Reciting a sonnet: production strategies and perceptual effects

Sandra Madureira

LIAAC, Department of Linguistics, Catholic University of São Paulo, Brazil

madusali@pucsp.br

Abstract

This paper examines, from a dynamically-oriented approach, speaking strategies used by two professional speakers in reciting a sonnet.

Spectrographic and perceptual analysis of prosodic and segmental elements are carried out. Uses of speaking strategies and their effects are contrasted to discuss relations between sound and sense.

Index Terms: sound and sense, speech expressivity, acoustic phonetic analysis, prosody, speaking styles.

1. Introduction

The objective of this paper is to focus on the analysis of speaking strategy choices used by two speech professionals in reciting a sonnet in order to discuss how relations between sound and sense are built up.

It examines the use of prosodic and segmental elements as employed by these two subjects in reciting a widely known Brazilian Portuguese sonnet and presents an analysis of their reciting styles. Considering relations between sound and sense requires referring to affective information, defining speaker roles, adopting a conceptual framework of speaking style and a phonetically-oriented approach which allows considering physical and symbolic properties of sound based on experimental analysis of speech samples.

Relying on acoustic instrumental analysis is crucial from production and perception points of view since it enables making inferences about the patterns of movements of the articulators and the vocal folds in the production of speech sounds, besides allowing to deal with the phonetic detail which characterizes speech.

In the present study, our analysis departs from some theoretical presuppositions, among them, that the speaker can be viewed as the animator (speaking machine), the author and the principal as [1] postulates, and that in building up stylistic strategies the speaker in his role of animator actively relates forms and meanings in such a way as proposed by [2] in his theory on style.

For the perceptual analysis of voice qualities, the Voice Vocal Profile Analysis Scheme (VPAS) developed by [3] is used and for analyzing the physical properties of speech, an acoustic phonetic analysis couched in a dynamic approach is carried out in an attempt to deal with the fact that, as [4] points out, speech not only informs but impresses listeners.

The fact that speech impresses and not only informs has to do with the potentiality of speech forms to express meanings related to processes of sound symbolism [5] and sound metaphors [6].

Principles from a dynamical approach to speech production and perception guided the analysis. From the point of view of speech production, it means that speech can be specified by classes of gestures taken as a set of movement patterns oriented towards a specific goal [7,8] and from the point of view of speech perception, it means that an active process which anticipates the sensory consequences of an action because it departs form the idea that sensory and motor patterns are congruent [9].

2. Methodological procedures

Commercially available recordings on CD of the "Soneto da Fidelidade" (Sonnet on Faithfulness) written by Vinicius de Moraes, a Brazilian diplomat, composer, song interpreter and poet recorded by an actor , identified in this work as S1 and an actress, identified as S2.

2.1 Corpus

Soneto de Fidelidade (Sonnet on Faithfulness)

De tudo ao meu amor serei atento
(I will be attentive to my love in everything)
Antes, e com tal zelo, e sempre, e tanto
(First and foremost, and always and with so much caring)
Que mesmo em face do maior encanto
(That even faced with the greatest fascination)
Dele se encante mais meu pensamento.
(It will be thy love captivating my attention.)

Quero vivê-lo em cada vão momento
(I want to feel love in each vain moment in my life)
E em seu louvor hei de espalhar meu canto
(And in praising love, my chant will spread far and wide)
E rir meu riso e derramar meu pranto
(And I will laugh my best laugh and shed my tears)
Ao seu pesar ou seu contentamento.
(According to my love's sorrow or delight)

E assim, quando mais tarde me procure
(And when, later on, there comes in search of me)
Quem sabe a morte, angústia de quem vive
(Either death, the anxiety of the living ones)
Quem sabe a solidão, fim de quem ama
(Or loneliness, the end of the loving ones)

Eu possa me dizer do amor (que tive):
(I can tell myself about the love (I had at one time)
Que não seja imortal, posto que é chama
(Let it not be immortal, since it is flame)
Mas que seja infinito enquanto dure.
(But let it be infinite while lasting through time)

2.2 Methodological procedures

The following analysis procedures were carried out: perceptual evaluation, phonetic acoustic analysis and measurement of f0 values in syllabic rhymes and duration values of speech segments, syllables, V_V units (from vowel onset to vowel onset, consonant and silent pauses in this interval included).

