	0.379	
17	L 1	delle scuole con ottimi insegnanti	18.944	20.978	M	2.034	12
18	L 1	(RESPIRO)	20.992	21.075	M	0.082	
19	L 1	che vengono: eh:m premiati	21.086	23.145	M	2.059	8
20	L 1	(RESPIRO)	23.145	24.009	M	0.864	
21	L 1	e troviamo	24.018	24.669	M	0.65	4
22	L 1	[PAUSA]	24.669	24.893	M	0.224	
23	L 1	ai agli insegnanti meno bravi	24.893	26.999	M	2.105	10
24	L 1	[PAUSA]	27.011	27.248	M	0.237	
25	L 1	altri posti di lavoro	27.259	28.407	M	1.147	8
26	L 1	(RESPIRO)	28.419	29.01	M	0.591	
27	L 1	dobbiamo dare la stessa scelta	29.01	30.705	M	1.694	10
28	L 1	[PAUSA]	30.705	30.894	M	0.189	
29	L 1	eh:	30.894	31.308	M	0.414	1
30	L 1	[PAUSA]	31.308	31.77	M	0.461	
31	L 1	ai genitori di tutta l'America eh le stesse scelte che hanno avuto il senatore Obama e sua moglie e quella che ho avuto io e mia moglie	31.77	37.519	M	5.748	49
32	L 1	(RESPIRO)	37.528	37.954	M	0.426	
33	L 1	di mandare i nostri figli alle scuole migliori	37.954	39.954	M	2	16
34	L 1	(RESPIRO)	39.962	40.861	M	0.898	
35	L 1	[PAUSA]	40.861	41.393	M	0.532	
36	L 1	bisogna garantire la concorrenza tra le scuole bisogna:	41.393	43.937	M	2.543	19
37	L 1	(RESPIRO)	43.946	44.75	M	0.804	
38	L 1	garantire quel tipo di concorrenza che ha migliorato sia le scuole pubbliche che quelle private	44.75	49.388	M	4.637	31
39	L 1	(RESPIRO)	49.394	50.009	M	0.615	
40	L 1	ora	50.009	50.293	M	0.283	1
41	L 1	[PAUSA]	50.293	51.547	M	1.253	
42	L 1	(RESPIRO)	51.557	52.74	M	1.182	
43	L 1	eh: gli investimenti il finanziamento non poss- ehm essere l'unica soluzione	52.74	57.472	M	4.731	26

## **APPENDIX 10:**

### **Frequencies of ratings (pilot survey)**





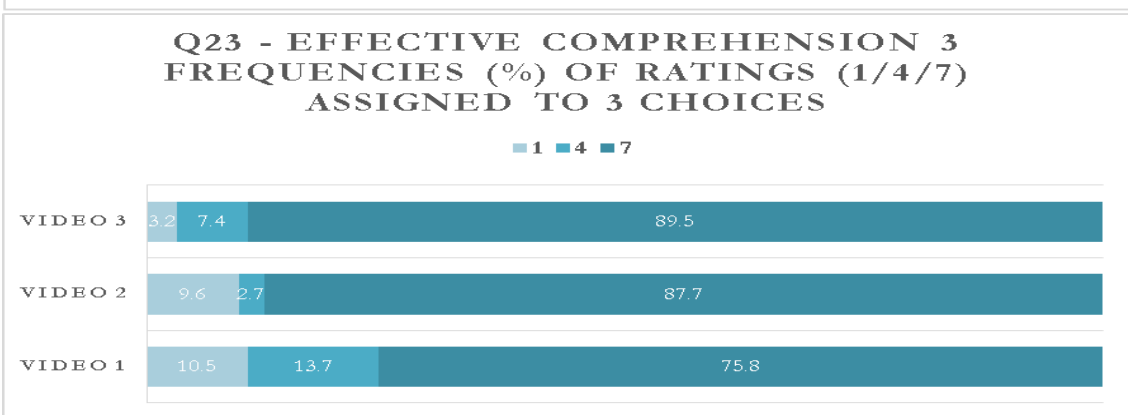
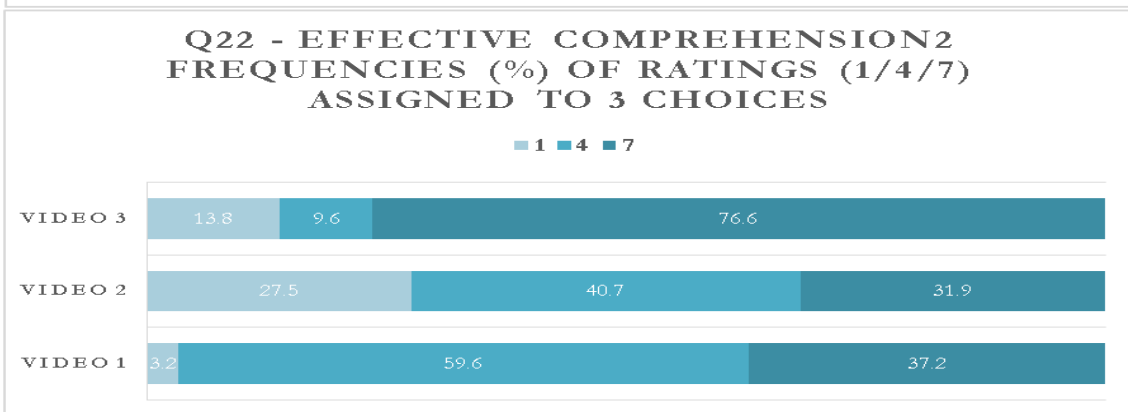
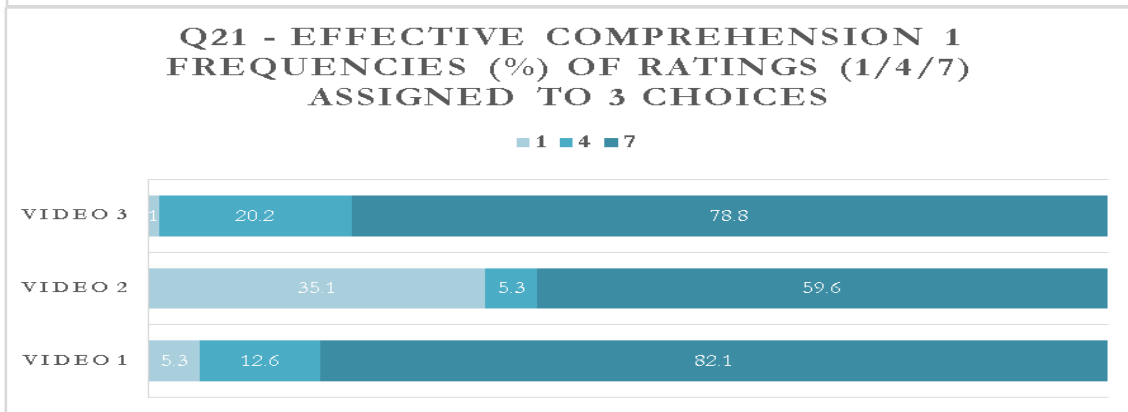
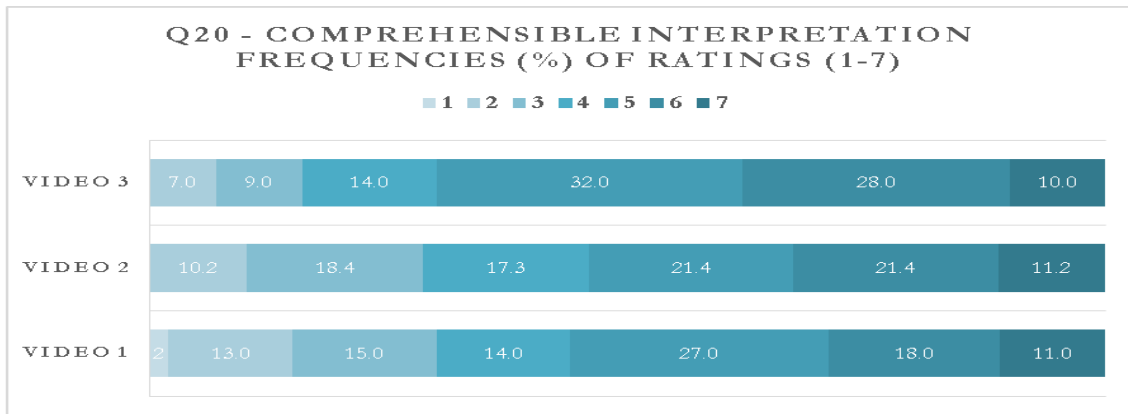












