A PROPOSAL FOR AN ASSESSMENT OF VOICE QUALITY IN TV BROADCAST SIMULTANEOUS INTERPRETING THROUGH A GESTALTIC APPROACH: THEORETICAL PARADIGM FOR A NEW QUESTIONNAIRE

Gregorio De Gregoris
gregorio.degregoris@phd.units.it
Università di Trieste

Abstract

After a literature review of questionnaire-based surveys on quality evaluation (both expectations or ideal preferences and assessment or judgment after a real experience) of simultaneous interpreting (De Gregoris 2014), Albano Leoni's model of speech perception (*Il volto fonico delle parole*, 2009), Chion's proposal for audio-vision perception (*L'Audio-vision*, 1994) and Fónagy's research on live speech (*La vive voix*, 1983) were studied in order to understand how to elicit an assessment of voice quality in television broadcast simultaneous interpreting from a gestaltic point of view.

Riassunto

Dopo aver eseguito una panoramica sulle indagini basate su questionari riguardanti la valutazione della qualità dell'interpretazione simultanea (sia le aspettative o le preferenze ideali che la valutazione o il giudizio in seguito ad un'esperienza reale; De Gregoris 2014), si è passati a studiare il modello di percezione del parlato proposto da Albano Leoni (*Il volto fonico delle parole*, 2009), la proposta di Chion per la percezione audiovisiva (*L'Audio-vision*, 1994) e lo studio di Fónagy sulla viva voce (*La vive voix*, 1983) al fine di capire come ottenere una valutazione gestaltica della qualità della voce nell'interpretazione simultanea televisiva.

Keywords: Survey. Gestalt. Assessment. Quality. Interpretation.

Parole chiave: Questionario. Gestalt. Valutazione. Qualità. Interpretazione.

MonTI Special Issue 3 (2016: 273-306). ISSN 1889-4178
DOI: 10.6035/MonTI.2016.ne3.10

Manuscript received on April 23, 2015 and accepted for publication on September 28, 2015.

1. Introduction

Interpreting Studies on both quality expectations (ideal evaluation or expression of preferences) of simultaneous interpreting (Bühler 1986; Kurz 1989, 1993; Kopczyński 1994; Chiaro & Nocella 2004; Moser 1995; Pöchhacker & Zwischenberger 2010) and quality assessment (evaluation after real experience or judgment) of simultaneous interpreting (Gile 1990; Marrone 1993; Vuorikoski 1993; Mack & Cattaruzza 1995; Garzone 2003; Russo 2005; Catana 2005; Collados Aís et al. 2007; García Becerra 2013) show that the 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, and use of appropriate style. These criteria were adopted in subsequent studies of the kind, 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; De Gregoris 2014).

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 simultaneous interpreting (SI). García Becerra (2013: 571) observed that "it looks as if 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 holística".

Considering the high ratings assigned to the voice as quality criterion in the survey on quality expectations by Kurz & 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 quality of a television broadcast simultaneous interpretation has to take into consideration, among other aspects, the influence the medium may have on perception.

2. Prosody and simultaneous interpretation

2.1. Cognitive rhythm and speech production

Goldman-Eisler (1968) conducted a series of "Experiments in spontaneous speech"; however, he also used reading and simultaneous interpreting to compare different cognitive activities, all related to speech production.

In one experiment concerning simultaneous interpretation, Goldman-Eisler (1968: 76-80) studied hesitations in speech (ratio pauses/speech) related to the structure of sentences (quantified through a "subordination index"). Results showed that both in spontaneous speech and in simultaneous interpreting, the two above-mentioned parameters seemed "independent". 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".

In the analysis of simultaneous interpretations (Goldman-Eisler 1968: 87-89), no relation was found between the temporal rhythm of output (interpretations) and input (source text). However, "highly significant relation" was found when the 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, having these a more rapid pace, among all other source texts.

The addition of pauses by the interpreter, even in the fast speed input, to create a rhythmic speech, led the researcher to revise his psycholinguistic hypothesis, which considered pausing related to "planning" and speech related to "achievement" or "execution". This phenomenon was then related to a "feedback-control" over the speech process, playing an "inhibitory" task; while from a behavioural point of view, it was considered as the manifestation of a "totality of attitude", a "specific neurophysiological set pervading the

whole situation" of speech production. The conclusive proposal was that of a "global tonigenic activation" functioning on the background of a "selective process" in speech production (Goldman-Eisler 1968: 90-93).

2.2. Simultaneous interpreting prosody

Shlesinger (1994) conducted a perceptual survey on the effect of simultaneous interpreting intonation on comprehension and recall. By comparing the differences in intonation between read-aloud and SI of the same texts (English-Hebrew and Hebrew-English), she found that SI intonation 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" (Shlesinger 1994: 229); occurrences of "stress incompatible with semantic contrast" (Shlesinger 1994: 231) were also found in a remarkable number. These aspects of SI intonation negatively affected comprehension and recall.

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) analysed the prosody of a SI corpus (72 min, English-German) acoustically. 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 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 the 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 the overall quality assessment of 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, low fluency had a negative impact on the assessment of the 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 among two experimental conditions and one control condition did not reach 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 a 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" (Tohyama & Matsubara 2006: 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" (Tohyama & Matsubara 2006: 896).

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 & 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 & Lenglet 2014) were also related to the difference among the two groups of subjects; in particular, the more fluent interpretation was

perceived as more accurate by translation students; the impact of interpretation fluency was lower for students of economics (because of their better knowledge of the subject matter, according to the researchers' assumption); however, translation students scored better than economics students.

2.3. 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 exposed, Golman-Eisler's "cognitive rhythm" should be marked by a formal, and not a temporal aspect. However, from Plato, 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 phonemes, inherited from the

previous psychologistic 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 of this research study recovers the original notions of "form" and "structure", at least as far as the analysis of speech perception is concerned, as will be better explained in the next paragraph.