Measurements of V-V units were taken into account due to their relevance as a psychoacoustic anchorage point as indicated in the phonetic literature on p-centers [10].

Measurements of acoustic parameters were made by means of Praat. From the duration values obtained, the mean, standard deviation, z-score and smoothed z-score were calculated by the SG detector, a software developed by Plinio Barbosa from the University of Campinas whose script is transcribed in [11].

Voice qualities, pitch accents, use of allophones; rhythmical and speech rate variations and distribution of silent pauses were analyzed perceptually and based on inspection of acoustic data. Voice qualities were further classified by means of the VPAS.

To analyze the distribution of pauses, the notion of attractor was used. In this paper, the word "attractor" is used in the sense of a *locus* towards which some kind of behavior tends. The end of the stanza , the end of a verse or the boundaries between syntactic constituents in a verse, for instance, were considered to attract the placement of silent pauses, since speakers tended to produce silent pauses at these *loci*.

Four kinds of attractors were taken into account: pragmatic (PA); discursive (DA); syntactic (SA) and morph-syntactic (MSA). The ends of the stanza and of the verse, for instance, were considered discursive attractors and the boundaries between syntactic constituents a syntactic attractor.

Qualitative judgments on the performance of the two subjects were asked from a group of 30 adults. A protocol was built up for evaluation purposes. The judges listened to the recordings in a quiet room and filled in the protocol forms. The protocol had two items. The first was a question to investigate the overall impression the speaker's interpretation caused on the listener. The second contained scalar degrees varying from 0 (absence) to 7 (presence in the highest degree) meant to be used to evaluate the conditions described by the following descriptors: sadness, anguish, softness and enthusiasm.

3. Results and Discussion

The presentation of the analysis data is divided into parts according to the phonetic event examined: silent pauses, spectral characteristics; f0 contours and duration contours. In the analysis of these events, correlations between sound and sense are discussed.

3.1 Analysis of silent pauses

The distribution of silent pauses by the subjects was described according to kinds of attractors mentioned in the methodological procedures: pragmatic (PA);

discursive (DA); syntactic (SA) and morph-syntactic (MSA).

The strongest attractor was the discursive one: both subjects placed a silent pause at the end of the stanza and S2 introduced a silent pause at the end of each verse. Syntactic attractors, such as coordinate constituents, phrases and subordinate clauses, were also strong since they were often separated by silent pauses.

S1 produced less silent pauses (12) than S2 (24). S1 took 42 ms and S2 68 ms to recite the poem. S1 used mainly breathing pauses while S2 used pauses as a stylistic speaking strategy.

The fact that S2 uses a great number of pauses has the effect of creating a feeling of expectancy. The duration of twelve of the silent pauses she produced were longer than the utterances which followed them and only 7 silent pauses were shorter than the utterances which followed them. Otherwise, S1 employed fewer pauses, and that had the effect of adding sonority and establishing coherence within utterances.

The duration in ms of all speech utterances and silent pauses produced by the subjects (S1's in Figure 2 and S2's in Figure 3) are displayed.
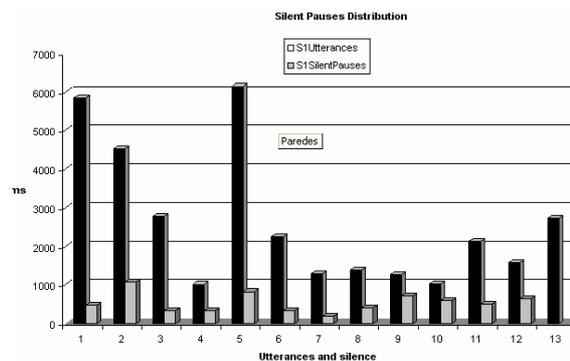