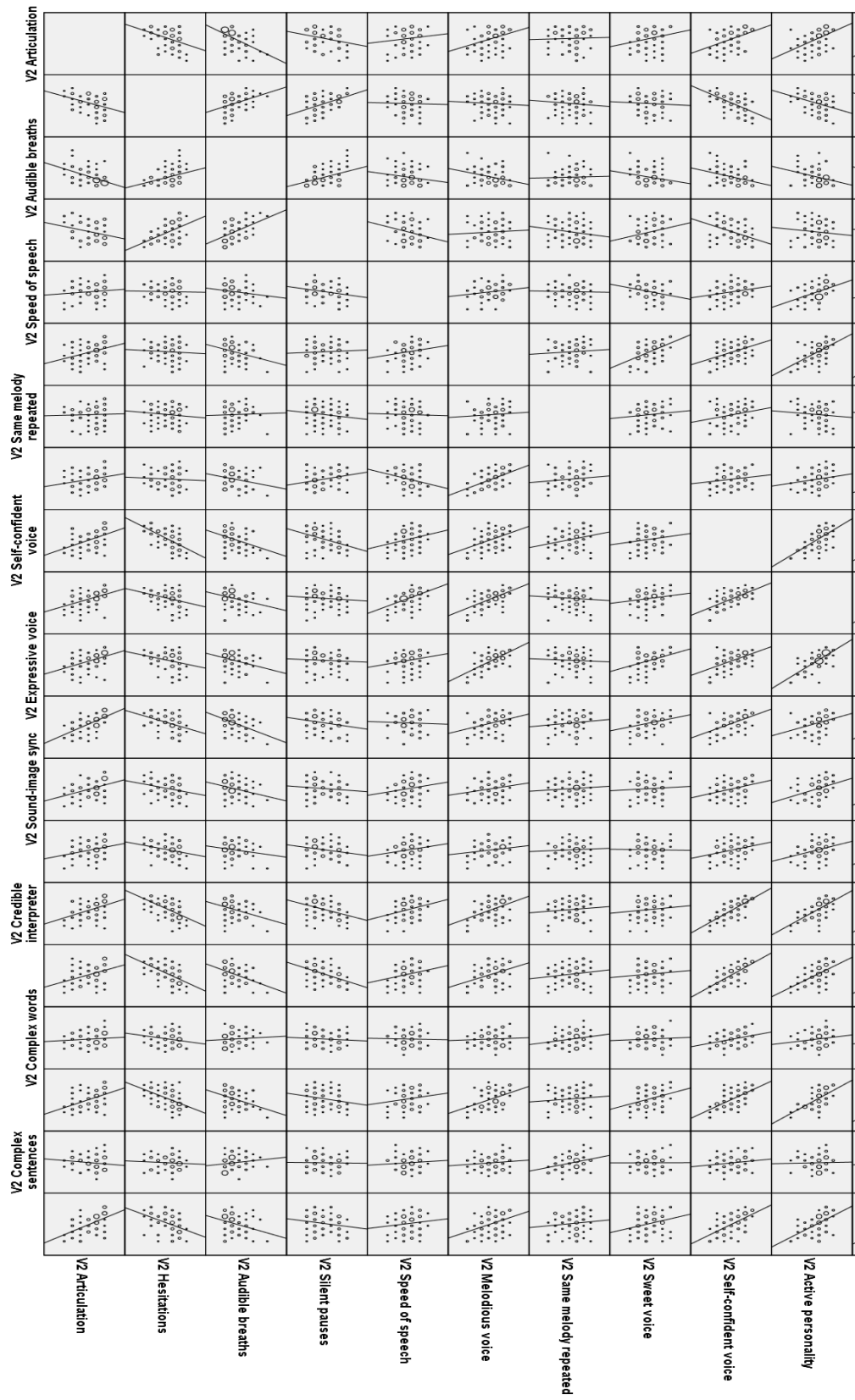
## **APPENDIX 11:**

### **Bivariate correlation of variables (pilot survey)**



	VI Complex sentences	VI Complex words	VI Credible interpreter	VI Sound-image sync	VI Expressive voice	VI Self-confident voice	VI Same melody repeated	VI Speed of speech	VI Audible breaths	VI Articulation
VI Articulation										
VI Hesitations										
VI Audible breaths										
VI Silent pauses										
VI Speed of speech										
VI Melodious voice										
VI Same melody repeated										
VI Sweet voice										
VI Self-confident voice										
VI Active personality										





