Significances (Signifying), 3(1), 98-113.
DOI : https://doi.org/10.18145/significances.v3i1.234

Sound symbolism, speech expressivity and crossmodality

Sandra Madureira, Mario Augusto de Souza Fontes, Zuleica Camargo[1]

Abstract

The basis of speech expressivity is sound symbolism, since the meaning effects attributed to the spoken mode by the listeners rest on the acoustic features deriving from the various articulatory maneuvers yielding breath, voice, noise, resonance and silence. Based on the impression caused by these features, listeners attribute physiological, physical, psychological and social characteristics to speakers. Among the prosodic elements, vocal quality settings have received the least attention regarding speech expressivity uses. In this paper, we introduce an innovative proposal for analyzing the expressive uses of voice quality settings. It concerns the grouping of the voice quality settings according to their shared acoustic output properties and vocal tract configurations so that correlations among the production, the acoustic and the semantic levels can be investigated and analogies between the speech phonic matter and meaning effects can be established.

Keywords: sound symbolism; speech expressivity; voice quality; acoustic analysis; perceptual analysis.

Résumé

L'expression de la parole est basée sur le symbolisme du son, car les effets de signification attribués au mode parlé par les auditeurs reposent sur les caractéristiques acoustiques découlant des diverses manœuvres articulatoires produisant respiration, voix, bruit, résonance et silence. Sur la base de l'impression provoquée par ces caractéristiques, les auditeurs attribuent des caractéristiques physiologiques, physiques, psychologiques et sociales aux locuteurs. Parmi les éléments prosodiques, les réglages de qualité vocale ont reçu le moins d'attention en ce qui concerne les utilisations de l'expressivité de la parole. Dans cet article, nous présentons une proposition innovante d'analyse des utilisations expressives des paramètres de qualité vocale. Il s'agit du regroupement des paramètres de qualité vocale en fonction de leurs propriétés de sortie acoustique et de leurs configurations du système vocal partagées, de sorte que les corrélations entre les niveaux de production, acoustique et sémantique puissent être étudiées et que des analogies entre la matière phonique et les effets de signification puissent être établies.

Mots-clés: symbolisme sonore; expressivité de la parole; qualité de la voix; analyse acoustique; analyse perceptuelle.

[1] Pontifical Catholic University of São Paulo, Brazil.

1. Introduction

Sound and meaning relations in speech communication have been debated for centuries since Plato discussed them in the Cratylus dialogue (Vieira, 2014). The controversial debate reported in this dialogue raised questions about the resemblance between form and meaning, contrasting conventional and nonconventional links between sound and meaning.

In Nobile (2011) the philosophical evolution of this theoretical debate from Plato´s to Humbolt´s views is commented. He provides arguments in favor of considering the sound systems of language as mirror images of its meaning systems. Other important contributions to this debate are found in the linguistic literature. These concern issues related to the following aspects: the sound shape of language integrating meaning and sound (Jakobson, 1978; Jakobson and Waugh, 1979); the specific phonetic features of speech sounds and their perceptual effects (Ladefoged, 1980); the expressive mode of speech (Tsur, 1979); the communicative and informative role of speech (Bolinger, 1986); the bidirectionality of the meaning to sound relation (Albano, 1988); the metaphorical view of speech production (Fónagy, 1983, 2001); the functional roles of form-to-meaning correspondences in language coding, decoding and communication (Dingemanse *et al*, 2015).

New insights on the motivation of the relations between sound and meaning have emerged thanks mainly to contributions from experimental studies (Xu and Chuenwattanapranithi, 2007; Chuenwattanapranithi et al, 2008; Xu, Kelly and Smillie, 2013; Noble and Xu (2011); Liu and Xu, 2014) which brought about evidence of sound symbolism, the direct links between sound and meaning (Hinton, Nichols, and Ohala 1994; Ohala 1997; Abelin, 1999).

Contini (2009) considers three types of motivations for the relations between sound and meaning: onomatopeic, phonosymbolic and iconic motivations. He points out that the phonosymbolic motivation is distinct from the others because speech sound characteristics convey semantic information and evoke extra-acoustic realities from a symbolic perspective.

A large amount of evidence in favor of sound symbolism has derived from the knowledge of the acoustic phonetic details of phonic segments and prosodic elements. According to the acoustic theory of speech production (Fant, 1970), speech is the acoustic output of voice and/or transient and continuous noise sources which resonate on the vocal tract. These voice and noise sources can characterize speech segments and phonation types as well (Barbosa and Madureira, 2015).

Speech segments can be classified into obstruents and resonants. Obstruent segments, which include the pulmonic sounds (stops, fricatives and affricates) and the non-pulmonic sounds (clicks, ejectives and implosives), are characterized by the presence of noise, while resonants, which include liquids, nasals, vowels and semivowels, are characterized by the subdivision of the vocal tract which creates cavities for the resonance of the voice source. There is also silence which makes part of stops. The silence is the acoustic output of the total obstruction produced by the articulators into contact. Breath, voice, noise, silence, are the segment-shaping features which are perceived by listeners and make impressions on them in diverse ways. Voice resonating in a non-obstructed cavity produces melodious sounds while a tightly constrained cavity generates turbulent noise and these characteristics affect the way pleasantness is evaluated by listeners.

Prosodic elements include stress, accent, intonation, rhythm, speech rate, pause and voice quality. Both prosodic elements and speech segments are used to express meaning effects in speech. Among these, the one which has received the least attention in the phonetic literature on speech expressivity is voice quality.

Voice quality is used to convey linguistic, paralinguistic and extralinguistic meanings. Creaky voice, for example, can be used to convey a linguistic meaning (the end of an utterance), a paralinguistic meaning (exasperation and an extralinguistic meaning (indicating the social group or the region the speaker belongs, for instance).