3. A proposal for eliciting a gestaltic perception of simultaneously interpreted speech

Albano Leoni's proposal of speech perception is titled *Il volto fonico delle parole* (2009) ("The phonic facet of words"; my translation), where "the linguistic unit of perception and processing is a phonological word or a word group or any other significant unit grasped in its essence in discourse" (Albano Leoni 2009: 165; my translation of quote). As the author recognises, the notion of "phonic facet of words" is not new; it was first introduced by the German psychologist and linguist Karl Bühler (1983 [1934]; in Albano Leoni 2009: 166). Nonetheless, it was not taken into account by phonology at that time (Albano Leoni 2009: 94). Since such a model does not accept segmentation, there is no distinction between segmental and suprasegmental features, or linguistic and paralinguistic aspects. The features of the "phonic facet of words" are "voice", "syllable" and "prosody"; other relevant aspects of this model are "sense" and "context" (Albano Leoni 2009: 183; my translation of quotes).

Albano Leoni's proposal of speech perception is not so far from Meschonnic's theory of "rhythm in language", where he considers rhythm as the "form" of discourse, drawing on the original definition of rhythm proposed by Benveniste (1966: 327-335), i.e. rhythm < gr. ῥυθμός < ῥειν ("to flow"), where ῥυθμός means "la forme dans l'instant qu'elle est assumée par ce qui est mouvant, mobile fluide, la forme de ce qui n'a pas consistence organique" (Benveniste 1966: 333). As Meschonnic (1982: 70) states, "from Benveniste on, rhythm may no longer be considered secondary to form, since it means 'organisation (disposition, configuration) d'un ensemble'". Rhythm is the form of a language, the way discourse is organised. Hence, rhythm is the organisation of the sense of a discourse, and it is peculiar to one discourse; therefore, it coincides with the sense of that discourse (cf. Meschonnic 1982: 70). Rhythm is not the result of the elements of a discourse, or the result of the processing of that discourse, it is the discourse itself (Meschonnic 1982).

Sense is created by discourse, and creates the discourse itself (Meschonnic 1982). As a consequence of this, there is no distinction between the so-called "segmental features" (phonemes) and suprasegmental aspects (voice and prosody). Thus, intonation is not excluded from sense; rather, it can make the whole sense of a discourse, even more sense than that of single words. Meschonnic (1982: 216) identifies discourse with rhythm, considering the latter as an "ensemble synthétique" of the elements that make up the discourse.

Both Albano Leoni's and Meschonnic's proposals reject the distinction between segmental and suprasegmental, linguistic and paralinguistic features. They do this by taking into account the "rhythm in language" as the form in the instant that is assumed by all the marks that display themselves in a particular way in a given discourse to produce the "signifiance", i.e. a specific semantics, which is not limited to the lexical meaning, constituted by all the values that make a discourse (Meschonnic 1982: 216-217). Considering the notion of "phonic facet of words", it is possible to think of a paradigm of speech perception where the marks of the *signifiance* are the same features as the "phonic facet of words" identified by Albano Leoni, i.e. voice, syllable and prosody, which, through the context and on the basis of the linguistic and non-linguistic knowledge of the world, contribute to create sense. Therefore, in the flow of the speech chain, the sense of a discourse develops according to the rhythm, i.e. the display, the organisation, the form of voice-syllable-prosody-sense-context-(linguistic) knowledge of the world assumed in subsequent instants of time.

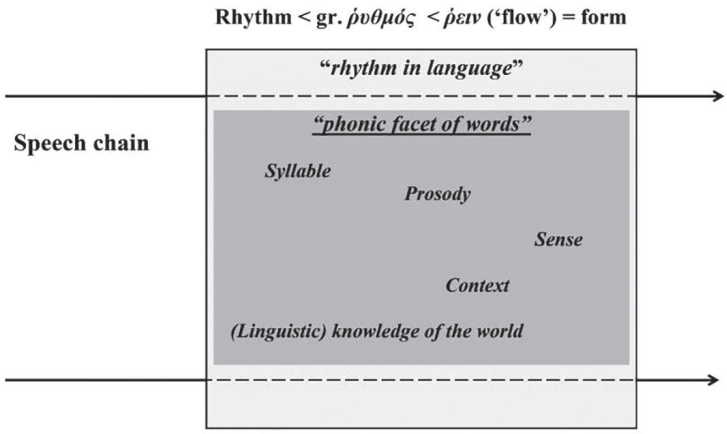

Figure 1. Albano Leoni's proposal of model of speech perception (2009) framed in Meschonnic's development (1982) of Benveniste's philological proposal (1966) of the original meaning of "rhythm"

4. 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 & 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.

4.1. The perception of "sound on screen"

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: 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) (Chion 1990). 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. Chion 1990: 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" (Chion 1990: 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: 12; emphasis in original).

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 (Chion 1990: 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. Chion 1990: 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 (Chion 1990: 5).[1] 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" (Chion 1990). 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" (Chion 1990: 63). Synchresis originates from a "point of synchronization", i.e. a moment when an auditory phenomenon and a visual one occur at the same time (Chion 1990:58). It is also "a function of meaning, and is organised according to gestaltist laws and contextual determinations" (Chion 1990: 63). Synchresis is the result of "dubbing, post-synchronization and sound-effects mixing" (Chion, 1990: 63).

In audio-vision, it is possible to distinguish three modes of listening (Chion 1990: 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. Reduced listening is the kind of listening "encouraged" by the "acousmatic situation" because it focuses attention on "sonic textures, masses and velocities" (Chion 1990: 32). A sound is "acousmatic" when the listener cannot see its source, or "originating cause" (Chion 1990: 71). Semantic listening, instead, is the kind of listening that prevails in cases of "audiovisual counterpoint", which occur

1. 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).

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 there is no connection between the images and the commentary (Chion 1990: 38-39).

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 sounds (cf. Chion 1990:73).

Thinking of an "audiologovisual poetics" (Chion 1990: 169-184), it is possible to "distinguish three modes of speech in film": "*theatrical speech*", "*textual speech*" and "*emanation speech*" (Chion 1990; italics in original). 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".

5. Voice as gesture: expression and perception

5.1. 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; my translation of quotes). It is evident that preverbal communication is still present in live speech, especially in prosodic elements of languages, mainly intonation and rhythm (cf. Fónagy 1983: 148). 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 (Fónagy 1983: 150; my translation of quotes).