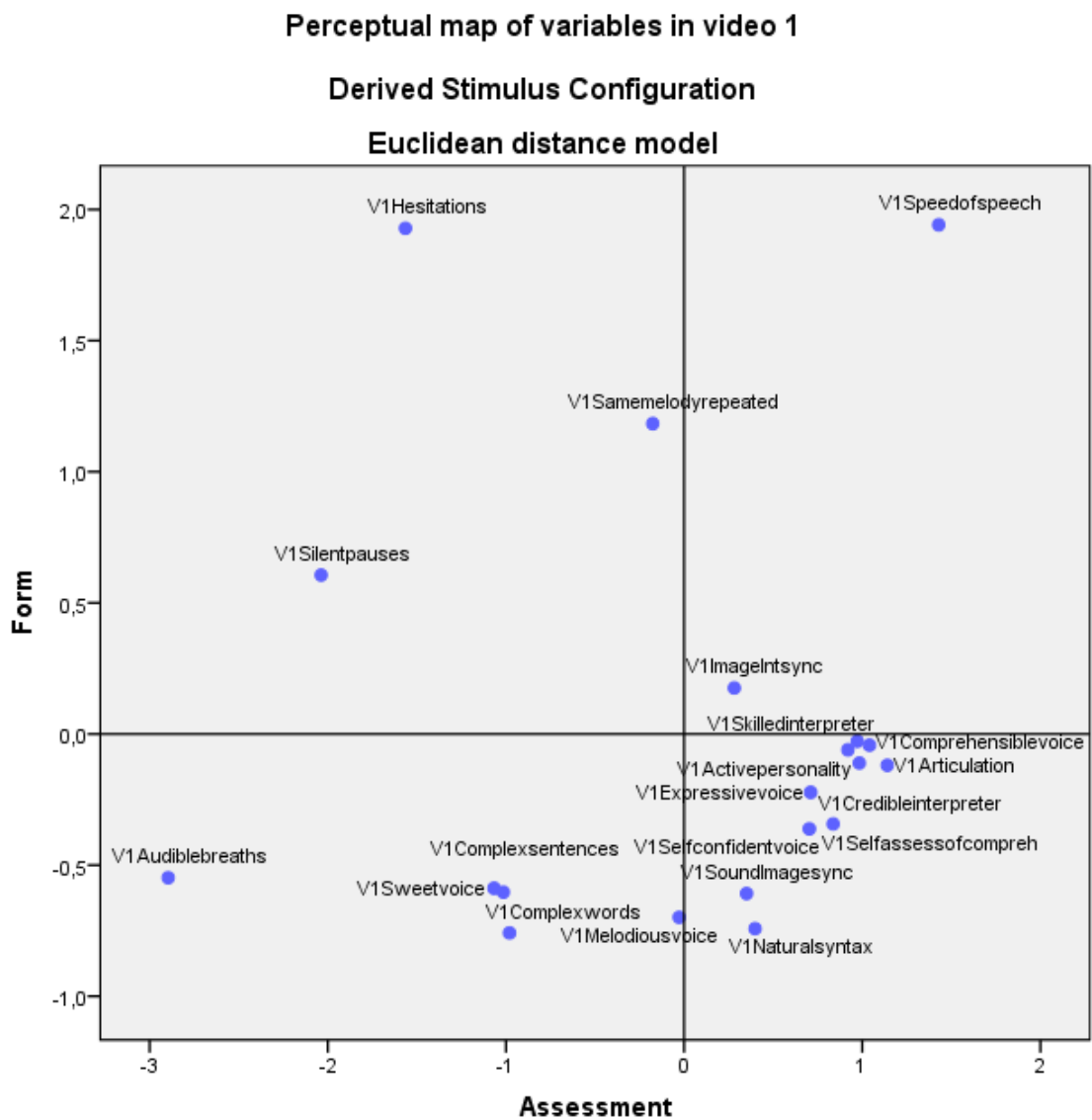


	V3 Complex sentences	V3 Complex words	V3 Credible interpreter	V3 Sound-image sync	V3 Expressive voice	V3 Self-confident voice	V3 Same melody repeated	V3 Speed of speech	V3 Audible breaths	V3 Articulation
V3 Articulation										
V3 Hesitations										
V3 Audible breaths										
V3 Silent pauses										
V3 Speed of speech										
V3 Melodious voice										
V3 Same melody/repeated										
V3 Sweet voice										
V3 Self-confident voice										
V3 Active personality										



**APPENDIX 12:**

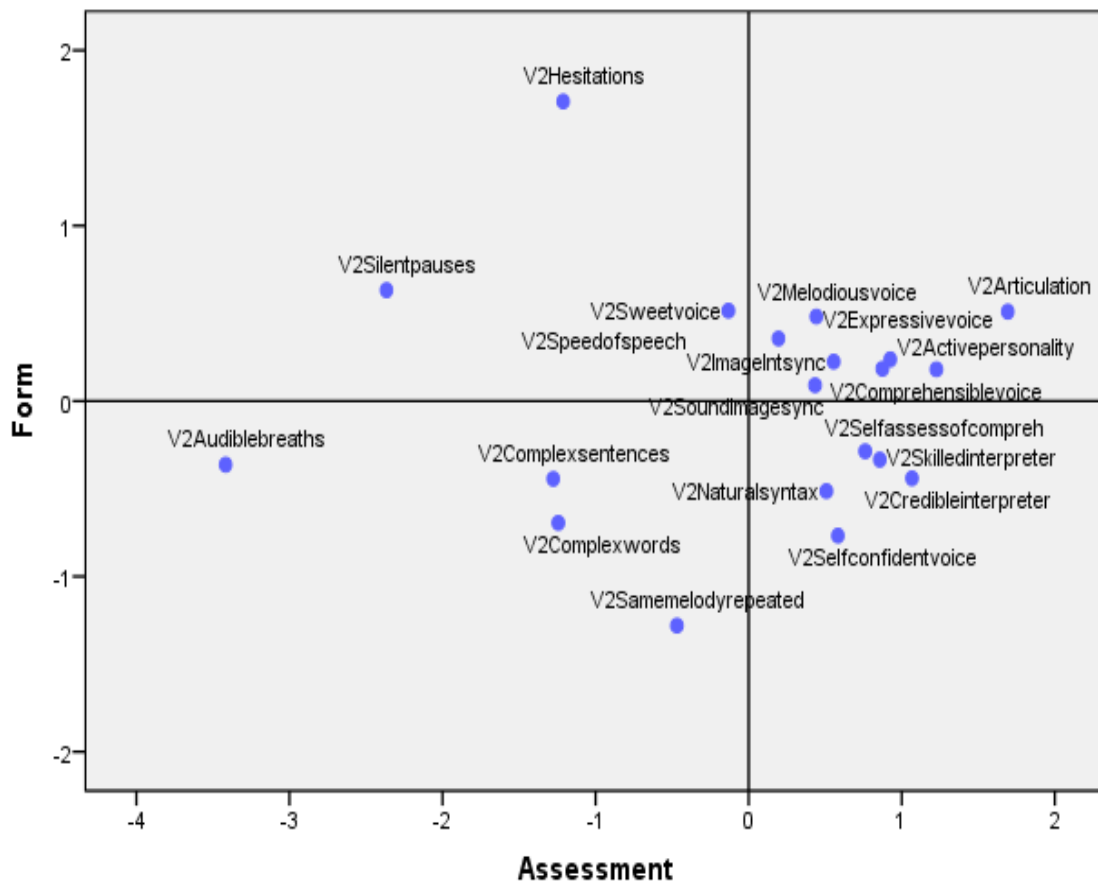
**Perceptual maps of variables by multidimensional scaling (pilot survey)**