Two main concepts of voice quality have been defined in the phonetic literature: voice quality as some type of phonation (Gordon and Ladefoged, 2001); voice quality as the product of phonatory, articulatory and tense settings (Laver, 1980). The latter concept is the one adopted in this work because what characterizes someone´s voice is not only the product of the vibration mode of the vocal folds. A speaker may keep the tongue body in an advanced position while speaking and that setting in combination or not with other settings characterizes his/her voice. As some of the vocal quality settings involve movements of the lips and the jaw and these movements are also described as Action Unities (AUs) in facial expression studies, two descriptive systems are considered: the Voice Profile Analysis Scheme, known as VPAS, (Laver and Mackenzie-Beck, 2007) which describes vocal quality settings, and the Facial Analysis Coding System, known as FACS, (Ekman, Friesen and Hagen, 2002), which describes facial movements, As Poyatos (1991:181) states they are "proof of the audible visual nature of speech".

Sound symbolism is taken to be the basis of speech expressivity because the meaning effects attributed to the spoken mode by the listeners are thought to be based on the acoustic features of sounds deriving from the various articulatory maneuvers yielding the aforementioned features "breath", "voice", "noise", "resonance" and "silence".

Concerning the links between sound and meaning , arbitrary, non arbitrary and motivated reasons have been debated in a good number of scholars´papers. As Svantesson (2017) points out, the literature on sound symbolism keeps growing and evidences concerning universal features have been found in works such as Blasi *et al* (2016).

Two insightful concepts which have been raised in the linguistic literature to explain the sound-meaning links are: mirror images (Lucas Nobile, 2011), sound metaphors (Fonagy (1983). The first emphasizes the iconic aspects of the link between sound and meaning, and the latter, the analogies between meaning and speech sound production characteristics.

Furthermore, relevant evidences for sound-meaning links are found in the symbolic codes characterizing speech usage. The frequency code (Ohala, 1983) refers to the fact that high fundamental frequency values are correlated with meanings such as fragility, submission, amenableness and docility while low f0 values are associated with powerfulness, dominance, hostility and agressivity. This code relies upon an evolutionary basis, the survival of the species (Morton, 1977), since larger and stronger animals produce sounds which are lower in frequency and higher in loudness.

Experimental evidence of the frequency code is found in Chuenwattanapranithi *et al* (2008). In their experiment, synthesized vowels with a 3D articulatory synthesizer with parameters derived from the frequency code were used.

The effort code (Gussenhoven, 2002) concerns articulatory effort and it implies that the higher the articulatory effort,the more precise/tense the articulationbecomes. Potential meaning effects related to higher articulatory effort are "strength", "determination" and "tenseness", 'while those related to lower articulatory effort convey lower articulatory effort "relaxation", "laziness" and "disinterest".

The same can be said in relation to phonation. Fonagy (2001) states that sounds produced with laryngeal tension present glottal irregularity and tend to be associated to negative states,

while regular patterns of vocal fold vibration are associated with musicality and tend to be associated with positive affective states.

The production or respiratory code (Gussenhove, 2002) concerns subglottal air pressure. At the end of utterances subglottal air pressure declines and at its beginning it rises. Some potential meaning effects related to declining subglottal pressure are "finality", "certainty" "scantiness" while potential meanings related to increasing subglottal pressure are "expectancy", "continuity" and "arousal" (Madureira, 2011).

The sirenic code (Gussenhoven, 2016) concerns phonation and it refers to whispery voice productions. Due to anatomical conditions women´s voices are commonly characterized by whisperiness. In the production of whispery voices, the vocal folds are slack and weakly adducted so the air escapes and causes glottal friction. Some potential meaning effects of this kind of phonation is "feminine", "good-looking", "timid", "friendly", and "immature".

The aforementioned codes are related to the acoustic parameters of frequency and intensity. Time is implicit since higher frequency as opposed to lower frequency has to do with a large amount of instances in the same temporal interval. Peaks of frequency and intensity can occur more or less frequently in time and this affects the way listeners atribute meaning to utterances. A huge amount of evidence of expressive uses of speech rate is found in the phonetic literature (for example, length of pauses and breathing distinguishing between emotions (Beller *et al*, 2006), slower articulation rates have been correlated with power and assertiveness (Duez, 1997) while rapid articulation rates was found to characterize young people´s speech (Jacewicz *et al*, 2009)). These findings about the relations between timing features and meaning effects suggest that speech rate acts as an effective symbolic code.

Based on the impression caused by the acoustic features, listeners attribute physiological, physical, psychological and social characteristics to speakers. In this way, speech can be considered both expressive and impressive, because it is used to convey meaning effects but it also makes an impression on listeners.

In this work, we argue that the investigation of the expressive uses of voice quality settings can be better approached if these settings are grouped according to their shared output acoustic properties. In such a way, their expressive usage can be interpreted in terms of the relation to the four sound symbolic codes aforementioned, which attribute iconic meanings to the acoustic parameters, and can also be viewed from a metaphorical point of view, since analogies alluding to the vocal tract configurations which characterize vocal quality settings and their acoustic outputs can be made. The kind of approach proposed in this work is intended to clarify the expressive power of voice qualities.

We also argue that the investigation of the expressive uses of voice quality settings must follow certain rigorous methodological criteria. Based on the results of experiments we have carried out on the expressive uses of voice quality methodological procedures involving description, measurement, tools and statistic tests have been developed.

Expressive uses of vocal qualities from a crossmodal perspective, that is, correlating among semantic, acoustic and visual features by means of multidimensional analysis, are reported and the expressive and impressive roles of vocal quality settings in spoken communication are discussed in relation to motivated links between sound forms and meaning effects.

On the following sections we summarize some of the findings and expose the material and the method we have applied to analyze expressive uses of voice quality settings.

