GESTURAL PROSODY AND THE EXPRESSION OF EMOTIONS: A PERCEPTUAL AND ACOUSTIC EXPERIMENT

Mario Augusto de Souza Fontes[1] *and Sandra Madureira*[2]
[1,2] LIAAC, [2]LAEL, [1,2]Department of Linguistics, PUCSP
marioasfontes@gmail.com and sandra.madureira.liaac@gmail.com

ABSTRACT

This paper presents a perceptual and acoustic experiment and introduces methodological procedures to deal with qualitative and quantitative variables. Its objectives are: investigating the functions of vocal and facial gestures in the appraisal of six basic emotions (Anger, Distaste, Fear, Happiness, Sadness and Shame) and valence (positive, neutral and negative); discussing the interaction between the visual, vocal and semantic dimensions in the evaluation of audio, visual and audiovisual stimuli corresponding to 30 utterances (10 of them semantically positive, 10 neutral and10 negative). The correlation among the variables was made by non-parametric tests applying FAMD and MFA. Among the perceptual and acoustic variables investigated, the most influential for the identification of valence/emotions were found to be the VPAS and the ExpressionEvaluator measures. Judgments concerning the positive, negative and neutral valence of the utterances and the type of emotion varied accordingly to the kind of stimuli (audio, visual or audiovisual).

Keywords: Vocal and Visual Prosody, Gestuality, Perceptual Evaluation, Acoustic Phonetics.

1. INTRODUCTION

This paper presents a perceptual and acoustic experiment whose objective is investigating the functions of the gestural prosody (vocal and facial) in the appraisal of six basic emotions [5] and discussing the interaction between the visual, vocal and semantic cues in the evaluation of a corpus consisting of 30 utterances so as to explore the links between gestural prosody and emotive expression.

Methodological procedures to relate qualitative and quantitative measures are introduced. Differences in the interpretation of the utterances are expected if audio and visual cues are considered separately or combined. It is also expected that depending on the nature of the emotion or the semantic load of the utterance, the visual or the vocal aspects will have a stronger influence on perceptual judgments.

There has been an increasing interest in exploring the speech expression of emotions and well as the facial expression of emotions because of their communicative power, social and clinical relevance and technological applications to affect computing.

A review on the theoretical perspectives concerning the expression of emotions in psychology can be found in [3]. The author discusses the Darwinian, the Jamesian, the cognitivists and the social constructivist theoretical perspectives and claims that these perspectives have begun to converge. Besides those philosophical [19] and physiological perspectives [13] present relevant contributions to be taken into account in the process of understanding how emotions are expressed by vocal and facial gestures.

Results of the research on the perception of emotions demonstrate that judges are able to infer speech with better-than-chance accuracy [16]. Among the works demonstrating correlations between vocal gestures and types of emotions [1, 17,18] are landmarks. Commonly reported research findings associating lower F0 values and minimized F0 extension were found in the expression of sadness, depression, boredom and disappointment and higher F0 values and maximized F0 extension in the expression of joy and hot anger [20,10]. Speech segment reduction, slow speech rate and lower values of formant frequencies in the speech expression of sadness and boredom have also been pointed out [11].

As far as the correlation of facial characteristics and the expression of emotions is concerned commonly reported associations are: happiness/joy: raised eyelid, lip corners upwards and open mouth; *anger*: stared eyes, wrinkled forehead, superior and inferior eyelids upwards, open mouth, lowered eyebrows and lips firmly pressed; *fear*: superior eyelid upwards and raised eyebrows; wrinkled forehead, eyes wide open and mouth slightly open; *sadness*: furrowed eyebrow, tears, and corners of the mouth downwards. Among the works which bring about those associations [4,6,7,15] can be highlighted.

This work faces the challenge of investigating how facial, vocal and semantic factors may affect the judgement of emotional expression.

The integration of visual, audio and semantic cues is an important issue in speech communication because conflicting information has to be resolved.

2. METHODS

2.1 The corpus, subjects and stimuli

The corpus consists of 30 utterances taken from the documentary "Jogo de Cena" by Eduardo Coutinho, a Brazilian filmmaker. It contains self- narratives of real life experiences told by women who had experienced them and retelling of the same narratives by actresses.