5.2. Prosody

"Intensity or dynamic stress articulates and organises speech. It divides the flow of speech chain into sequences, into rhythmic groups" (Fónagy 1983: 107; my translation of quote). "Stressed syllables are the product of a special effort by phonatory […] and, especially, of a strong contraction of expiratory muscles, mainly intercostal and abdominal" (Fónagy 1983: 108; my translation of quote). 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 listener experiences the articulatory effort produced by the speaker" (Fónagy 1983). 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, physiological effort is inversely proportional to the intensity of the sound produced.

Emphatic stress is characterised 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. Fónagy 1983: 110). Strong and frequent emphatic stressing is a feature of angry and hate-filled attitudes (cf. Fónagy 1983: 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. Fónagy 1983: 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 both as 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) (Fónagy 1983). 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 (Fónagy 1983). 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. Fónagy 1983: 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 (Fónagy 1983). When the vibrations of tensed folds accelerate, then the tone rises, when fold relaxes and vibration decelerates, then tone decreases (Fónagy 1983). Prosodic elements of language, intonation and rhythmic patterns, have a gestural behaviour: they not only signify attitudes, but *are* the same attitudes, they convey these attitudes *per se* (cf. Fónagy 1983: 149).

5.3. Attitudes in voice

Fónagy (1983: 122-137; my translation of quotes) 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". For reasons of space, only few cases are reported, i.e. those considered relevant to the present study, especially for the construction of questionnaire (see § 6.1.2. and 6.2.); namely: tenderness, anger and aggressiveness, which is related to argumentative speeches.

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. Fónagy 1983: 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. Fónagy 1983: 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 (Fónagy 1983).

Aggressiveness 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 (cf. Fónagy 1983: 151).

Body attitude at the basis of explanation or logical operation is not very different from aggressiveness (cf. Fónagy 1983: 138). Both cases are characterised 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 (Fónagy 1983).

5.4. 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 listeners due to the "law of accommodation" (Fónagy 1983: 155; my translation of quote). Therefore, vocal gestures that form vocal style detach from emotional attitude to attach themselves to the speaker's personality, becoming personal messages (Fónagy 1983).

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 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 were 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).

QUESTIONNAIRE BIOGRAPHIQUE (Voix Féminines)

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

Texte	GM %	SS %
d) Fait l'école buissonnière en permanence, doit changer deux fois d'école	22 (31,4)	3 (4,5)
6. *a*) S'habille avec beaucoup de goût, elle est élégante, même dans des robes peu coûteuses	38 (50,0)	10 (11,9)
b) S'habille avec maladresse et peu de goût, elle n'est jamais élégante, même dans une robe relativement chère	2 (2,6)	44 (52,4)
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	32 (44,7)	6 (7,1)
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.....	2 (2,6)	24 (28,6)
7. *a*) Elle était une fille à maman	11 (16,2)	21 (30,4)
b) Elle était une fille à papa.....	41 (60,3)	12 (17,4)
c) Elle était orpheline	1 (1,5)	17 (24,6)
d) Ses parents ne s'occupaient pas d'elle	15 (22,1)	19 (27,3)
8. *a*) Elle était la sœur aînée	31 (52,5)	34 (54,8)
b) Elle était la cadette	28 (47,4)	28 (45,2)

Figure 3. Questionnaire on the vocal style of Gaby Morlay (G.M) and Simone Signoret (S.S.) (Fónagy 1983: 162-163) – part II

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; my translation of quotes) conducted a series of experiments called "jeu d'écho", used in experimental psychology, where subjects were asked to mimic utterances after a pause of a few seconds; they were recorded, analysed acoustically, 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" (Fónagy 1983: 251-253), the Hungarian expression "Az én hibám volt" ["etait-ce (ou: c'etait) 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 to 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 levels (cf. Fónagy 1983: 254-255).

Further analyses on a 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 (Fónagy 1983: 299; my translation of quotes) (an example is reported in the figures 4 and 5 – below).

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 (Fónagy 1983: 306). The evocative technique of vocal artists adds a third dimension, the melodic movement: "melodicity" – that adds to time (duration) and height (tone) (Fónagy 1983: 310; my translation of quotes – 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 regularité 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).

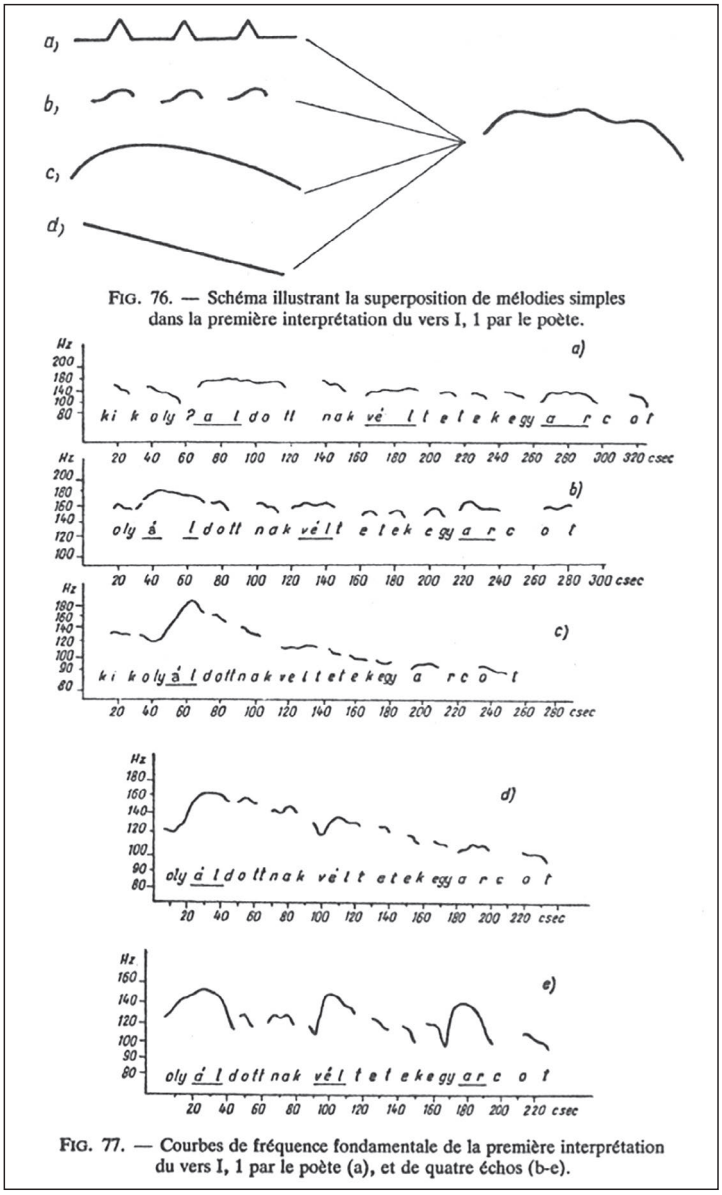

FIG. 76. — Schéma illustrant la superposition de mélodies simples dans la première interprétation du vers I, 1 par le poète.