2. The descriptions of vocal quality settings

The vocal quality setting, a long-term muscular adjustment of the vocal apparatus, is an analytical unit proposed by Laver (1980, 2000) and which is part of the Voice Profile Analysis Scheme (VPAS) as developed by Laver *et al* (1981) and Laver and Mackenzie-Beck (2007). It refers to articulatory, phonatory, tense and prosodic features which characterize a person´s voice. It is used for describing voice qualities from a perception point of view and it is rated according to a 6-point Likert scale (See Appendix 1).

The term setting was coined by Honikman (1964) and applied by Laver to the description of types of voice qualities. The phonetic model of description of voice qualities developed by Laver (1980) presents a lot of advantages over other descriptive models: it considers the influence of factors other than phonation features in the building up of vocal profiles; it is componential and that allows the consideration of components individually and in combination with others; settings are described from physiological, perceptual and acoustic perspectives.

The characteristics constituting Laver´s model make it very adequate to consider expressive usages of voice qualities which have been reported in the phonetic literature as paralinguistic qualifiers (Poyatos, 1991), as pragmatic signaler (Kreiman and Sidtis, 2011), as indices of personality (Sapir, 1927; Scherer, 1979), affective states(Gobl and Ní Chasaide 2003; Johnstone and Scherer, 1999, 2000; Laver 1980; Scherer, 2005; Scherer *et al*, 2015; Tolkmitt and Scherer, 1986; van Bezooijen, 1984), speakers´ degree of nativeness (Piske, MacKay, and Flege, 2001), and physical characteristics (Kreiman and Sidtis, 2011).

Details concerning the application of VPAS, the principles governing its settings, the key segments to judge the settings and training in its use are found in Mackenzie-Beck (2005), Camargo and Madureira (2008) and Camargo, Madureira and Rusilo (2015).

Vocal and visual cues help identifying voice quality settings. Visual characteristics play a relevant role in speech perception and that is due to the fact that the brain motor areas responsible for speech production are also involved in body expressions. The 7th cranial nerve, innervating the facial muscles, is also connected to the brain motor areas responsible for speech production. Vocal and visual gestures are integrated as evidenced by the experiments on the so called McGurk effect (McGurk and MacDonald, 1976).

As some of the settings in VPAS involve face and neck movements (lip spreading, rounded and protuded lips, close jaw, open jaw, labiodentalization, advanced tip tongue position, lowered larynx and raised larynx), their relation to the facial action movement unities are worth considering in investigating expressive uses of voice quality settings.

3. The description of facial action unities

Facial action unities (AUs) are the analytical unities of the Facial Action Coding System (FACS). They can be categorized in terms of lower to higher intensities (A, B, C, D and E) and laterality (left or right). FACS is a perceptual componential protocol developed by Ekman and Friesen (1976) and Ekman, Friesen and Hagen (2002) to describe controlled facial movements and uncontrolled millisecond-lasting micro expressions. Sixty four AUs are listed in the 2002 version of the protocol.

Both FACS and VPAS are relevant tools to be applied in the analysis of speech expressivity because they provide descriptive analytical unities which can be correlated with other aspects,

as for instance, AU12 which refers to lip corner puller, corresponds to voice quality setting lip spreading and both of them are found in expressions of joy.

At present, the analysis of the AUs can be made automatically in systems which use facial expressions to identify emotions. Their level of accuracy reported in the literature is high and advances in the automatic recognition of facial expression has been acknowledged (Martinez, Valstar and Pantic, 2017). Systems which integrate facial expression analysis with vocal expression analysis have also appeared (Tian, Kanade and Kohn, 2011) and that must improve non-verbal expression recognition.

Both FACS and VPAS have been incorporated in the works which yielded the methodological procedures for the analysis of the expressive uses of voice qualities reported in this work. In the following the works which yielded the materials and methods to analyze the expressive uses of voice qualities are presented.

Similarities between FACS and VPAS are described in Madureira and Fontes (in press). Their analytical unities refer to muscular adjustments viewed in reference to a neutral setting or an action unity. Their characteristics in terms of spatiality, directionality, temporality, regularity and irregularity affect the way listeners and viewers perceive speech expressivity. In fact, speech expressivity profiles should include vocal and visual features (Ohala, 1980).

4. The path yielding to a methodological procedure proposal for the analysis of expressive uses of vocal quality settings

In this section, conceptual and methodological issues concerning our investigation on speech expressivity are presented, since they can help to demonstrate the foundations of the proposal introduced in this paper.

In Madureira (2008) the use of VPAS and a semantic differential scale (Osgood, Suci and Tannenbaum, 1957) enabled describing the voice quality settings used by an actor and an actress to recite a same poem and evaluating a group of listeners´ perception of meaning effects. The actor's performance was judged enthusiastic and the actress' performance as sad and anguished. The impressive effects on the listeners were interpreted as stemming from prosodic differences, mainly voice quality.

Madureira and Camargo (2010) and Madureira (2011) used the VPAS, acoustic analysis, the sound symbolism typology as proposed by Hinton Nichols, and Ohala (1994), and a one-way ANOVA test to analyze specific uses of sound symbolism. The use of synesthesic, imitative, and metalinguistic types of sound symbolism were identified, the latter related to specific uses of voice quality settings to express affective states.

Fontes and Madureira (2015) analyzed the expression of emotions in expressive speech by analyzing facial, vocal and semantic features.To perform the acoustic measures, the ExpressionEvaluator (Barbosa, 2019), designed to extract automatically acoustic measures relevant to analyze speech expressivity was applied.

Besides the ExpressionEvaluator script, other tools and methods were used: a profile to analyze facial expressions; the Gtrace to analyze the expression of emotions and emotion primitives; the Elan to annotate and synchronize the visual and vocal data; and multivariate statistical analysis (FAMD and MFA methods) to correlate qualitative and quantitative data.