The corpus had 10 utterances with positive qualifiers such as "lindo" (beautiful), 10 with negative qualifiers such as "horroroso" (awkward) and 10 without qualifiers. Utterances from 1 to 10 contained positive qualifiers, from 11 to 20 no qualifiers and from 21 to 30 negative qualifiers.

The utterances were produced by 11 subjects, 7 of them being actresses and 4 non-actresses. The latter were the women who narrated facts they have experienced and the former interpreted the stories these women had narrated.

There were 3 types of stimuli: audio, visual and audiovisual. The same utterance was presented in audio, visual and audiovisual form in separate sessions.

2.2 The methodological procedures

The relations between gestural prosody and emotional expression were investigated by means of the following analytical procedures and methods: acoustic analysis; perceptual analysis of voice quality, emotions, valence and facial gestures and multivariate statistical analysis.

Therefore, two kinds of variables were concerned: qualitative and quantitative. To correlate them, non-parametric tests applying the FAMD and MFA methods were used [16].

Acoustic measures were automatically extracted by the ExpressionEvaluator Script developed by [2] and running in PRAAT. The script extracts 13 measures: f0 measures: f0 median (mdnf0), inter-quartile semi amplitude (sampquartisf0), skewness and 0.995 quantil (quan995f0); f0 derivative: df0 mean (medderivf0), standard deviation (desvpaddf0), skewness (assimdf0div10); intensity measures: intensity skewness (assimint); promptness (the difference between the acoustic energy of the integral signal and the intensity of the low pass filtered signal, upper band limit equal to 1,5* average f0 of the acoustic signal under analysis); spectral tilt: spectral tilt mean (medinclinespec), standard deviation (desvadinclinespec), skewness assiminclinespec); and LTAS: LTAS frequency standard-deviation (desvapadltas).

Perceptual evaluation tests to identify valence (positiveness, negativeness and neutrality) and a set of 6 basic emotions (happiness, anger, shame, fear, sadness and distress) applied to a group of 34 judges (undergraduate and graduate students) using the Gtrace developed by [14].

The test comprised 3 sessions, being one for the presentation of the audio stimuli, one for the visual stimuli and one for the audiovisual stimuli. The stimuli in each session were randomly presented.

Each session lasted for 20 minutes. In the first session either audio or visual stimuli were presented. The last session was always the audiovisual. Between the first and the second session the subjects had a 15 minute interval. Each subject took about 2 hours to answer the tests.

A descriptive profile was built to annotate the movements and directionality of the facial organs. The inspection of the facial gestures was performed with the help of Elan from Max Planck Institute of Psycholinguistics.

In order to identify the vocal quality settings the Vocal Profile Analysis Scheme (VPAS) developed by [12] was used. The settings were described by a phonetician with great expertise in the use of the scheme.

In order to correlate the qualitative and quantitative measures, two methods of explorative multivariate analysis the multiple functional analysis (MFA); and Factor Analysis of Mixed Data (FAMD) were applied. The data were analyzed with the software R, Rcommander, and FactoMinerR [9].

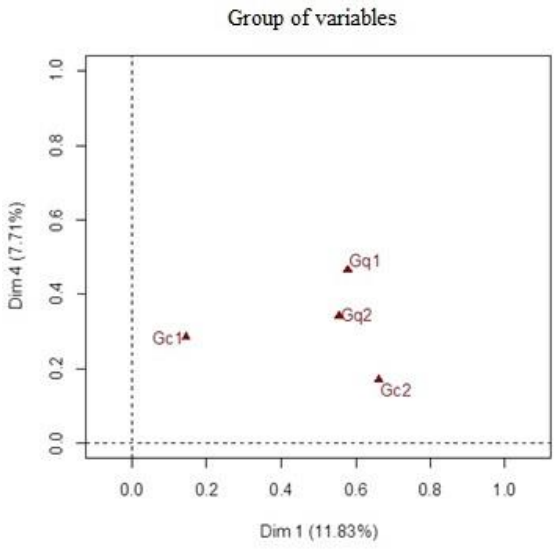
3. RESULTS AND DISCUSSION

The analysis of the role of the vocal and visual gestures in the identification of valence and emotions showed that VPAS variables and the ExpressionEvaluator measures were quite influential. VPAS was the most robust factor (MFA=3.47) to represent the vector space of the variables under study.

