Vocal and facial expressions and meaning effects in speech expressivity

Mario Fontes, Sandra Madureira
Pontifical Catholic University of São Paulo, Brazil

Abstract

The objective of this work is to investigate the congruence between non-verbal and verbal cues in persuasive speech. The selected corpus comprises video excerpts in which artists from divergent political perspectives provide support to the minister of the Brazilian Supreme Federal Court. The research methodology comprises: annotation of the video excerpts; text analysis; automatic analysis of the speakers' facial expressions and emotions by means of the FaceReader; analysis of the vocal quality and prosodic settings by means of the VPA; acoustic analysis of the data by means of the ExpressionEvaluator (Barbosa, 2009); and multivariate statistical analysis, applying MFA in R, package FactorMinerR, to correlate qualitative and quantitative variables. Results indicate the interplay among facial and vocal prosodies and intended persuasiveness.

Key words: verbal and non-verbal language, gestural prosody, multimodality, persuasion, speech expressivity

Introduction

The objective of this work is to investigate the congruence between non-verbal and verbal cues in persuasive speech. The non-verbal aspects are key factors in oral communication reinforcing the semantic content, contradicting the semantic content or adding extra information.

Method

Corpus

The selected corpus is composed of video excerpts in which famous artists from divergent political perspectives provide support to the minister of the Supreme Federal Court, using a same utterance "Resist, Minister", with two different meanings: "Resist to changing the decision about the conviction at the second instance" and "Resist to the pressure of the ones who want you to change the decision about the conviction at the second instance".

The choice of the corpus was motivated by three factors: the semantic and pragmatic content of the text; the quality of the videos; and the possibility of comparing the oral interpretation of two groups of artists having different opinions about a same subject.

Subjects

The subjects are four male and four female artists. According to their political beliefs, they can be divided into two groups (two males and two females in each one). Some of the artists are pro and others are contra conviction at the second instance, that is, in favor or not of incarcerating people who, found guilty and again on appeal to higher courts, receive a five-year jail sentence at the second instance.

The analysis of facial and vocal expressions

The analysis of the facial action unities was performed automatically with the use of the FaceReader and the vocal quality and prosodic settings by means of VPA (Laver and Mackenzie Beck, 2007).   FACS and VPA systems are contrasted in Madureira and Fontes (2019).

The acoustic measures were automatically extracted by the ExpressionEvaluator script developed by Barbosa (2009) for Praat. The script extracts 12 acoustic measures related to fundamental frequency (f0), intensity, spectral tilt and Long Term Average Spectrum (LTAS).

Statistical procedures

In order to correlate the quantitative and qualitative variables, the statistical factorial method called Multiple Factor Analysis (MFA) has been applied with FactorMInerR (Husson et al, 2013).

The MFA was used to study similarities among stimuli relative to the 34 research variables structured in the three groups studied in this work: Gc1 (FACS), Gc2 (ExpressionEvaluator measures) and Gq2 (VPA). All measures were normalized by z-score. In order to verify the correlation among the groups of variables the Pearson Lg and RV coefficients were used.

Results

The application of the MFA allows the analysis of the variables in several dimensions. The variables analyzed were displayed in five dimensions. Table 1 exhibits the contributions of the variables in these dimensions. In Dimension 1 Gc2 (ExpressionEvaluator) is more representative, in Dimension and 3 it is Gc1 (FACS) and in Dimension 4 and 5 is Gq2 (VPA).

Table 1. Significant variables in the five dimensions of analysis.

Groups	Dim. 1	Dim. 2	Dim 3.	Dim. 4.	Dim 5.
Gc1	31.5841	46.9774	47.8636	15.4637	34.8319
Gc2	38.0663	10.7865	17.1027	34.1301	21.8097
Gq2	30.3496	42.2361	35.0337	50.4063	43.3584

In Dimension 1, Disgust was the main important factor influencing the division between the two groups. In Dimension 2, it was Contempt. Artists pro

conviction at the second instance did not display disgust or contempt but the ones contra conviction did. These findings can be interpreted as the artists pro conviction at second instance were happy with the Supreme Court decision and/or angry against impunity while the group of artists contra conviction at second instance showed an aversive feeling against the decision, although semantically and orally tried hard to maintain neutrality.

The representativeness of the acoustic (ExpressionEvaluator) and perceptual variables (VPA) and of the emotions (FaceReader) can be seen in the graphic displayed in Figure 2. The most representative emotional variables are (Disgust, Sadness and Contempt). Neutrality is also a strong factor. This is accordingly to the idea it was important for the artists to be persuasive without spoiling the appealing effect of the message they wanted to convey. The most representative acoustic measures are f0 related.

Figure 2. Acoustic, emotion and vocal quality variables.

In Dimension 1, significant correlations (p=0) were found between VPA mandibular and labial settings and Disgust (98.3). In Dimension 2, "Sadness and Contempt" were correlated to the VPA variables "Open Jaw" and "Lip Spreading" to happiness, the former characterizing the subjects contra conviction at second instance and the latter the ones pro conviction. In terms of interpreting these meaning effects, "Open Jaw" can be associated to screaming while "Lip Spreading" characterizes smiles. The use of these two settings is thought to be accordingly to the interpretation of the intended meanings conveyed by the persuasive kind of messages conveyed by the two groups: complaint against happiness.

In Table 2, the Lg Pearson coefficient is provided. The Gc1 (FACS) and the Gq2 (VPAS) are the most representative as indicated by the MFA coefficient. The Gq2 (VPAS) e Gc1 (FACS) groups are related (RV=0,81) and Gq1 é the one which is closer to the other groups (RV=0,94).

Table 2. LG Pearson coefficient values.

Groups	Gc1	Gc2	Gq2	MFA
Gc1	2.5342	1.2434	1.9847	2.2332
Gc2	1.2434	1.5271	1.4019	1.6170
Gq2	1.9847	1.4019	2.3686	2.2304
MFA	2.2332	1.6170	2.2304	2.3566

Conclusion

The analysis of the video excerpts revealed that the non-verbal features, were influential in separating the two groups analysed. Findings provided pieces of evidence for meaning effects not possible to be interpreted without the use of multidimensional analysis. This was thought to be fundamental to disclose the paralinguistic meanings of the persuasive discourse.

References

Barbosa, P. A. 2009. Detecting Changes in Speech Expressiveness in participants of a radio program. Proceedings of Interspeech, Brighton, UK, 1, 2155-2158.

Husson, F., Josse, J., Lê, S., Mazet, J. 2013. FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R (version 1.25) [R package].

Laver, J., Mackenzie-Beck J. 2007. Vocal Profile Analysis Scheme. Edinburgh: Queen Margareth University College, Speech Science Research Centre.

Madureira, S., Fontes, M.A.S. 2019. The analysis of facial and speech expressivity: tools and methods. In J.M. Lahoz-Bengoechea, R. Pérez Ramón (Eds.), Subsidia. Tools and resources for speech sciences (pp. 19–26). Málaga: Universidad de Málaga.