FIG. 77. — Courbes de fréquence fondamentale de la première interprétation du vers I, 1 par le poète (a), et de quatre échos (b-e).

Figure 4. Melodic curves compared (Fónagy 1983: 294)

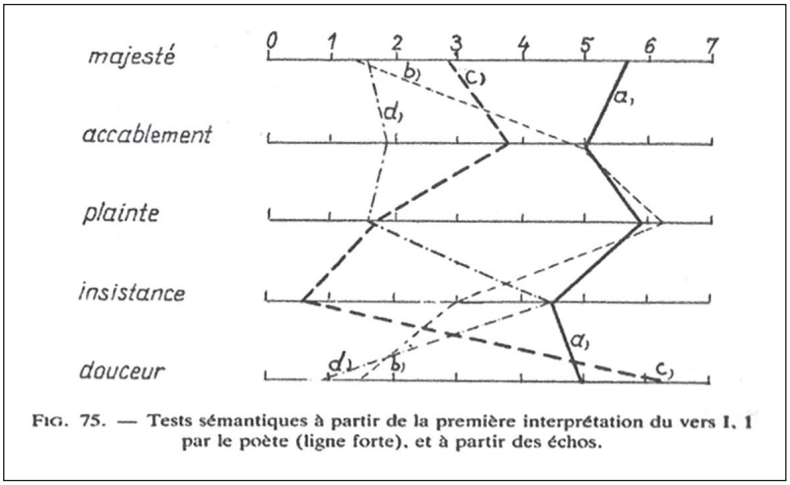

FIG. 75. — Tests sémantiques à partir de la première interprétation du vers I, 1 par le poète (ligne forte), et à partir des échos.

Figure 5. Semantic tests compared (Fónagy 1983: 293)

5.5. Poetry and vocal art

Fónagy's (1983: 316-321; my translation of quotes) discussion and conclusion 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 versions 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 poesie 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 recurrant aux moyens du langage primitif que poésie et art vocal se constituent en poiesis, en creation et action (Fónagy 1983: 320).

Poetry and vocal art share a "certain balance 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".

6. 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 through 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 fonic facet of words") (Albano Leoni 2009), *Audio-vision* (Chion 1990) and *La vive voix* (Fónagy 1983). Thus, a number of categories,[2] considered relevant to this study, were extracted from the above mentioned proposals. Items and categories of the questionnaire are reported in table 1 (see below).

6.1. Deep structure of the questionnaire: categories related to the sources of the theoretical paradigm

The traits of *Il volto fonico delle parole* were all taken as categories of reference related to this 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 category, 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 categories: "theatrical speech", "textual speech" and "emanation speech". As to the categories extracted from Fónagy's *La vive voix*, these were: "expressivity", "comprehensibility", "melodicity", "vocal attitude" and "vocal personality".

6.2. Superficial structure of the questionnaire: items related to the categories extracted from "The phonic facet of words" (Albano Leoni 2009)

With reference to the category "syllable" taken from "The phonic facet 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,

2. The word "category" means, in general, "whatever notion that can serve as a rule for an investigation or its linguistic expression, no matter the field". Plato, who first theorised categories, considered them "primarily as determinations of reality and, secondarily, as notions needed to investigate and understand reality". He defined categories as "supreme genres", namely: "being, movement, repose, identity, otherness". Plato also proposed that categories make it possible to establish a correspondence between reality and discourse. (Abbagnano 1993: 115, my translation of quotes).

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 ascribed to the category of "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 § 5.4). Actually, the item "same melody repeated" was an attempt to elicit the perception of "melodicity" that Fónagy detected in the artistic use of voice, but we assumed that a voice used in television is expected to have artistic features (Straniero Sergio 1997: 54). Under the category of "prosody", we ascribed the items "audible breaths", "silent pauses", "natural/non-natural syntax", "simple/complex sentences", "melodious/monotonous voice", "sweet/aggressive voice". Prosody is strictly linked to syntax, since in orality prosody conveys the meaning of noun groups, phrases, clauses and sentences. Silent pauses, be them 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. Now, we are well aware that intonation is just one component of voice, and that in many cases the main feature of this may be timbre, which depends on physical and physiological elements. But, again, for reasons of simplification of the language of the items, considering the subjects, the timbre was not even included in the aspects of voice (see below). The item "sweet/aggressive voice" is also a function of "melodious/monotonous voice" (see section 5.3), 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 5.2), and then by rhythm.

Under the category "voice" we ascribed items that, in our opinion, can be intuitively related to voice. We did not consider the distinction between breathing, timbre, voice, prosody, speech, etc., not only because such (theoretical) distinction may not be always accessible to non-expert subjects, but mainly because, in perception, such distinction may not be possible due to overlapping features (e.g. tone for voice and intonation; timbre for tone and intonation). We assume that, in perception, such distinction may not be possible simply because in production it is not; because breathing, phonation,

articulation, intonation are all physically, physiologically and psychologically interrelated; this interrelation also concerns text and context (see § 2 and 5). For these reasons, the items ascribed under the "voice" category 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 demonstrated by Fónagy (1983) (see § 5.3 and 5.4). The items "comprehensible voice" and "expressive voice" refer to the aesthetic aspects of artistic voice (see § 5.5). The item "credible voice" is mainly related to the impression created by the medium and the context. In fact, 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).