Among the variables studied, VPAS variables "Raised Larynx" and "High Pitch", the acoustic measures "mednf0, quan99.5f0, intensity skewness and spectral tilt mean" and the visual variables lips and eyes movements were found to be more influential. These variables showed statistically significant differences ($p < 0.05$) and were the most influential factors in associating vocal and visual gestures to emotional expression.

In Figure 1 Gq1 refers to the VPAS variables, Gq2 refers to the facial variables, Gc1 to the emotion variables and Gc2 to the acoustic measures.

Figure 1: Groups of variables in dimensions 1 and 4 (DIM1 and Dim4): Gq1 refers to the VPAS variables, Gq 2 to the facial variables, Gc1 to the emotion variables and Gc2 to the acoustic measures.



Both visual and vocal prosody were found to interact with the semantic domain, emphasizing or changing the semantic load of the utterances.

In 14 out of 30 utterances judgments of valence were the same in the three kinds of stimuli and in two of them no coincidence was found. These two utterances had positive qualifiers and this semantic feature conflicted with the visual and vocal cues of the stimuli: head downwards, teary eyes hyper-functional voice quality, high pitch, fast speech rate and tongue clicking

No utterances with positive qualifiers were judged negatively based on the audio stimuli presentations and only 1 when audiovisual stimuli were presented, but in visual stimuli presentations 4 of them were.

The fact that there were more negative judgments concerning visual stimuli can be interpreted in relation to the lack of semantic cues. In audio and audiovisual stimuli the oral or the oral and visual cues interact with the semantic ones and conflicting information can arise.

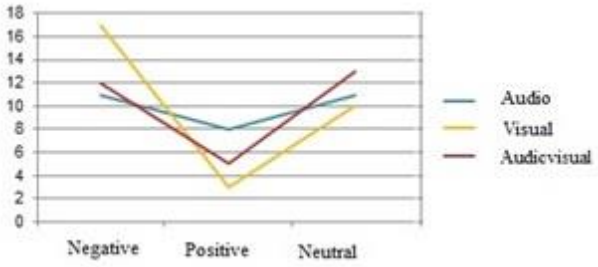
No audio, visual or audiovisual stimuli relative to the utterances with negative qualifiers were judged positively. In this case variation was not orthogonal, since both stimuli and semantic characteristics converged.

Differences in valence judgment were also found in relation to utterances without qualifiers. Two utterances without qualifiers were judged as negative, 1 as positive and 7 as neutral when audio stimuli were presented. When visual stimuli were presented 5 utterances were judged as negative, 1 as positive and 4 as neutral. When the audiovisual stimuli were presented 3 utterances were judged as negative, 1 as positive and 7 as neutral.

These differences reveal levels of semantic orthogonality in relation to the utterances without qualifiers: 70% of the audio stimuli, 40% of the visual stimuli and 60% of the audiovisual stimuli were judged as neutral.

Figure 2 indicates the levels of semantic orthogonality found in our data. When visual stimuli were presented 17 utterances out of 30 were evaluated in a negative way, but when audiovisual and audio stimuli were presented the results were respectively 12 and 11. Only 3 utterances were judged as positive when visual stimuli were presented.

Figure 2: Negative, positive and neutral appraisal of audio, visual and audiovisual stimuli. In the ordinate axis the number of utterances and in the abscissa axis the kind of valence judgements.



The explorative multivariate MFA and FAMD methods were applied to correlate the variables. Both methods yielded equivalent results. Redundancy was found in the arrangement of the utterances in four of the dimensions of the vector space.

According to the results of the FAMD, the quantitative variables on dimensions 1 and 4 were found to be more representative. In these dimensions the VPAS and the acoustic measures were found to be more representative of the vector space than the other groups of variables.

Table 1 shows the distribution of the 30 utterances in 4 dimensions as analyzed by means of MFA and FAMD. In bold the numbers of the utterances which emerged in the same positions in both methods.