In Madureira, Fontes and Fonseca (2016), the ExpressionEvaluator, the VPAS, a semantic componential questionnaire and multivariate statistical analysis were used in an acoustic and

perceptual experiment exploring the expressive characteristics of voice quality settings and contrasting professional and non-professional speech styles. The results indicated that the two styles were differentiated by means of fundamental frequency (f0) related measures, the f0 inter-quartile semi-amplitude and the f0 median, revealing varying f0 and fast f0 variation speed in professional speech style.

Menegon and Madureira (2016) investigated voice quality settings in singing data. Three kinds of singing instructional metaphors were analyzed. The analysis combined the use of the ExpressionEvaluator, analysis of facial expressions and the VPAS. The vocal tract configurations used for singing after the metaphorical instructions were found to be iconically related to the forms suggested by the metaphors.

In Da Lomba, Madureira and Fontes (2016), singing, reciting and speech data were contrasted using a semantic differential questionnaire, a perceptual task, the ExpressionEvaluator and multivariate statistical analysis. The results indicated that singing, reciting and speech data affect listeners in diverse ways.

In Barbosa, Camargo and Madureira (2016) five Praat scripts (BeatExtractor, SGdetector, SalienceDetector, ProsodyDescriptor and ExpressionEvaluator) developed by Barbosa for analysing prosodic elements in expressive and pathological speech are presented. Automatic, language-independent detection of vowel onsets is performed by the BeatExtractor script; language-dependent, semi-automatic detection syllable-sized normalised duration peaks for the study of prominence and boundary marking by the SGDetector; language-independent automatic detection of syllable-sized normalised duration peaks for the study of prominence and boundary by the SalienceDetector script; duration, fundamental frequency and intensity measures to analyze prosodic elements in terms of rate and prominence by the ProsodyDescriptor; prosodic parameters related to duration, F0, LTAS and spectral emphasis by the ExpressionEvaluator for the study of rhythm, intonation and voice qualities.

In Camargo *et al* (2019) links among perception, acoustic characteristics, voice quality profiles and gender are examined by statistical analysis weighing the relative proximity of controlled factors. Fundamental frequency was found to play a relevant role in establishing perceptual and acoustic correlations.

In Madureira (2019), the expressive uses of Brazilian Portuguese rhotic varieties and voice quality settings were analyzed. The methodological procedures involved the use of VPAS, description of articulatory-acoustic features of rhotic sounds, a semantic differential questionnaire and multivariate statistical analysis. The results indicate uses of sound symbolism in expressing meaning effects.

In Madureira e Fontes (2019) the VPAS, the FACS, the ExpressionEvaluator script, a semantic componential analysis questionnaire, the Affectiva SDK tool and multivariate statistical analysis were used to correlate semantic, visual and acoustic features to investigate speech expressivity. FACS and VPAS structural similarities are pointed out.

5. Methodological proposal for analyzing expressive uses of voice qualities

In the works reported in the previous section, methodological procedures were developed to cope with the challenge of describing vocal quality settings and correlating them to meaning effects, acoustic measures and perceptual judgments. These are the cornerstones of the methodological approach proposed.

For the description of voice qualities, the VPAS must be used due to its comprehensiveness and phonetic precision. Some difficulties in correlating voice quality settings, acoustic characteristics and meaning effects can be overcome by grouping voice quality settings according to their acoustic outputs and vocal tract configurations as proposed in this work.

For the analysis of meaning effects, semantic descriptors must be selected. Choice of descriptors must be well structured, considering the complex combinations of semantic components which constitute lexical items.

For the analysis of acoustic properties, scripts such as the ExpressionEvaluator and ProsodyDetector, running in PRAAT, are recommendable, since they provide means to access phonetic details.

The ExpressionEvaluator Script was developed by Barbosa (2009) to analyze speech expressivity data. It runs in the software PRAAT, developed by Paul Boersma and David Weenink from the University of Amsterdam, and extracts 12 acoustic measures: *fundamental frequency*: f_0 median (mednf0), inter-quartile semi-amplitude (sampquartisf0), skewness, and 0.995 quantile (quan995f0); *fundamental frequency derivative*: df0 mean (medderivf0), standard deviation (desvpaddf0), and skewness (assimdf0div10); - *intensity measures*: intensity skewness (assimint); *spectral tilt*: spectral tilt mean (medinclinespec), standard deviation (desvadinclinespec), and skewness (assiminclinespec); *Long-Term Average Spectrum*: LTAS frequency standard deviation (desvapadltas).

For the analysis of perception meaning effects, perceptual tests containing rating scales to evaluate semantic descriptors and their application to large groups are desirable from the statistical point of view. Semantic differential questionnaires and tools to perform perception of evaluation tests such as the GTrace, developed by McKeown *et al* (2012), are useful for that purpose.

For establishing correlations among the semantic, acoustic and perceptual domains, multivariate statistics is desirable. A friendly tool to apply these statistical tests is FactorMinerR (Husson *et al* 2013). It runs in the freely available statistical software "R".

Multivariate statistical analysis includes the Multiple Factor Analysis (MFA), the Multiple Correspondence Analysis (PCA), the Multiple Correspondence Analysis (MCA) and the (FAMD) methods (Husson *et al*, 2013).

The Multi Factor Analysis (MFA) is a statistical factorial method which applies to variables structured in groups: a group of quantitative variables and a group of qualitative variables. The MFA can be viewed as an extension of two kinds of analysis: PCA, calculating the principal components of quantitative variables and MCA, calculating qualitative variables.

The FAMD is applied simultaneously to both quantitative and qualitative variables and calculates proximity relations among the groups of quantitative and qualitative variables.