To identify the items to be ascribed under the category of "sense", we considered 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). Therefore, we considered it fundamental to ascribe under this category the items "comprehensible" and "expressive voice", "credible voice" and "comprehensible interpretation", as 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 category of "sense", 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"; but 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 ascribed to the category "context" 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 assigned to the category "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 many categories recur in many items. This argument is understandable, but such structure was inevitable, considering the theoretical approach used, which is based on the total perception of speech production, audiovision 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 § 2) and in perception (Fónagy 1983: 160-169; 1993; see also § 5.4); 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 § 4). In addition, there are the proposals of: i) rhythm as form by Benveniste (1966; see § 3); ii) rhythm as form of a language (Meschonnic 1982; see § 3); and iii) gestaltic perception of speech (Alban Leoni 2009; see § 3).

<table>
<tr><th colspan="3">DEEP STRUCTURE OF QUESTIONNAIRE</th><th>SUPERFICIAL STRUCTURE</th><th rowspan="3">QUESTIONNAIRE FLOW</th></tr>
<tr><th colspan="3">Categories for:</th><th rowspan="2">Quality criteria / Items</th></tr>
<tr><th>Il volto fonico delle parole</th><th>Audio-vision</th><th>La vive voix</th></tr>
<tr><td rowspan="3">Syllable</td><td>Synchresis (synchronised synthesis)
Textual speech
Theatrical speech
Emanation speech</td><td>Expressivity
Comprehensibility
Melodicity
Vocal attitude
Vocal personality</td><td>Articulation</td><td rowspan="15">1. articulation
(phono-syllabic scanning)
2. hesitations
3. audible breaths
4. silent pauses
5. speed of speech
6. melodious / monotonous voice
7. same melody repeated (sung speech)
8. sweet / aggressive voice
9. self-confident/ insecure voice
10. Active / self-defeating personality of the interpreter
11. (in)expressive voice
12. (un) comprehensible voice
13. interpretation-image synchrony
14. informativity of interpretation with respect to the image</td></tr>
<tr><td rowspan="2">Synchresis
Textual speech
Theatrical speech
Emanation speech</td><td rowspan="2">Expressivity
Comprehensibility
Vocal attitude
Vocal personality</td><td>Hesitations</td></tr>
<tr><td>Speed of speech</td></tr>
<tr><td>Synchresis
Textual speech</td><td>Expressivity
Comprehensibility
Melodicity
Vocal attitude
Vocal personality</td><td>Same melody repeated (sung speech)</td></tr>
<tr><td rowspan="5">Prosody</td><td>Synchresis
Textual speech</td><td rowspan="4">Expressivity
Comprehensibility
Vocal attitude
Vocal personality</td><td>Audible breaths</td></tr>
<tr><td rowspan="2">Synchresis
Textual speech
Theatrical speech</td><td>Silent pauses</td></tr>
<tr><td>Natural / non–natural syntax</td></tr>
<tr><td>Synchresis
Textual speech</td><td>Simple / complex sentences</td></tr>
<tr><td rowspan="2">Synchresis
Textual speech
Theatrical speech</td><td rowspan="2">Expressivity
Comprehensibility
Melodicity
Vocal attitude
Vocal personality</td><td>Melodious / monotonous voice</td></tr>
<tr><td>Sweet / aggressive voice</td></tr>
<tr><td rowspan="5">Voice</td><td>Synchresis
Textual speech
Theatrical speech
Emanation speech</td><td>Expressivity
Comprehensibility
Melodicity
Vocal attitude
Vocal personality</td><td>Articulation</td></tr>
<tr><td rowspan="2">Synchresis
Textual speech
Theatrical speech
Emanation speech</td><td rowspan="2">Expressivity
Comprehensibility
Vocal attitude
Vocal personality</td><td>Hesitations</td></tr>
<tr><td>Speed of speech</td></tr>
<tr><td>Synchresis
Textual speech
Theatrical speech</td><td rowspan="2">Expressivity
Comprehensibility
Melodicity
Vocal attitude
Vocal personality</td><td>Melodious / monotonous voice</td></tr>
<tr><td>Synchresis
Textual speech</td><td>Same melody repeated</td></tr>
</table>

	Synchresis *Textual speech* *Theatrical speech*		Sweet / aggressive voice	15. (in)credibile interpreter 16. (un)skilled interpreter 17. simple / complex words 18. non-natural / natural syntax 19. simple / complex sentences 20. Comprehensible interpretation (self-assessment of comprehension) 21. Real comprehension of text (multiple-choice test)
			Active / self-defeating personality of the interpreter	
	Synchresis *Textual speech*	*Comprehensibility*	Comprehensible voice	
		Expressivity	Expressive voice	
	Synchresis *Textual speech* *Theatrical speech*	*Expressivity* *Comprehensibility* *Melodicity* *Vocal attitude* *Vocal personality*	Credible voice	
Sense	*Synchresis* *Textual speech*	*Comprehensibility*	Comprehensible voice	
	Synchresis *Textual speech*	*Expressivity*	Expressive voice	
	Synchresis *Textual speech* *Theatrical speech*	*Expressivity* *Comprehensibility* *Melodicity* *Vocal attitude* *Vocal personality*	Credible voice	
			Comprehensible interpretation	
	Synchresis *Textual speech*	*Expressivity* *Comprehensibility*	Complex / simple words	
	Synchresis *Textual speech*	*Expressivity* *Comprehensibility*	Complex / simple sentences	
	Synchresis	*Vocal personality*	Active / self-defeating personality of the interpreter	
Context	*Synchresis* *Textual speech* *Theatrical speech* *Emanation speech*	*Expressivity* *Comprehensibility* *Melodicity* *Vocal attitude* *Vocal personality*	Melodious / monotonous voice	
			Same melody repeated	
	Synchresis *Textual speech* *Theatrical speech*		Comprehensible interpretation	
	Synchresis *Textual speech*	*Expressivity*	Expressive voice	
	Synchresis *Textual speech* *Theatrical speech*	*Expressivity* *Comprehensibility* *Melodicity* *Vocal attitude* *Vocal personality*	Credible voice	
	Synchresis		Sound-image Synchrony	
	Synchresis *Textual speech* *Theatrical speech*		Informativity of interpretation with respect to the image	