As mentioned before, utterances numbered from 1 to 10 contained positive qualifiers and from 21 to 30 negative qualifiers. Utterances from 11 to 20 had no qualifiers.

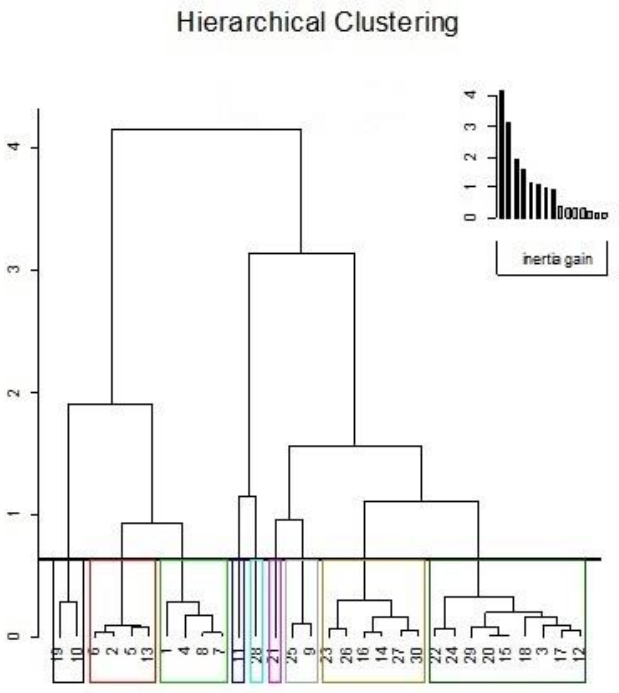
Table 1: Distribution of the 30 utterances in the 4 dimensions as analyzed by MFA and FAMD.

MFA				FAMD			
Dim.1	Dim.2	Dim.3	Dim.4	Dim.1	Dim.2	Dim.3	Dim.4
10	**28**	9	**21**	**10**	**28**	9	**21**
19	21	8	11	**19**	11	27	25
11	3	19	9	**11**	26	21	9
9	26	7	28	4	17	19	11
25	5	11	25	**25**	3	8	5

A higher degree of precision in grouping the utterances evaluated with the same kind of emotion was obtained with FAMD deriving 9 clusters. A smaller number of clusters was found to be insufficient to separate the different types of emotion.

Figure 3 shows a hierarchical cluster derived with FAMD, grouping the utterances into 9 clusters.

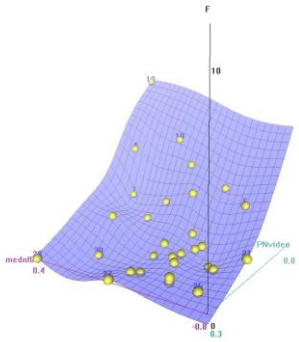
Figure 3: Hierarchical clustering showing the 30 utterances grouped into 9 clusters



These 9 clusters were obtained by reducing the number of variables, that is, including only the most influential qualitative and quantitative factors to identify the types of emotion as well as by applying linear regression.

Figure 4 shows the results one of the 4 smooth linear regressions taking into account three influential variables: mednf0, pnvideo and happiness**.**

Figure 4: Graphic derived from a smooth linear regression, taking into account Mednf0, pnvideo and happiness as variables. The dots represent the 30 utterances.



Explorative multivariate analysis made it possible to correlate qualitative and quantitative variables and identify the most influential factors in determining the type of emotion and valence. They were the VPAS variables (Raised Larynx Voice and Loudness) and the acoustic measures (desvpaddf0, mdnf0, quant99.5f0 and assmf0).

To identify sadness the relevant factors were Raised Larynx voice quality setting and desvpaddf0 (38%) in MFA. The same factors emerged with FAMD: desvpaddf0 (69%) and Raised Larynx voice quality setting (32%). To identify happiness the relevant factors were quan99.5f0 (72%) and mednf0 (76%) in MFA. The same factors emerged with FAMD: mdnf0 (66%) and quant99.5f0 (58%). To identify shame the relevant factor was Loudness (41% in MFA and 51% in FAMD) in dimension 4. To identify distaste the relevant factors were Loudness (41% in MFA and 52% in FAMD) and assmf0 (-48% in MFA and - 44% in FAMD). To identify fear the relevant factor was Loudness (41% in MFA and 52% in FAMD).