### Perceptual map of variables in video 2

#### Derived Stimulus Configuration

#### Euclidean distance model

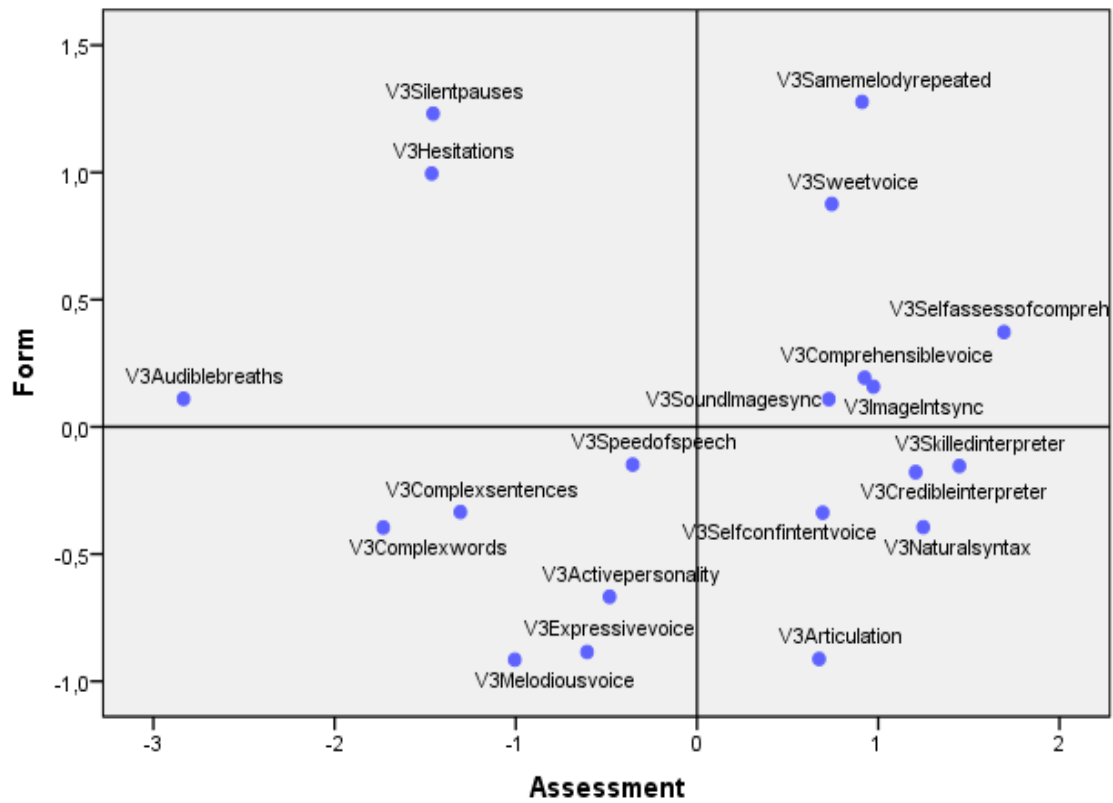




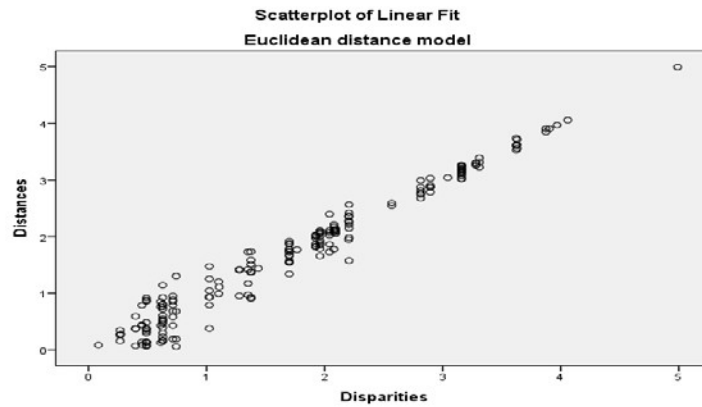
**Perceptual map of variables in video 3**

**Derived Stimulus Configuration**

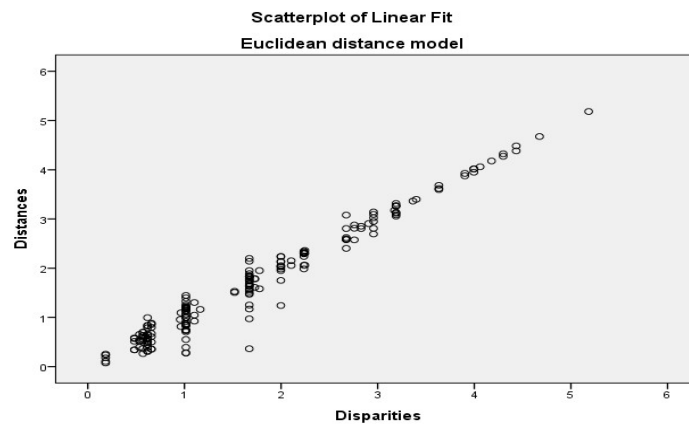
**Euclidean distance model**



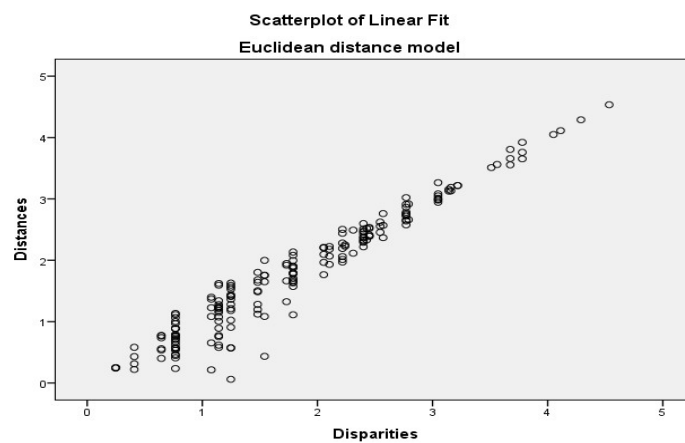
Shepard diagram – video 1



Shepard diagram – video 2



Shepard diagram – video 3



Note: scree plots are reported in section 4.7.3.4.4 (figures 4.30 – 4.32)

**APPENDIX 13:**

**Analysis of principal components – following scree plots (figures 4.30 – 4.32)**

Video 1

**Rotated Component Matrix<sup>a</sup>**

	Component	
	1	2
V1 Articulation	<b>,691</b>	,324
V1 Hesitations	<b>-,689</b>	,149
V1 Audible breaths	<b>-,449</b>	,397
V1 Silent pauses	-,451	,564
V1 Speed of speech	-,073	<b>-,374</b>
V1 Melodious voice	<b>,649</b>	,294
V1 Same melody repeated	,037	<b>,548</b>
V1 Sweet voice	,348	,132
V1 Self-confident voice	<b>,858</b>	,098
V1 Active personality	<b>,775</b>	,168
V1 Expressive voice	<b>,788</b>	,333
V1 Comprehensible voice	<b>,813</b>	,049
V1 Sound-image sync	<b>,685</b>	,138
V1 Image-Int sync	<b>,485</b>	,326
V1 Credible interpreter	<b>,882</b>	,187
V1 Skilled interpreter	<b>,863</b>	,097
V1 Complex words	,259	<b>,534</b>
V1 Natural syntax	<b>,823</b>	,151
V1 Complex sentences	,250	<b>,563</b>
V1 Self-assess of compreh	<b>,792</b>	,037
V1 Effective compreh 1	-,068	<b>-,246</b>
V1 Effective compreh 2	,019	,088
V1 Effective compreh 3	<b>,235</b>	,072

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser

Normalization.<sup>a</sup>

a. Rotation converged in 3 iterations.