<table>
<tr><td></td><td>Synchresis
Textual speech
Theatrical speech</td><td>Expressivity
Comprehensibility
Melodicity
Vocal attitude
Vocal personality</td><td>Skilled / unskilled interpreter</td><td></td></tr>
<tr><td rowspan="4">(Linguistic) knowledge of world</td><td rowspan="2">Synchresis
Textual speech
Theatrical speech</td><td rowspan="2">Expressivity
Comprehensibility</td><td>Complex / simple words</td><td rowspan="4"></td></tr>
<tr><td>Complex / simple sentences</td></tr>
<tr><td>Synchresis
Textual speech
Theatrical speech</td><td rowspan="2">Expressivity
Comprehensibility
Melodicity
Vocal attitude
Vocal personality</td><td>Self-assessment of comprehension</td></tr>
<tr><td>Synchresis
Textual speech
Theatrical speech
Emanation speech</td><td>Real comprehension of text</td></tr>
</table>

Table 1. Proposal for a questionnaire to elicit a gestaltic assessment of voice quality in TV broadcast simultaneous interpreting

7. Application of the questionnaire

The questionnaire was designed for a quality interpreting assessment of Italian interpretations of the 2012 US Presidential Debate, within the author's PhD research study. The televised interpretations of US Presidential Debates constitute a sub-corpus of the Italian Corpus of Television Interpreting (*CorIT*), created and developed at the University of Trieste (cf. Dal Fovo 2013). The questionnaire was built to be administered, through a web-based platform, to professional interpreters, television experts, television viewers, actors and musicians. A pilot survey has already been carried out, having BA and MA translation and interpreting students as respondents (101 subjects); in this case the questionnaire was administered *in praesentia*. This survey included three excerpts of Italian interpretations of 2008 US Presidential Debates, and not of 2012 Debates, because when the material for the pilot study was prepared, the sub-corpus of Italian (and Spanish) interpretations of 2012 Debates had not been completely transcribed yet. The pilot survey also included an experimental variable: one of the three video excerpts was manipulated by replacing the original Italian interpretation with an imitation of it, executed by a professional TV dubber, working mainly as narrator of TV documentaries. The dubbing-actor listened to the original interpretation first, and then performed an imitation of it by reading the original transcript (where original disfluencies were also indicated) and listening to the original English speaker on earphones while producing his own imitation. The aim of such an experimental variable was to isolate the prosodic and vocal aspects of the

interpretation (in line with previous experiments – see above, § 2); in order to elicit the effect, in perception (and consequently, in assessment) of the performance of an "artistic voice" (Fónagy 1983 – see above, § 5.4 and 5.5), a voice of a dubber with an experience of use of voice for television.

The questionnaire for the pilot study was developed according to the linguistic approach described above and following the methods of social research (Bailey 1995:103-207; Gillham 2000: 1-45, 395-424; Blanke & al. 2006: 1-57). The "questionnaire flow" (Blanke & al. 2006: 47-48; Bailey 1995: 163-170) was built in such a way to help respondents to move from sound or phonic perception to sense construction, and finally to self-assessment of the subject's comprehension; it ends with a comprehension test.

As to "questions formats" (Blanke & al. 2006: 35-43), the majority of questions were "closed" questions. Questions on comprehension were "multiple choice"; those on possible aspects not considered in the questions and on comments on the questionnaire were "open" questions; while final questions on personal data were all "factual" questions. All closed questions were "scaled-response" questions, with "numeric (endpoint-labelled) scales" (Blanke & al. 2006: 40-41).

The survey included three video excerpts (one min each) and three sets of the same questions for each video excerpt; therefore each questionnaire was made up of 3 blocks, each block being constituted by one video plus one set of questions, and a final block of questions on personal data. Sets of questions were the same, except for the last three questions of each set, which were related to the comprehension of the relative interpretation.

The questionnaire flow was developed in seven stages. Particular attention was paid to the wording and phrasing of questions. Considering that not all subjects for which the questionnaire was built (see above) were experts in linguistics, technical terms (e.g. prosody, intonation, tone, singsong, syntax, terminology, etc.) were avoided and replaced by common words referring to the same aspect (e.g. same melody repeated, melodious voice, words, sentences, etc.). In addition, to reduce the circularity of declarative questions, sometimes with positive attributes (e.g. *high* speed of speech; *pleasant* tone variation; etc.), that could have had an excessive influence on respondents, these were restructured in a more neutral way 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 (see below – figure 6).

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) permitted an evaluation pattern that is a numeric endpoint-labelled scale, but the scale is hidden to the respondent, thus the graphic solution results in a cursor initially positioned in the middle that can be moved by the subject towards the two labelled endpoints. The hidden scale is 1 to 7 points; all the "negative" labels of the questions were placed on the left, corresponding to point 1 of the scale. At the end, the questionnaire was administered *in praesentia*, and no longer on-line, due to technical reasons related to the video excerpts. Nevertheless, the graphic trick of the hidden scale was transferred to the paper; respondents, in fact, were asked to indicate their choice with a sign on a segment (or intensity line) (see below – figure 6) having the labels as endpoints; the number on the scale was found by the researcher, when he measured it with the scale printed on transparent paper having the same length of the segment. The web based platform also allowed a randomized administration of the three blocks linked to the video excerpts, in order to avoid the response set phenomenon. In the *in praesentia* administration the randomization was created by distributing the possible combinations of the 3 videos (in sequence) to groups of respondents as much as possible equal in number.

Results and discussion of the pilot survey will be matter of a future paper.