The application of multivariate analysis involves three steps: finding a common structure among variables; describing the specificity of each group of variables by means of correlation analysis; and comparing the resulting values by means of the individual analyses of the variables (Fontes, 2014).

After considering the theoretical and methodological issues concerning the investigation of speech expressivity, we move on to introduce, in the following section, our proposal concerning expressive uses of vocal quality settings.

6. Our proposal for analyzing expressive uses of vocal qualities

In order to analyze the expressive uses of vocal qualities, one of the challenges is grasping how links between form and meaning are motivated since acoustic characteristics and vocal tract configurations must be assigned and confronted to the sound symbolism codes underlying speech expressive uses. In adopting the phonetic model of description of voice qualities developed by Laver, one is faced with 52 settings and the challenge is to interpret them in terms of symbolic codes.

Our proposal is that form meaning links can be established considering shared acoustic properties and vocal tract configurations among voice quality settings. That would organize the 52 voice quality settings into 6 groups as following explained.

The first group concerns voice quality settings which diminish the vocal tract (lip spreading, labiodentalization, raised larynx and pharyngeal constriction), or which are characterized by increased overall muscular tension and increased subglottal pressure (tense vocal tract and tense larynx settings) or yet increased vibration of the vocal folds (falsetto)since all of them cause frequency to become higher.

The second group concerns voice quality settings which enlarge the vocal tract (protruded and round lips, protruded jaw, lowered larynx, pharyngeal constriction), or which are characterized by decreased overall muscular tension and decreased subglottal pressure (lax larynx setting) or decreased vibration of the vocal folds (creak and creaky voice) since all of them cause frequency to become lower.

The meaning effects of these vocal qualities mentioned in the two previous paragraphs can be interpreted in terms of the frequency code as proposed by Ohala (1994) and the effort code as proposed by Gussenhoven (2002) and are according to the principle of iconicity.

The third group concerns vocal quality settings which diminish the oral cavity (raised tongue body, fronted tongue tip and body tongue position) or its aperture causing the proprioception of diminutive size.

The fourth group concerns vocal quality settings which enlarge the oral cavity (lowered tongue body, retracted tip and backed body tongue position) or its opening causing the proprioception of enlarged size.

The meaning effects of these vocal qualities mentioned in the two previous paragraphs can be interpreted in terms of the sound metaphors as proposed by Fonagy (1983). The same can be thought of the vocal quality settings which are characterized by some kind of articulatory obstruction: the denasal voice quality setting and the close jaw voice quality setting.

The fifth group concerns vocal quality settings which are characterized by air escaping through the vocal folds (whisper, whispery and breathy voices) or through the nasal cavity and can be interpreted in terms of the sirenic code (Gussenhoven, 2016).

In her study on the relations between gender and voice, Biemans (2000) mentions that nasality correlated significantly with feminine behaviour: higher scores for feminine behaviour on the gender identity questionnaire are related to a more nasal rating of the speech.

The sixth group concerns vocal quality settings (harsh and tremor) which are characterized by irregularity of the vibration of the vocal and vocal quality settings which can be interpreted in terms of the iconicity principle. These are associated with negative meanings (Poyatos, 1991).

The proposed grouping of vocal quality settings from the acoustic output and vocal configuration perspectives is thought to be a considerable original contribution to the study of

the expressive uses of voice qualities since it allows to interpret them in terms of the speech symbolic codes, iconic and metaphorical aspects. Otherwise, links between sound and meaning effects in relation to expressive uses of voice qualities are difficult to be considered.

Conclusion

Voice quality is thought to be a major way of communicating paralinguistic and extralinguistic meanings, hence its relevance to the study of speech expressivity. Grasping its expressive usage demands a thorough methodological approach.

The innovative proposal reported in this work to describe expressive uses of voice qualities, emphasizing the grouping of vocal quality settings according to acoustic output and vocal configuration features, is meant to allow the correlations among the perceptual, acoustic and semantic domains involved in speech expressivity, taking into account, by means of analytical resources and tools, vocal and visual features, acoustic properties and meaning effects in relation to sound symbolic codes and sound symbolism types.

References

ABELIN, Asa. (1999). *Studies in Sound Symbolism*. Doctoral dissertation. Gothenburg: Göteborg University.

ALBANO, Eleonora. (1988). "Fazendo sentido do som, *Ilha do Desterro". A Journal of English Language, Literatures in English and Cultural Studies 19*, 11–26.

BARBOSA, Plinio, CAMARGO, Zuleica, MADUREIRA, Sandra. (2016). "Scripts for the Acoustic Analysis of Speech Data", in S. Madureira (org) *Sonoridades – Sonorities*, São Paulo: Edição da Pontifícia Universidade Católica de São Paulo.

BARBOSA, Plinio A., MADUREIRA, Sandra. (2015). *Manual de Fonética Acústica Experimental: Aplicações a Dados do Português*. São Paulo: Cortez.

BARBOSA, Plinio A. (2009). "Detecting changes in speech expressiveness in participants of a radio program". *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009)*, Brighton, 2155–2158.

BELLER, Gregory, HUEBER, Thomas, SCHWARZ, Diemo and Rodet, Xavier. (2006). "Speech Rates in French Expressive Speech", *Speech Prosody 2006*, Dresden.

BIEMANS, Monique. (2000). *Gender variation in voice quality*. Utrecht: Landelijke Onderzoekschool Taalwetenschap LOT, Utrecht University.

BLASI, Damian E., WICHMANN, Soren, HAMMARSTRÖM, Harald, STADLER, Peter F. and CHRISTIANSEN, Morten H. (2016). "Sound-meaning association biases evidenced across thousands of languages". *Proceedings of the National Academy of Sciences of the United States of America*, *113*(39), 10818–10823. doi:10.1073/pnas.1605782113

BOLINGER, Dwight. (1986). *Intonation and its Parts: Melody in Spoken English*. London: Edward Arnold.