.

4. CONCLUSION

This work tackled issues concerning the vocal and the facial gestuality in the expression of emotion. Explorative multivariate analysis was applied to investigate the interactions among visual, vocal and semantic domains.

The most influential variables for the identification of valence/emotions were found to be the VPAS variables and the ExpressionEvaluator measures.

Visual stimuli had a higher tendency to be evaluated in a negative way than audiovisual or audio stimuli.

Judgments of emotions and valence were found to differ accordingly to the kind of stimuli (audio, visual or audiovisual) concerned.

5. REFERENCES

[1] Banse, R., Scherer, K. R. 1996. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70:614-636.

[2] Barbosa, P. A. 2009. Detecting changes in speech expressiveness in participants of a radio program. *Proceedings of Interspeech*. Brighton: United Kingdom, p. 2155-2158.

[3] Cornelius, R. R. 1996. *The science of emotion*. Research and tradition in the psychology of emotion. Upper Saddle River, NJ: Prentice-Hall.

[4] Darwin, C. 1872/1965. *The expression of the emotions in man and animals*. Chicago University of Chicago Press.

[5] Ekman, P. 1994. All Emotions are basic. In: Ekman, P Davidson, R. (ed.) *The nature of emotion: fundamental questions*. New York: Oxford University.

[6] Ekman, P. Friesen, W. V. 1978. *Facial Action Coding System*: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press: Palo Alto.

[7] Fridlund, A. J. 1994. *Human facial expression*: An evolutionary view. San Diego, CA: Academic Press.

[8] Husson, F.; LÊ, S.; Pagès, J. 2011. *Exploratory Multivariate Analysis by Example Using R*. CRC Press.

[9] Husson, F.; Josse J.; LÊ S. Mazet, J. 2013. *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R*. R package version 1.25. Available in: http://CRAN.R-project.org/package=FactorMineR

[10] Juslin, P. N.; Scherer, K. R.; Harrigan, J.; Rosenthal, R.; Scherer, K. R. (2005) Vocal Expression of Affect. In: Harrigan, R. Rosenthal, &. Scherer, (Eds.) *The New Handbook of methods in nonverbal behavior research*, Oxford University Press, Oxford, UK, p 65-135.

[11] Kienast, *M.;* Sendlmeier, *W. 2000.* Acoustical analysis of spectral and temporal changes in emotional speech. *Proc. of ITRW on Spheech and Emotion*. Newcastle, Northern Ireland, UK. 92-97.

[12] Laver, J. Mackenzie-Beck, J. 2007. *Vocal Profile Analysis Scheme – VPAS*. Edinburgh: Queen Margareth University College.

[13] Lövheim, H. 2012. A new three-dimensional model for emotions and monoamine neurotransmitters. *Medical Hypotheses*. 78(2):341-8.

[14] Mckeown, G.; Valstar, M.; Cowie, R.; Pantic, M. Schröder, M. 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.

[15] Mortillaro, M., Mehu, M. Scherer, K. R. 2011. Subtly different positive emotions can be distinguished by their facial expressions. *Social Psychological & Personality Science*. 2:262-271.

[16] Scherer, K. R. 1986. Vocal affect expression: A review and a model for future research. *Psychological Bulletin*. Washington: American Psychological Association, 99:143-165.

[17] _____. 2003. Vocal communication of emotion: a review of research paradigms. *Speech Communication*. 40:227-256.

[18] _____. 2005. What are emotions? And how can they be measured?. *Social Science Information*. 44 (4):693-727.

[19] Spinoza, B. 2009. *Ética, demonstrada segundo a ordem geométrica*. Tradução do Latim: Tomas Tadeu. Belo Horizonte: Edit. Autêntica.

[20] Van Bezzoijen, R. 1984. *The characteristics and recognizability of vocal expression of emotions*. Drodrecht: Foris. 141-145.