Video 2

**Rotated Component Matrix<sup>a</sup>**

	Component				
	1	2	3	4	5
V2 Articulation	,408	,510	-,138	,362	,224
V2 Hesitations	-,104	<b>-,770</b>	-,136	-,093	,131
V2 Audible breaths	-,240	<b>-,626</b>	,176	-,123	-,249
V2 Silent pauses	,119	<b>-,774</b>	-,024	,205	-,265
V2 Speed of speech	,383	,024	-,141	<b>-,670</b>	,247
V2 Melodious voice	<b>,764</b>	,034	-,071	,125	,085
V2 Same melody repeated	-,175	,156	<b>,647</b>	,067	,123
V2 Sweet voice	,404	-,115	,141	<b>,709</b>	-,106
V2 Self-confident voice	,544	,647	,253	-,045	-,033
V2 Active personality	<b>,855</b>	,173	-,151	-,136	,017
V2 Expressive voice	<b>,832</b>	,135	-,072	,165	,119
V2 Comprehensible voice	,209	,612	,037	,461	,084
V2 Sound-image sync	,529	,140	,288	,108	,468
V2 Image-Int sync	,432	,208	,187	-,083	,577
V2 Credible interpreter	,664	,577	,195	-,087	,011
V2 Skilled interpreter	,537	,680	,234	-,049	-,007
V2 Complex words	,198	,125	<b>,675</b>	,136	-,048
V2 Natural syntax	<b>,646</b>	,481	,116	,122	-,114
V2 Complex sentences	,014	-,041	<b>,746</b>	-,086	-,034
V2 Self-assess of comreh	,548	,502	,051	,267	-,113
V2 Effective compreh 1	-,092	-,025	-,091	-,025	<b>,700</b>
V2 Effective compreh 2	,158	,133	-,457	,412	,332
V2 Effective comreh 3	,033	,342	-,204	<b>,540</b>	,301

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization. <sup>a</sup>

a. Rotation converged in 8 iterations.

Video 3

**Rotated Component Matrix<sup>a</sup>**

	Component				
	1	2	3	4	5
V3 Articulation	<b>,602</b>	,045	,314	-,223	-,404
V3 Hesitations	-,096	-,386	<b>-,549</b>	,233	-,114
V3 Audible breaths	-,067	,061	<b>-,677</b>	,269	,185
V3 Silent pauses	,055	-,110	<b>-,771</b>	-,124	-,135
V3 Speed of speech	-,025	,002	,504	,462	,135
V3 Melodious voice	<b>,705</b>	,071	-,153	-,070	,099
V3 Same melody repeated	-,052	<b>,429</b>	,036	-,247	-,430
V3 Sweet voice	,197	,192	-,172	<b>-,483</b>	,023
V3 Self-confident voice	,402	,320	,562	,150	-,052
V3 Active personality	<b>,721</b>	,208	,032	,194	,197
V3 Expressive voice	<b>,788</b>	,235	,058	,271	,156
V3 Comprehensible voice	<b>,652</b>	,186	-,008	-,427	-,240
V3 Sound-image sync	,111	<b>,787</b>	,192	,159	-,009
V3 Image-Int sync	,009	<b>,849</b>	-,019	-,040	,066
V3 Credible interpreter	,542	,598	,247	,021	,045
V3 Skilled interpreter	,584	,527	,441	,047	-,130
V3 Complex words	,047	,092	-,149	<b>,666</b>	-,125
V3 Natural syntax	,534	,373	,403	-,143	-,055
V3 Complex sentences	,260	,009	-,073	<b>,709</b>	,038
V3 Self-assess of compreh	,413	,610	,099	-,212	-,240
V3 Effective compreh 1	,166	-,042	-,029	-,017	<b>,724</b>
V3 Effective compreh 2	<b>-,334</b>	,147	-,043	-,072	,064
V3 Effective comreh 3	-,230	,107	,237	-,262	<b>,561</b>

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization. <sup>a</sup>

a. Rotation converged in 9 iterations.

## APPENDIX 14:

### Open-ended questions (pilot survey)

VIDEO 1 (A)	VIDEO 2 (B)	VIDEO 3 (C)
		<p>Problema legato alla credibilità: in una pausa vuota si è sentito parlare di "problemi", ma il corrispettivo in italiano non è emerso: problema dovuto al décalage lungo o omissione? / <i>problem related to the credibility: in a silent pause I heard "problem", but I did not hear the italian equivalent: too a long décalage or omission?</i></p>
<p>Per una migliore valutazione sarebbe stato utile ascoltare interpretazioni diverse dello stesso estratto. / <i>for a better evaluation it would have been more useful to listen to different interpretations of the same excerpts</i></p>		
<p>D24 - Si ha la sensazione che l'interprete sia continuamente in ansia per la preoccupazione di "star dietro" all'oratore. / <i>Q24 It seems to me that the interpreter is anxious because he is worried about lagging behind.</i></p>		
	<p>Poco accento sulle proprietà di eloquio. L'interprete ha una parlata molto gradevole e accattivante, ma ciò nondimeno l'interpretazione è scarsa, con scarsa proprietà di</p>	

	<p>linguaggio e pertanto poco naturale per un discorso di tal genere. Not a well-spoken Interpreter. / <i>S/he has a pleasant and attractive [accattivante] speech; nevertheless, the interpretation is not good, the use of language is not correct; therefore, the interpretation is not natural for such a type of discourse</i></p>	
		<p>Not a good start for the interpreter - it was difficult to grasp the words. Then the interpretation improved to worsen towards the end.</p>
<p>Con le domande di comprensione mi sono resa conto che alla fine del video non mi è rimasto impresso il succo del discorso. Non so se questo sia dovuto al fatto che mi sono concentrata troppo sulla forma dell'interpretazione, perché volevo "prepararmi" alle domande del questionario o perché l'interprete parlava troppo velocemente e in parte stentavo a seguirlo. Credo che sia dovuto in parte a entrambi i motivi./ <i>Thanks to the questions on comprehension I realized</i></p>	<p>Questa interpretazione era molto più fruibile della prima, soprattutto perché il ritmo era meno serrato. / <i>This interpretation was much more usable than the first one, especially because the rhythm was slower.</i></p>	<p>Questa risulta nel complesso l'interpretazione più comprensibile e che ispira maggior sicurezza, perché non ha né la velocità esagerata della prima né le esitazioni della seconda. / <i>This is the most comprehensible interpretation and the one that inspires more confidence, because it did not present neither the speedness of the first one nor the hesitations of the second one.</i></p>

<p><i>that I do not remember the content of the video. I do not know if this is due to the fact that I paid too much attention to the form of the interpretation, because I wanted to be more prepared to answer the questions or rather because the interpreter spoke too fast so I could not follow him. I think it is due to both the reasons.</i></p>		
<p>Ci sono delle domande un po' vaghe: D.14 → differenza con D.13?; D7. / <i>Some questions are a little vague: Q14 (what is the difference between this and Q13?); Q7.</i></p>		
<p>Pause troppo lunghe seguite da frasi pronunciate troppo velocemente. / <i>Too long pauses followed by sentences pronounced too fast.</i></p>		
<p>In alcuni passaggi è difficile seguire il filo del discorso a causa dell'alternanza di rallentamenti e accelerazioni nel discorso dell'interprete. Il ricorso dell'interprete a termini generici o inadeguati al registro rende la resa italiana un po' goffa. / <i>In some passages it is difficult to follow the thread (of what the interpreter is saying) because of the alternation of decelerations and accelerations in the interpreter's speech. The interpreter uses generic terms or improper terms to</i></p>	<p>Alcune frasi sono state lasciate aperte, il che ostacola la comprensione generale del discorso, soprattutto dalla metà in poi. All'inizio, al contrario, avevo ricevuto un'impressione di credibilità e fiducia nei confronti dell'interprete. / <i>Some sentences were not ended, this hampered my general comprehension of the speech, especially the second part. Conversely, at the beginning I had a good impression of credibility and confidence</i></p>	