CAMARGO, Zuleica, MADUREIRA, Sandra, REIS, N., RILLIARD, A. (2019). "The phonetic approach of voice qualities: challenges in corresponding perceptual to acoustic descriptions".

In J. M. Lahoz-Bengoechea, R. P. Ramón (eds) *Subsidia: Tools and Resources for Speech Sciences*, University of Malaga, 11-17.

CAMARGO, Zuleica, GOMES, P.C. , MADUREIRA, Sandra, RUSILO, L. C. (2015). "Voice quality from a phonetic perspective: supralaryngeal and muscular tension settings". In: *18th International Congress of Phonetic Sciences*, 2015, Glasgow. *18th International Congress of Phonetic Sciences-Proceedings*. London: International Phonetic Association, v. 1, 117-120.

CAMARGO, Zuleica, MADUREIRA, Sandra. (2008). "Voice quality analysis from a phonetic perspective: Voice Profile Analysis Scheme Profile for Brazilian Portuguese (BP-VPAS)", in *Proceedings of the Fourth Conference on Speech* Prosody, Campinas. Campinas: Capes, Fapesp, CNPq, v. 1, 57-60.

CHUENWATTANAPRANITHI, Suthathip, XU, Yu, THIPAKORN, Bundit, MANEEWONGVATANA, Songrit. (2008). "Encoding emotions in speech with the size code -- A perceptual investigation". *Phonetica 65*, 210-230.

CONTINI, M. (2009). "Les phonosymbolismes: continuité d'une motivation primaire ?", in R. van Deyck (éd.), *Travaux de Linguistique, Revue internationale de linguistique française*, *59*, Bruxelles: Duculot, 77-103.

DA LOMBA, Juliana A., MADUREIRA, Sandra, FONTES, Mario Augusto S. (2016). "As Palavras Cantada, Falada e Declamada: Um Estudo Sobre a Relação Entre, Fala e Declamação e seus Efeitos Impressivos", in S. Madureira (org) *Sonoridades* Canto – *Sonorities* (p.17-30), São Paulo: Edição da Pontifícia Universidade Católica de São Paulo.

DINGEMANSE, Mark, BLASI, Damian E., LUPYAN, Gary, CHRISTIANSEN; Morten H., MONAGHAN, Padraic. (2015). "Arbitrariness, iconicity and systematicity in language". *Trends in Cognitive Sciences 19*(10), 603–615.

DUEZ, Danielle. (1997). "Acoustic markers of political power". *Journal of Psycholinguistic Research 26*, 641-654.

EKMAN, Paul, FRIESEN, Wallace V. and HAGER, J. C. (eds.) (2002). *Facial Action Coding System* (2nd ed.). Salt Lake City, UT: Research Nexus eBook.

EKMAN, Paul and FRIESEN, Wallace V. (1976). "Measuring Facial Movement". *Environmental Psychology and Nonverbal Behavior*, *1*(1), 56-75.

FANT, Gunnar. (1970). *Acoustic Theory of Speech Production*. Mouton: The Hague.

FÓNAGY, Iván (2001). *Languages within Language: an Evolutive Approach.* Amsterdam: John Benjamins.

FONAGY, Iván (1983). *La vive voix: Essais de psycho-phonétique*. Paris: Payot.

FONTES, Mario Augusto S., MADUREIRA, Sandra. (2015). "Gestural Prosody and the Expression of Emotions: a Perceptual and Acoustic Experiment". The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, UK: The University of Glasgow. Paper number 0390 retrieved from http://www.icphs2015.info/pdfs/Papers/ICPHS0390.pdf

FONTES, Mario Augusto S. (2014). "Gestualidade vocal e visual, expressão de emoções e comunicação falada". Tese de Doutorado inédita. São Paulo : Pontifícia Universidade Católica de São Paulo.

GOBL, Christer, CHASAIDE, Ailhbe N. (2003). "The role of voice quality in communicating emotion, mood and attitude", *Speech Communication 40*(1–2), 189–212.

GORDON, Matthew, LADEFOGED, Peter. (2001). "Phonation types: a cross linguistic overview". *Journal of Phonetics 29*, 383-406.

GUSSENHOVEN, Carlos. (2016). "Foundations of intonation meaning: anatomical and physiological factors", in Oller, D. K., Dale, R. and U. Griebel, U. (eds), *Topics in Cognitive Science, 8*(2), 425–434.

GUSSENHOVEN, Carlos. (2002). "Intonation and Interpretation: Phonetics and Phonology". *Proceedings of the 1st International Conference on Speech Prosody*, Aix-en-Provence, 47-57.

HINTON, Leanne, NICHOLS, Johanna, OHALA, John J. (eds.) (1994). *Sound Symbolism.* Cambridge: Cambridge University Press.

HONIKMAN, B. (1964). "Articulatory settings", in D. Abercrombie, D.B. Fry, P.A.D. MacCarthy, N.C. Scott, J. L. M. Trim (eds), *In Honour of Daniel Jones*, London: Longman, 73-84.

HUSSON, François, JOSSE, Julie, LE, S, MAZET, Jeremy M. (2013). *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R* (version 1.25) [R package]. URL : http://factominer.free.fr.

JACEWICZ, Eva, FOX, Robert A., O'NEILL, Caitlin. (2009). "Articulation rate across dialect, age, and Gender", *Language Variation and Change 21*, 233–256.

JAKOBSON, Roman (1978). *Six Lectures on Sound and Meaning*, Cambridge, MA: MIT Press.

JAKOBSON, Roman, Waugh, Linda R. (1979). *The Sound Shape of Language*, Indiana University Press and Harvester Press.