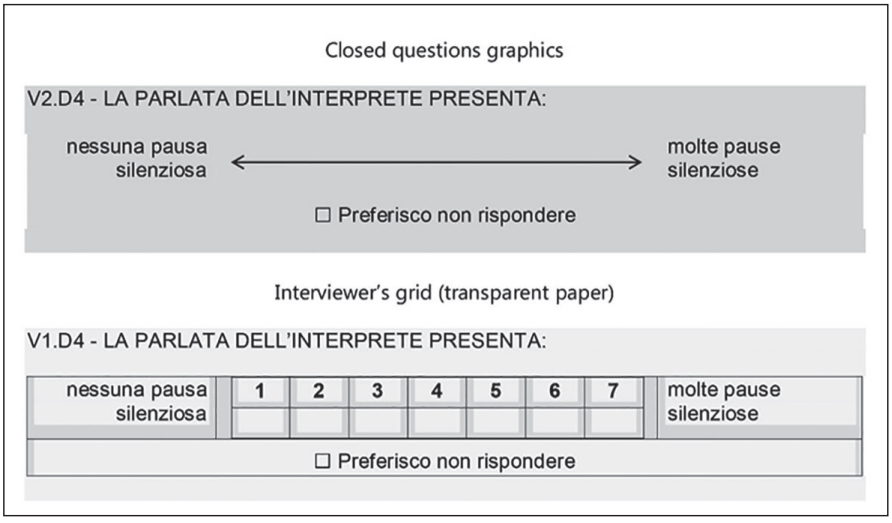

Figure 6. Example of a closed question format and the relative assessment pattern

References

ABBAGNANO, Nicola. (1993) *Dizionario di filosofia*. Milano: TEA.

AHRENS, Barbara. (2004) "Non-verbal phenomena in simultaneous interpreting: causes and functions." In: Hansen, Gyde; Kirsten Malmkjær & Daniel Gile (eds.) 2004. *Claims, Changes and Challenges in Translation Studies*. Amsterdam & Philadelphia: John Benjamins, pp. 227-237.

AHRENS, Barbara. (2005) "Prosodic phenomena in simultaneous interpreting: A conceptual approach and its practical application." *Interpreting* 7:1, pp. 51-76.

ALBANO LEONI, Federico. (2009) *Dei suoni e dei sensi. Il volto fonico delle parole*. Bologna: Il Mulino.

BAILEY, Kenneth D. (1995) *Methods of Social Research*. New York: The Free Press. Italian translation by Maurizio Rossi: *Metodi della ricerca sociale*. Bologna: Il Mulino, 1995.

BENVENISTE, Émile. (1966) *Problèmes de linguistique générale*. Paris: Gallimard.

BLANKE, Karen; Giovanna Brancato; Jürgen H. P. Hoffmeyer-Zlotnik; Thomas Körner; P. Lima; Stefania Macchia; Manuela Murgia; Anja Nimmergut; R. Paulino; Marina Signore & Giorgia Simeoni (2006) *Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System*. Electronic version: <http://www.istat.it/it/files/2013/12/Handbook_questionnaire_development_2006.pdf>

BÜHLER, Hildegund. (1986) "Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters." *Multilingua* 5:4, pp. 231-235.

BÜHLER, Karl. (1934) *Sprachtheorie. Die Darstellungsfunktion der Sprache*. Jena: Gustav Fisher. Italian translation by Serena Cattaruzza Derossi: *Teoria del linguaggio. La funzione rappresentativa del linguaggio*. Roma: Armando, 1983.

CATANA, Cinzia. (2005) *Le qualità ben pronunciate: la dizione dell'interprete e la percezione dell'utente in simultanea*. Università di Bologna, sede di Forlì: SSLMIT. Unpublished Master's Degree dissertation.

CHIARO, Delia & Giuseppe Nocella. (2004) "Interpreters" Perception of Linguistic and Non-Linguistic Factors Affecting Quality: A Survey through the World Wide Web." *Meta* 49:2, pp. 278-293.

CHION, Michel. (1990) *L'Audio-Vision*. Paris: Editions Nathan. English translation by Claudia Gorbman: *Audio-vision: sound on screen*. New York: Columbia University Press, 1994.

CHRISTODOULIDES, George & Cédric Lenglet. (2014) "Prosodic correlates of perceived quality and fluency in simultaneous interpreting." *Proceedings of the Speech Prosody 7* Conference, Dublin, 20-23 May 2014, pp. 1002-1006.

CHRISTODOULIDES, George. (2013) "Prosodic features of simultaneous interpreting." *Proceedings of the Prosody-Discourse Interface Conference 2013 (IDP-2013)*, Leuven, 11-13 September 2013, pp. 33-37.

COLLADOS AÍS, Ángela; Esperanza Macarena Pradas Macías; Elisabeth Stévaux & Olalla García Becerra (eds.) (2007) *La evaluación de la calidad en interpretación simultánea: parámetros de incidencia*. Granada: Comares.

COLLADOS AÍS, Ángela. (2002) "Quality assessment in simultaneous interpreting: The importance of nonverbal communication." In: Pöchhacker, Franz & Miriam Shlesinger (eds.). *The Interpreting Studies Reader*. London & New York: Routledge, pp. 326-335.

DAL FOVO, Eugenia (2013) "Il progetto CorIT: corpus e prospettive di ricerca." *RITT* (*Rivista internazionale di tecnica della traduzione*) 15, pp. 45-62.

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." *RITT* (*Rivista internazionale di tecnica della traduzione*) 16, pp. 57-87.

FÓNAGY, Ivan. (1983) *La vive voix*. Paris: Payot.

FÓNAGY, Ivan. (1993) "Il significato dello stile vocale." *Phoné Semantiké*, special issue of *Il Verri*, IX:1-2 (maggio-giugno), pp. 7-29.

FÓNAGY, Ivan. (2001) *Languages within language: an evolutive approach*. Amsterdam & Philadelphia: John Benjamins.

GARCÍA BECERRA, Olalla. (2013) *La incidencia de las primeras impresiones en la evaluación de la calidad de la interpretación simultánea: un estudio empírico*. Universidad de Granada. PhD dissertation.

GARZONE, Giuliana. (2003) "Reliability of quality criteria evaluation in survey research." In: Collados Aís, Ángela; María Manuela Fernández Sánchez & Daniel Gile (eds.). *La evaluación de la calidad en interpretación: investigación*. Granada: Comares, pp. 23-30.

GILE, Daniel. (1990) "L'évaluation de la qualité de l'interprétation par les délégués: une étude de cas." *The Interpreter's Newsletter* 3, pp. 66-71.