<p><i>the register required by the context; for these reasons the Italian interpretation sounds a little clumsy.</i></p>	<p><i>towards the interpreter</i></p>	
<p>Ho trovato difficile rispondere alle domande sulla melodiosità della voce dell'interprete e sulla sincronizzazione tra le immagini e il discorso dell'interprete. / <i>I found it difficult to answer the questions on: i) the melodiousness of the interpreter's voice and ii) the synchronization of the images with the interpreter's speech.</i></p>		
<p>Tendevo ad essere più concentrata sulla qualità che sugli argomenti trattati nel video. / <i>I tended to be more concentrated on the quality than on the topics related to each video excerpt.</i></p>		
<p>Troppe pause piene e vocali trascinate ./ <i>Too many filled pauses and vowels dragged out.</i></p>		
		<p>La ridondanza nell'interpretazione fa presumere che ci sia stata una generalizzazione perché in alcuni punti si ripete e rimane sul vago. / <i>A redundancy in the interpretation makes me assume that there was a generalization, because in some points the interpreter repeats the same notion being vague.</i></p>
<p>La durata dei video è il minimo per poter dare una valutazione. / <i>The duration of video is the minimum for an evaluation.</i></p>		
<p>Nel caso del Video 2 (il primo che abbiamo guardato) non ho risposto alle domande sull'argomento perché mi sono concentrata solo sull'interpretazione, non perché l'interprete non fosse chiaro. Poi per i video 1 e 3 sapevo che ci sarebbero state domande sull'argomento, perciò ascoltando mi sono concentrata anche su quell'aspetto. / <i>In the case of Video 2 (the first one we watched) [??] I did not answer the questions related to the questions on the topic dealt with, because I focused only on the interpretation, but not because the interpreter was not clear. As to Videos 1 and 3, I knew that I would have found questions on the topic; therefore, in that case, while listening I also focused on that aspect.</i></p>		
	<p>Corrispondenza inglese-italiano: "we</p>	<p>Presenza di parole riconoscibili nelle pause</p>

	<p>have a free trade agreement with", interprete glissa. / <i>English-Italian correspondence: "we have a free trade agreement with..." - the interpreter skirts the issue.</i></p>	<p>dell'interprete poi non rese (es. "problemi" alla fine del video) / <i>During the interpreter's pauses I detected some words that were not rendered in the interpretation (eg. "problems" at the end of the video)</i></p>
<p>Forse sarebbe stato meglio una risposta multipla anche per valutare la performance dell'interprete, invece che la barra d'intensità (troppo dispersiva). / <i>Instead of an intensity bar (too much dispersive), a multiple choice answer would have been a better choice in order to assess the interpreter's performance.</i></p>		
<p>La spiegazione scritta della barra di intensità non è molto chiara. La D13 non sembra avere molto a che fare con i video mostrati. Molto interessante (e geniale) la parte di comprensione del testo. / <i>The written explanation of the intensity bar is not very clear. It seems to me that question n. 13 is not much related to the videos. Very interesting (and brilliant) the part related to the comprehension of text.</i></p>		
<p><i>Perhaps the video is too short [1 min] to perceive well all the aspects analyzed in the questionnaire.</i></p>		
	<p>Auto-correzioni, specialmente cambio di pronome personale a inizio frase. / <i>Self-repairs - especially the replacement of personal pronoun at the beginning of the sentence.</i></p>	
<p>È stato difficile percepire la differenza tra "dolce" e "aggressiva" in riferimento alla voce dell'interprete. Con un secondo ascolto probabilmente sarei stata più precisa nella scelta delle risposte. / <i>It was difficult do distinguish the difference between "aggressive" and "sweet" voice of the interpreter. A second listening might have me allowed to be more precise in selecting the answers.</i></p>		
<p>Nessuna delle tre interpretazioni risulta pienamente comprensibile. / <i>None of the three interpretations turned out fully comprehensible.</i></p>		
<p>Non è sempre stato facile capire subito l'argomento degli spezzoni di video presentati. / <i>It was not always easy to immediately understand the topic of the video excerpts.</i></p>		
	<p>Il migliore dei 3 interpreti (forse per la voce più piacevole). / <i>The best of the three interpreters (maybe</i></p>	

	<i>thanks to the most pleasant voice).</i>	
<p>La coerenza e la coesione del discorso dell'interprete sono aspetti ai quali avrei dedicato più spazio. / <i>I would have paid more attention to the coherence and the cohesion of the interpreter's speech.</i></p>		
<p>Alcune domande lasciano spazio all'interpretazione personale. Mancavano domande sui fattori di distrazione nel video. / <i>Some questions leave room to personal interpretation. There was a lack of questions on factors of distraction in the video.</i></p>		
<p>Aveva un "time-lag" notevole. / <i>[The interpreter] had a notable time-lag.</i></p>		
<p>Le scelte linguistiche mi hanno lasciato perplessa. / <i>The linguistic choices made [by the interpreter] puzzled me.</i></p>	<p>L'interprete non utilizza un tono noioso, nonostante faccia un paio di riprese la resa è abbastanza piacevole. The interpreter does not use a boring tone [intonation?]. / <i>Although a couple of repairs (riprese) the delivery is not too bad (abbastanza piacevole).</i></p>	
<p>Consiglio di far vedere video più lunghi per capire di più il senso del discorso. / <i>My advise is to show longer video excerpts to allow a better understanding of the meaning of the speech.</i></p>		
<p>Ho notato una leggera incongruenza nella costruzione della frase "importiamo dalla Cina dando i soldi all'Arabia Saudita", ma si comprende comunque il senso. / <i>I noticed a slight inconsistency in the construction of the sentence 'we import from China giving money to Saudi Arabia', nevertheless the meaning is comprehensible.</i></p>		

<p>Talvolta la comprensione di alcune frasi è risultata difficoltosa, ma probabilmente ciò non è dovuto solamente ad un'incertezza da parte dell'interprete ma anche al fatto che gli estratti erano troppo brevi per poter meglio inferire ciò che l'interprete stava interpretando. <i>Sometimes I found it difficult to comprehend some sentences; it might be due not only to the interpreter's hesitations but also to the duration of the video excerpts, which were too short to make me infer what the interpreter was interpreting.</i></p>		
	<p>Non sempre le strutture dell'italiano erano perfette, ma nell'insieme il messaggio era chiarissimo. / <i>The structures of Italian language were not always perfect, however the global message was very clear.</i></p>	<p>La voce dell'interprete sembra talvolta essere un po' sottotono; questo potrebbe dare al pubblico una sensazione di insicurezza. / <i>It seems to me that sometimes the interpreter speaks in a low tone of voice, this could give the audience a feeling of insecurity.</i></p>
<p>Molto utile anche per me in fase di studio delle tecniche di Interpretazione, poiché mi permette di riflettere su più aspetti, sia di forma che di contenuto. / <i>[I found the questionnaire] very useful since I am studying interpretation techniques; in fact, it allows me to focus on some aspects related both to the form and the content.</i></p>		
<p>Pareva che parlasse a sproposito di "energia" ed "energetico". / <i>It seemed to me that the interpreter talked about 'energy' without understanding what he was talking about.</i></p>		
<p>Credo che le domande siano complete e ampie, forse è difficile rispondere bene a tutte perché la durata dei video è troppo breve e non si riesce a focalizzare bene l'attenzione sull'interpretazione in questione. / <i>I think that the question are complete and exahustive, maybe it is difficult to answer correctly to all of them because the duration of the video excerpts is too short and it is not easy to focus one's attention on the relative interpretation.</i></p>		
<p>Non so se sia un'informazione pertinente ma forse sarebbe da chiedere se il tono (volume) di voce dell'interprete sia troppo alto o troppo basso. Per esempio io l'ho trovato troppo alto nel primo e troppo basso nel secondo. / <i>I don't know if the following is a relevant information: it should be asked wether the tone (volume) of the interpreter's voice is too loud or too low. For example, I found it too loud in the first video and too low in the second one.</i></p>		