JOHNSTONE, T., SCHERER, K. (1999). "The effects of emotions on voice quality", in *Proceedings of the 14th International Congress of Phonetic Sciences*. San Francisco, CA: University of California, 2029–2032.

JOHNSTONE, T., SCHERER, K. (2000). "Vocal communication of emotion", in M. Lewis and J. M. Haviland-Jones (eds.), *Handbook of Emotions.* New York: Guilford Press, 220–223.

KREIMAN, Jody, SIDTIS, Diana. (2011). *Foundations of voice studies: Interdisciplinary approaches to voice production and perception.* Boston: Wiley-Blackwell.

LADEFOGED, Peter. (1980). "What are linguistic sounds made of language", *Language, 56*, 485-502.

LAVER, John. (2000). "Phonetic evaluation of voice quality", in R. D. Kent and M. J. Ball (eds.), *Voice Quality Measurement*. San Diego, CA: Singular, 37–48.

LAVER, John. (1980). *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

LAVER, John, MACKENZIE-BECK, Janet. (2007). *Vocal Profile Analysis Scheme–VPAS* [handout]. Edinburgh: Queen Margareth University College, Research Centre.

LAVER, John, WIRZ, Sheila, MACKENZIE-BECK, Janet, M., HILLER, Steven M. (1981). "A perceptual protocol for the analysis of vocal profiles: work in progress". *Edinburg University Department of Linguistics 14*, 139-55.

LIU, Xialuan and XU, Yu (2014). "Body size projection and its relation to emotional speech—Evidence from Mandarin Chinese". *Speech Prosody 2014* (p. 974-977), Dublin,.

MACKENZIE-BECK, Janet. (2005). Perceptual analysis of voice quality: the place of the Vocal Profile Analysis, in W. J. Hardcastle and J. Mackenzie-Beck (eds.), *A Figure of Speech: a Festschrift for John Laver* (p. 285–322). Mahwah, NJ: Lawrence Erlbaum,.

MADUREIRA, Sandra. (2019). "Portuguese rhotics in poem reciting: perceptual, acoustic, and meaning-related issues", in M. Gibson and J. Gil (orgs.) *Romance Phonetics and Phonology* (p. 191-215). Oxford : Oxford University Press,.

MADUREIRA, Sandra. (2011). "The Investigation of Speech Expressivity", in H. Mello, A. Panunzi, T. Raso (Eds.), *Illocution, modality, attitude, information patterning and speech annotation*. Firenze: Firenze University Press, *1*, 101-118.

MADUREIRA, Sandra. (2008). "Reciting a sonnet: production strategies and perceptual effects", in P. A. Barbosa, S. Madureira, and C. Reis (eds.), *Proceedings of the 4th Conference on Speech Prosody* (*Speech Prosody 2008*). São Paulo: RG, 697–700.

MADUREIRA, Sandra, FONTES, Mario Augusto S. (2019). "The analysis of facial and speech expressivity: tools and methods", in J. M. Lahoz-Bengoechea, R. P. Ramón (eds) *Subsidia: Tools and Resources for Speech Sciences* (p. 19-26), University of Malaga,

MADUREIRA, Sandra, FONTES, Mario Augusto S., FONSECA, B. C. (2016). "Voice Quality and Speaking Styles", in *Dialectologia*, Special Issue VI, 171-190.

MADUREIRA, Sandra, CAMARGO, Zuleica. (2010). Exploring sound symbolism in the investigation of speech expressivity, in A. Botinis (ed.), *Proceedings of the 3rd ISCA Tutorial and Research Workshop on Experimental Linguistics* (*ExLing 2010*). Athens: ISCA and University of Athens, 105-108.

MARTINEZ, Brian, VALSTAR, Michel, JIANG, Bihan, PANTIC, Maja. (2017). "Automatic Analysis of Facial Actions: A Survey", *Transactions on Affective Computing*, *10*(3), 325-347.

MCGURK, Harry, MACDONALD, John. (1976). "Hearing lips and seeing voices". *Nature 264*, 746-748.

MCKEOWN, Gary, VALSTAR, Michel, COWIE, Roddie, PANTIC, Maja, SCHRÖDER, Marc. (2012). "The SEMAINE database: annotated multimodal records of emotionally coloured conversations between a person and a limited agent IEEE". *Transactions of Affective Computing 3*, 165-183.

MENEGON, Paulo, MADUREIRA, Sandra. (2016). "Metáforas no Ensino de Canto e seus Efeitos na Qualidade Vocal: um Estudo Acústico-Perceptivo", in S. Madureira (org.) *Sonoridades - Sonorities*. São Paulo: Edição da Pontifícia Universidade Católica de São Paulo.

MORTON, Eugene S. (1977). "On the occurrence and significance of motivation-structural rules in some bird and mammal sounds". *American Naturalist 111*, 855-869.

NOBILE, Luca. (2011). "Words in the mirror: analysing the sensorimotor interface between phonetics and semantics in Italian", in P. Michelucci, O. Fischer and C. Ljungberg (eds), *Iconicity in Language and Literature 10*: *Semblance and Signification*, Amsterdam/Philadelphia, John Benjamins, 101-131.

NOBLE, Luca, XU, Yu. (2011). "Friendly Speech and Happy Speech – Are they the same?", *The 17th International Congress of Phonetic Sciences*, Hong Kong: 1502-1505.

OHALA, John J. (1997). "Sound symbolism", in *Proceedings of the 4th Seoul International Conference on Linguistics [SICOL]* 11-15 Aug 1997, 98-103.

OHALA, John J. (1994). "The frequency codes underlies the sound symbolic use of voice pitch", in L. Hinton, J. Nichols & Ohala J.J. (eds.), *Sound symbolism* (p. 325-347). Cambridge: Cambridge University Press.