GILLHAM, B. (2000) *Developing a Questionnaire*. London: Continuum.

GOLDMAN-EISLER, Frieda. (1968) *Psycholinguistics. Experiments in spontaneous speech*. London & New York: Academic Press.

IGLESIAS FERNÁNDEZ, Emilia. (2013) "Unpacking Delivery Criteria in Interpreting Quality Assessment." In: Van Deemter, Roelof & Dina Tsagari (eds.) 2013. *Assessment Issues in Language Translation and Interpreting*. Frankfurt: Peter Lang, pp. 51-66.

KOPCZYŃSKI, Andrzej. (1994) "Quality in conference interpreting: some pragmatic problems." In: Snell-Hornby, Mary; Franz Pöchhacker & Klaus Kaindl (eds.) 1994. *Translation Studies: An Interdiscipline*. Amsterdam & Philadelphia: John Benjamins, pp. 189-198.

KURZ, Ingrid & Franz Pöchhacker. (1995) "Quality in TV Interpreting." *Traslatio-Nouvelles de la FIT- FIT Newsletter* 14:3-4, pp. 350-358.

KURZ, Ingrid. (1993) "Conference interpretation: expectations of different user groups." *The Interpreter's Newsletter* 5, pp. 13-21.

MACK, Gabriele & Lorella Cattaruzza. (1995) "User surveys in SI: a means of learning about quality and/or raising some reasonable doubts." In: Tommola, Jorma (ed.) 1995. *Topics in Interpreting Research*. Turku: Centre for Translation and Interpreting, University of Turku, pp. 37-49.

MACK, Gabriele. (1999) "L'interpretazione in TV: vecchie e nuove ipotesi di ricerca." Paper presented at the founding conference of the Associazione Italiana di Linguistica Applicata (AItLA), Pisa, October 22-23, 2009. Electronic version: <www.sslmit.unibo.it/aitla/pisa/papers/html>

MARRONE, Stefano. (1993) "Quality: a shared objective." *The Interpreter's Newsletter* 5, pp. 35-41.

MEAK, Lidia. (1990) "Interprétation simultanée et congrès medical: attentes et commentaires." *The Interpreter's Newsletter* 3, pp. 8-13.

MESCHONNIC, Henri. (1982) *Critique du rythme. Antropologie historique du langage*. Lagrasse: Verdier.

MOSER, Peter. (1995) "Survey on expectations of users of conference interpretation. Final report commissioned by AIIC." Electronic version: <file:///C:/Users/HP%206730S/Downloads/Full-report-in-PDF%20(2).pdf>

NIKOLAJ, Trubeckoj. (1929) "Zur allgemeinen Theorie der phonologischen Vokalsysteme." In: *Travaux du Cercle Linguistique de Prague* Vol. I, pp. 39-66.

PÖCHHACKER, Franz & Cornelia Zwischenberger. (2010) "Survey on quality and role: conference interpreters' expectations and self-perceptions." Electronic version: <http://aiic.net/page/3405/survey-on-quality-and-role-conference-interpreters-expectations-and-self-perceptions/lang/1>

PRADAS MACÍAS, Esperanza Macarena. (2006) "Probing quality criteria in simultaneous interpreting: The role of silent pauses in fluency." *Interpreting* 8:1, pp. 25-43.

RUSSO, Mariachiara. (2005) "Simultaneous film interpreting and users' feedback." *Interpreting* 7:1, pp. 1-26.

SHLESINGER, Miriam. (1994) "Intonation in the Production and Perception of Simultaneous Interpretation." In: Lambert, Sylvie & Barbara Moser-Mercer (eds.) 1994. *Bridging the Gap: Empirical Research in Simultaneous Interpretation*. Amsterdam & Philadelphia: John Benjamins, pp. 225-236.

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*. Barcelona: Universitat Pompeu Fabra. PhD dissertation.

STRANIERO SERGIO, Francesco. (2007) *Talkshow Interpreting. La mediazione linguistica nella conversazione spettacolo*. Trieste: Edizioni Università di Trieste.

TOHYAMA, Hitomi & Shigeki Matsubara. (2006) "Influence of pause length on listeners' impressions in simultaneous interpretation." Proceedings of the *Ninth International Conference on Spoken Language Processing* (*Interspeech 2006 – ICSLP)*, Pittsburgh, Pennsylvania, 17-21 September, pp. 893-896.

VILCHES, Lorenzo (1989). *Manipulación de la información televisiva*. Barcelona: Paidós.

VUORIKOSKI, Anna Riitta. (1993) "Simultaneous interpretation – user experience and expectations." In: Picken, Catriona (eds.) 1993. *Translation - The Vital Link. XIII FIT World Congress, Proceedings*. Brighton: Institute of Translation and Interpreting, pp. 317-327.

WILLIAMS, Sarah. (1995) "Observations on anomalous stress in interpreting." *The Translator* 1:1, pp. 47-64.

BIONOTE / NOTA BIOGRAFICA

GREGORIO DE GREGORIS has an MA degree in Translation and Interpretation from the University of Bologna, where he attended the Conference Interpreting study program (Italian, Spanish and English) and wrote his final degree dissertation on "The Body and The Voice in Orality". From 2003 to 2006, he participated in the activities of the Center of Theatre Studies of the Department of Interdisciplinary Studies in Translation, Languages and Cultures (*SITLeC*) of the University of Bologna. Since 2009 he is a free-lance translator. In January 2012 he enrolled in the PhD program in Interpreting and Translation Studies at the University of Trieste.

GREGORIO DE GREGORIS è laureato all'Università di Bologna (SSLMIT, Forlì) in Traduzione e Interpretazione, indirizzo Interpretazione di Conferenza, con una tesi su "Il corpo e la voce nell'oralità". Ha partecipato alle attività del Centro di Studi Teatrali del Dipartimento di Studi Interdisciplinari su Traduzione, Lingue e Culture (SITLeC) dell'Università di Bologna tra il 2003 e il 2006. Dal 2009 è traduttore free-lance. Dal gennaio 2012 è iscritto alla Scuola Dottorale in Studi Umanistici dell'Università di Trieste, indirizzo Studi in Interpretazione.