<p>L'interprete dava l'impressione di annaspire alla ricerca delle parole, probabilmente un qualcosa di dovuto alla velocità del presidente Obama. E' come se volesse dire tutto e alla stessa velocità. / <i>The interpreter gave me the impression of fumbling for the words, probably because of the president Obama's speaking speed; it's like he wanted to say everything at the same speed.</i></p>	<p>Diverse omissioni, anche delle parole chiave dell'estratto (es: trade agreement → verso la fine). / <i>I found some omissions, even relating to key words of the excerpt (eg. trade agreement → at the end)</i></p>	
<p>Credo che potrebbe essere inserita una domanda circa la coerenza delle frasi. / <i>I think that a question on the coherence of the sentences should have been included.</i></p>		
	<p>I nessi logici tra le diverse parti del discorso non erano molto chiari. / <i>the logical links among the different parts of speech were not very clear.</i></p>	<p>La seconda parte del video risultava comprensibile, la prima metà invece difficile da seguire e capire per due motivi: il discorso in inglese era trasmesso con un volume che non consentiva di sentire bene in italiano, i due discorsi erano sovrapposti. L'interpretazione era molto frammentata. / <i>The second part of the video was comprehensible, while the first part was difficult to follow and understand for two reasons: i) the speech in English had a volume that did not allow to listen well to the Italian speech - the two speeches were overlapped; ii) the interpretation was fragmented.</i></p>

<p>C'è stata una pausa abbastanza imbarazzante nei primi minuti, quando Obama parlava delle leggi di Bush e lo spettatore italiano poteva di sicuro rendersi conto dell'incertezza dell'interprete. Poi alla fine, invece l'interprete era molto in ritardo rispetto a cosa veniva detto da Obama e questo ritardo è diventato palese quando Obama, ormai avendo finito di parlare se n'era andato, mentre l'interprete ha ancora parlato per qualche secondo. / <i>In the first minutes there was a very embarrassing pause, it occurred when Obama was talking about Bush's Acts and the Italian spectator surely could detect the interpreter lack of confidence. Instead, at the end of the video excerpt, the interpreter was very behind with respect to what Obama was saying; this delay became clear when Obama, after finishing his speech, went away; while the interpreter went on speaking for a few seconds.</i></p>	<p>L'interprete ha usato qui una voce monotonale e noiosa e nell'interpretazione ha cercato di riprodurre il TP senza svolgere un vero lavoro d'interpretazione nella LA, facendo spesso delle costruzioni atipiche per l'italiano. / <i>The interpreter used here a monotonous and boring voice and, in the interpretation, tried to reproduce the source text without executing a real interpretation in the target language, since he often used unusual Italian structures.</i></p>	<p>La cosa che mi ha colpito di più del video è la voce dell'interprete che mi è sembrata molto inespressiva, piatta e poco comunicativa. / <i>What impressed me most in the video is the interpreter's voice - I found it very unexpressive, monotonous and little communicative.</i></p>
<p>Trovo questo distionario [sic] molto utile all'interno dello studio dell'interpretazione, ma anche nell'ambito della traduzione, soprattutto per quanto riguarda la resa in italiano, non sempre chiara e precisa. / <i>I found this</i></p>		

<p><i>questionnaire very useful for the purpose of studying interpretation, but also for studying translation, especially for what concerns the Italian rendering, not always clear and accurate.</i></p>		
		<p>L'interprete si è mostrato esitante solo verso la fine, mentre nella parte iniziale e centrale è riuscito a stare bene al passo con l'oratore e non ha avuto grosse esitazioni. / <i>The interpreter proved himself hesitant only at the end [of the excerpt], while in the initial and central parts he managed to follow the speaker and did not have notable hesitations.</i></p>
		<p>L'interprete ha tralasciato diverse informazioni, soprattutto nell'ultimo passaggio dove l'intera frase di McCain è stata interpretata con poche parole. / <i>The interpreter omitted several pieces of information, especially in the final part where an entire sentence by McCain was interpreted using few words.</i></p>
<p>Ho l'impressione che non sempre l'interprete stia capendo il messaggio dell'oratore, motivo per cui alcune riformulazioni risultano piuttosto goffe (vorrei spendere molti soldi sull'argomento). / <i>My impression is that the interpreter does not understand the speaker's message all the time; therefore, some restructurings turned out</i></p>	<p>Il contenuto informativo non è del tutto chiaro (alla fine dello spezzone ascoltato non mi è chiaro il messaggio e l'intento comunicativo dell'oratore). Questo può tuttavia dipendere da una scarsa pianificazione del TP e da formulazioni criptiche dell'oratore, che hanno causato</p>	<p>Risulta faticoso seguire il discorso dell'interprete (specialmente se si considera una durata superiore a quella considerata nel presente studio). / <i>Following the interpreter's speech turned out a difficult task - especially considering a longer duration than that considered in this study.</i></p>

<p><i>very awkward (I would like to spend a lot of money on this topic).</i></p>	<p>difficoltà di riformulazione nel testo interpretato. / <i>The content of the information is not very clear (at the end of the excerpt that I listened to it is not clear to me what is the message and the communicative intention of the speaker). However, this may be due to a poor planning of the source text and to cryptical structures by the speakers, that made it difficult to restructure the interpreted text.</i></p>	
<p>La domanda D7 è particolarmente difficile perché non è possibile ascoltare veramente la melodia dell'oratore, se non durante le pause dell'interprete. Tuttavia manca una visione d'insieme sui tratti prosodici dell'oratore. / <i>The Question 7 is particularly difficult because it is not possible to really listen to the speakers's melody, except during the interpreter's pauses. However, there is a lack of global vision on the speaker's prosodic features.</i></p>		
<p>Trattandosi di un dibattito credo che la relazione con le immagini sia difficilmente valutabile. Quando il discorso presenta difetti in italiano è possibile dare già un giudizio sulla competenza, ma se sembra corretto non emergono eventuali problemi di corretta resa del contenuto, che non è possibile valutare. / <i>I think that it is difficult to evaluate the relationship with the images in a debate. When the Italian speech has something wrong it is possible to express a judgment on competence, but it seems to me it is correct, since I didn't find anything wrong in the correct rendering of the content, that is possible to assess.</i></p>		
	<p>L'interpretazione era chiara, nonostante alcuni tentennamenti dell'interprete. Ha un po' balbettato, ma questo non disturbava troppo la comprensione. Purtroppo ho ascoltato solo la voce e il tono</p>	