OHALA, John J. (1984). "An ethological perspective on common cross - language utilization of F0 of voice". S. Karger AG, *Phonetica, 41*, Basel, Switzerland, 1-16.

OHALA, John J. (1983). Cross-language use of pitch: an ethological view. *Phonetica 40* (1), 1-18.

OHALA, John J. (1980). "The acoustic origin of the smile," *Journal of the Acoustical Society of America 68*, S33, np. DOI : https://doi.org/10.1121/1.2004679.

OSGOOD, Charles E., SUCI, Georges J., TANNENBAUM, Percy H. (1957). *The Measurement of Meaning*. Urbana, IL: University of Illinois Press.

PISKE, Thorsten, MACKAY, Ian R.A., FLEGE, James E. (2001). "Factors affecting degree of foreign accent in an L2: A review," *Journal of Phonetics 29*, 191–215.

POYATOS, Fernando. (1991). Paralinguistic modifiers: our many voices. Language and Communication, *1*(3),181-195, Printed in Great Britain, Pergamon Press plc.

SAPIR, Edward. (1972 [1927]), "Speech as a Personality Trait in J. Laver, S. Hutcheson (orgs.). Communication in Face-to-Face Interaction. Great Britain: Penguin Books,71-81.

SCHERER, Klaus R. (2005). "What are emotions? And how can they be measured?", *Social Science Information 44*(4), 695-729.

SCHERER, Klaus R. (1979). "Non-linguistic indicators of emotion and psychopathology", in C. E. Izard (ed), *Emotions in personality and psychopathology* (p. 495-529), New York: Plenum.

SCHERER, Klaus R., Sundberg, Johan, Tamarit, Luca, Salomão, Glaucia L. (2015). "Comparing the expression of emotion in the speaking and the singing voice". *Computer Speech & Language 29*(1), 218-235.

SVANTESSON, Jan Olof. (2017). "Sound symbolism: the role of word sound in meaning". *WIREs Cognitive Science*, 8: n/a, e1441. doi:10.1002/wcs.1441.Tian, Yi., Kanade, T., Cohn, J. F. (2011). Facial Expression Recognition, *Handbook of Face Recognition*, 487-519.

TOLKMITT, Franck J., Scherer, Klaus L. (1986). "Effects of experimentally induced stress on vocal parameters", *Journal of Experimental Psychology: Human Perception and Performance*, *12*(3), 302–313.

TSUR, Reuvent. (1992). *What Makes Sound Patterns Expressive? The Poetic Mode of Speech Perception*. Durham, NC: Duke University Press.

VAN BEZOOIJEN, Renée. (1984). *Characteristics and Recognizability of Vocal Expressions of Emotion*. Dordrecht: Foris.

VIEIRA, C. Oliveira. (2014) *Crátilo, ou sobre a correção dos nomes.* São Paulo: Paulus.

XU, Yu, Kelly, Andrew and Smillie, Cameron. (2013). "Emotional expressions as communicative signals", in S. Hancil and D. Hirst (eds.), *Prosody and Iconicity* (p. 33-60). Amsterdam : John Benjamins.

XU, Yu, Chuenwattanapranithi, Suthathip. (2007). "Perceiving anger and joy in speech through the size code". *The 16th International Congress of Phonetic Sciences*, Saarbrucken, 2105-2108.

Appendix

VPAS Laver & Mackenzie-Beck (2007)

Speaker: Date of recording: Judge: Recording ID:

	FIRST PASS		SECOND PASS							
	Neutral	Non-neutral	SETTING	moderate			extreme			
				1	2	3	4	5	6	
A. VOCAL TRACT FEATURES										
1.Labial			Lip rounding/protrusion							
			Lip spreading							
			Labiodentalization							
			Minimised range							
			Extensive range							
2. Mandibular			Close jaw							
			Open jaw							
			Protruded jaw							
			Extensive range							
			Minimised range							
3.Lingual tip/blade			Advanced tip/blade							
			Retracted tip/blade							
4. Lingual body			Fronted tongue body							
			Backed tongue body							
			Raised tongue body							
			Lowered tongue body							
			Extensive range							
			Minimised range							
5.Pharyngeal			Pharyngeal constriction							
			Pharyngeal expansion							
6.Velopharyngeal			Audible nasal escape							
			Nasal							
			Denasal							
7. Larynx height			Raised Larynx							
			Lowered Larynx							
B. OVERALL MUSCULAR TENSION										
8. Vocal tract tension			Tense vocal tract							
			Lax vocal tract							
9. Laryngeal tension			Tense larynx							
			Lax larynx							
C. PHONATION FEATURES										
	SETTING		**Present**		**Scalar Degree**					
			Neutral	**Non-neutral**	**Moderate**			**Extreme**		
					1	**2**	**3**	**4**	**5**	**6**
10. Voicing type	Voice									
	Falsetto									
	Creak									
	Creaky									
11. Laryngeal frication	Whisper									
	Whispery									
12.Laryngeal irregularity	Harsh									
	Tremor									
		Neutral	**SETTING**		**moderate**			**extreme**		
					1	**2**	**3**	**4**	**5**	**6**
D. PROSODIC FEATURES										

13.Pitch	**Mean**		High						
			Low						
	Range		Minimised range						
			Extensive range						
	Variability		High						
			Low						
14. Loudness	**Mean**		High						
			Low						
	Range		Extensive range						
			Minimised range						
	Variability		High						
			Low						
E. TEMPORAL ORGANIZATION									
15. Continuity			Interrupted						
16. Rate			Fast						
			Slow						
F. OTHER FEATURES									
17. Respiratory support			Adequate						
			Inadequate						
18. Dyplophonia			Absent						
			Present						