	<p>dell'interprete, senza prestare attenzione al contenuto del discorso di Obama. / <i>The interpretation was clear, despite some hesitations by the interpreter. He stammered out a little bit, but this did not disturb too much comprehension.</i></p> <p><i>Unfortunately, I listened only the interpreter's voice and tone, I did not pay attention to the content of Obama's speech.</i></p>	
		<p>Si nota la presenza di calchi dall'inglese che sono evidenti anche se il TP non è stato ascoltato. / <i>I noticed the presence of calques from English, that were evident even if I didn't listen to the source text.</i></p>
<p>Ogni tanto l'interprete si autocorreggeva cambiando le parole e la struttura della frase. / <i>Once in a while the interpreter corrected himself by chainging the words and the structure of the sentence.</i></p>		
		<p>Voce poco coinvolta/coinvolgente, a tratti fastidiosa. / <i>The voice was a little captivating, sometimes annoying.</i></p>
<p>L'interprete risultava molto sicuro prendendo in considerazione il tono della voce e la sua attitudidine. In alcune</p>		

<p>occasioni ho però riscontrato la presenza di alcune frasi che non sono state debitamente completate. / <i>The interpreter appeared very self-confident considering the tone of his voice and his aptitude. However, in some cases I detected some sentences that were not duly finished.</i></p>		
<p>Le domande del questionario prendono in considerazione tutti gli aspetti principali dell'interpretazione, quindi mi sembra uno studio abbastanza completo. Anche le domande di comprensione sono ben mirate sull'argomento principale dell'interpretazione. / <i>The questions concern all main aspects of interpretation, therefore the research appears exhaustive. Questions on comprehension are well focused on the main topic of the interpretation.</i></p>		
<p>Le domande presentate nel questionario ricoprono i principali aspetti che avrei considerato anche io nel "giudicare" le tre interpretazioni, mi è sembrato piuttosto completo. Le domande di comprensione andrebbero però messe prima di tutte le altre perché concentrandosi sulle precedenti, si arriva poi alle ultime tre con le idee non più così chiare. <i>The questions concern all the main aspects that I myself would have considered for "judging" the three interpretations - the questionnaire seems to be exhaustive. The questions on comprehension should be placed before all the others, because at the end of the questionnaire the attention is absorbed by the previous questions and the ideas are not so clear as to answer them.</i></p>		
<p>Décalage ampio. False partenze. / <i>Long décalage. False starts.</i></p>	<p>Uso di molte pause piene con intercalari. / <i>A lot of filled pauses and stock phraese have been used.</i></p>	
<p>Mancherebbe un confronto finale tra le tre interpretazioni. / <i>A final comparison of the three interpretations is missing.</i></p>		
<p>Delle tre, è l'interpretazione migliore. / <i>This is the best interpretation of the three.</i></p>	<p>Positivo l'uso della prosodia. / <i>Good use of prosody.</i></p>	<p>Il maggior difetto evidente è la ridotta scansione delle parole. / <i>The major fault is the poor scansion of words [pronunciation].</i></p>
<p>In quanto non studentessa di Interpretazione di conferenza, mi è più difficile giudicare la qualità dell'attività di interpretazione. Ho meno conoscenze e risorse per fornire un commento completo. Tuttavia una persona meno</p>		

<p>esperta può giudicare chiaramente: il senso logico di ciò che viene detto, la prosodia e le parole utilizzate. / <i>Since I am not an interpreting student, I find it difficult to judge the quality of the interpretation. My competences and abilities prevent me from express an exhaustive commentary. However, the less competent one is, the clearer may her/his judgment be: on the logical sense of what is said, prosody and words used.</i></p>		
<p>Le frasi dell'interprete erano molto sconnesse, si fatica a comprendere il senso a livello globale, sono frasi molto brevi che da sole potrebbero avere senso, ma non capisco se fosse Obama a parlare così. / <i>The interpreter pronounced unconnected sentences, therefore global comprehension requires high effort; sentences are short, taken separately they could make sense - I don't understand wether the Obama's speech had the same feature.</i></p>		<p>Nonostante la voce dell'interprete risulti meno chiara rispetto al video precedente, i contenuti sono molto più comprensibili e c'è meno esitazione. / <i>Despite the interpreter's voice sounds less clear than the previous video, the content [in this video] is more comprehensible, and there are less hesitations.</i></p>
<p>Avrei voluto sentire anche la voce dei politici per fare un confronto più oggettivo. / <i>I would have bear the politician's voice to make a more objective comparison.</i></p>		
<p>Sarebbe preferibile sapere i contenuti delle domande prima di guardare i video in modo da avere presenti gli aspetti da valutare. / <i>It would be preferable to know the content of questions before watching the video excerpts, in order to consider the aspects that have to be assessed.</i></p>		
		<p>Alcune parole erano scandite male. / <i>Some words were poorly pronounced.</i></p>
<p>alcune risposte alle domande di comprensione potrebbero risultare in contrasto con i giudizi sull'interpretazione perché i video sono molto brevi e risulta difficile seguire il filo del discorso essendo a metà dibattito. / <i>Some answers to the questions on comprehension could appear in contrast to the judgment on the interpretation because the video are very short and it is difficult to follow the thread of the argument in the middle of the debate.</i></p>		

	<p>Siccome l'interpretazione viene fatta in voice-over e pertanto non si sente il testo d'origine, si possono fare domande anche su questo aspetto, ad esempio se si preferisce avere la possibilità di fare un confronto col TO e quindi poter controllare la delivery dell'interprete, oppure se ai fini della comprensione è meglio ascoltare solo il testo interpretato. / <i>Since the interpretation is received in voice-over mode, and consequently the source text can't be heard, some questions on this aspect could have been asked; for example whether the subject would have liked to compare [the interpretation] with the source text, in order to control the interpreter's delivery, or rather would have preferred to listen to only the interpreted text.</i></p>	
<p>Non sono molto chiare le domande "l'interpretazione è/non è in sincronia con le immagini" e "l'interpretazione dà/non dà conto di quanto mostrato dalle immagini. Anche la scala d'intensità come metro di giudizio è un po' fuorviante, forse sarebbe meglio dare punteggi. / <i>The questions "the interpretation is (not) synchronized with the images" and "the interpretation does (not) account for what is shown in the images" are not very clear. Also the intensity bar as evaluation scheme is a bit misleading - a numerical rating scale would have been better off.</i></p>		

	Alcune false partenze e riformulazioni a metà frase. / <i>Some false starts and restructurings in the middle of sentences.</i>	
False partenze e riformulazioni, calchi. Magari spostare le domande di comprensione all'inizio. / <i>False starts and restructurings, calques. I suggest to move the questions on comprehension at the beginning of the questionnaire.</i>		
Caotico, confuso. Scandiva bene le parole, ma il contenuto era confuso e poco comprensibile. / <i>Chaotic, confused. He pronounced words with distinction, but the content was confused and little comprehensible.</i>	L'interpretazione è molto buona, l'interprete è a suo agio e anche se esita un po' in un punto non perde il filo del discorso e risolve la frase coerentemente. / <i>The interpretation is very good, the interpreter is at ease and even if he hesitates a bit in a given point, he does not lose the thread of his talk and resolves the sentence in a coherent way.</i>	La voce dell'interprete è piuttosto impastata, il che rende un po' faticosa la comprensione e condiziona la valutazione sui contenuti: certe frasi erano incomprensibili e dette troppo velocemente. / <i>The interpreter doesn't speak with distinction, consequently the comprehension requires a bit of effort and has an impact on the assessment of the content: some sentences were incomprehensible and pronounced too fast.</i>
Qualche domanda sul contenuto, sulla completezza e la fruibilità del messaggio da parte del destinatario. Per il resto, è una figata! / <i>Some questions on the content, the completeness and the usability of the message [could be improved?], but for the rest it rocks!</i>		