Figure 2. Duration in ms of S1's speech utterances in black and duration in ms of silent pauses in a light shade of gray.
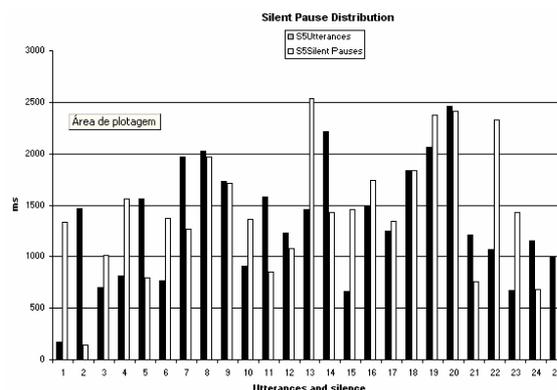

Figure 3. Duration in ms of S2's speech utterances in black and duration in ms of silent pauses in a light shade of gray.

No silent pause is produced by S1 in the final verse of the poem. On the other hand, S2 introduces silent long pauses after "mas" (1047 ms) and "infinito" (674 ms).

The absence of the silent pause in that verse can be thought as highlighting the sense of conflict between "finity and infinity" alluded in the utterance. The

presence of pauses, separating the two semantic constituents, has a weakening effect: the conflict is mitigated.

3.2 Analysis of spectral characteristics

Congruence between sense and meaning was found in speaking strategies based on sound symbolism such as the one used by S2 (a female subject) in producing the word "chama" (flame). Her use of the voice quality setting of *whispery creaky voice* (Figure 4) turned out to emphasize the movement and the instability of the flame.
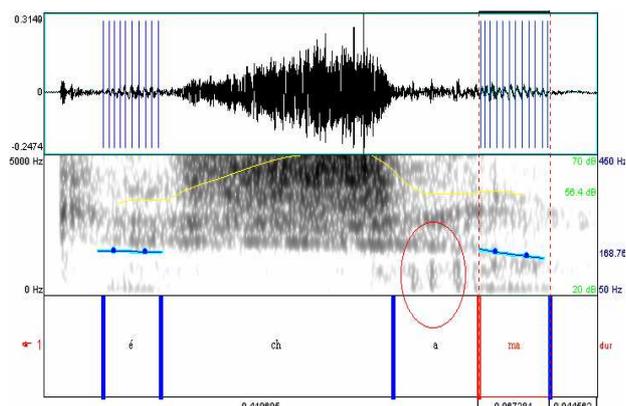

Figure 4. The circle refers to the vowel of the stressed syllable "cha" in the word "chama" (flame). The setting of creaky voice (VPAS) was found to characterize the S2's production.

S2 used *whispery voice* as a kind of permanent setting and *whispery creaky voice* as a kind of intermittent setting during the production of the sonnet and S1 used the setting of modal voice.

S2 placed a silent pause after producing the word "infinito", which caused the last syllable of the word to be lengthened (prepausal lengthening). Otherwise, S1 lengthened the first syllable of the word, whose rhyme has a continuant phonetic segment (a nasalized vowel) which was symbolically explored by the actor to emphasize infinity.

3.3 Analysis of f0 contours

Analysis of f0 contours were based on the extraction of one value in syllabic rhymes and from 3 to 4 values on pitch accented vowels.

F0 contours of the utterance " e com tal zelo e sempre e tanto" (and with care and always and so much) as produced by S1 displayed a continuous rising towards the end of the intonation group contrasting with the local rising on the word "tanto" (so much) produced by S2. The effects of these strategies are quite diverse, since climax is better achieved with the continuous rising trajectory.

S1 varied intonation patterns as an expressive speaking strategy but S1 did not. In coordinated phrases and sentences, f0 was used by S1 as a cohesive element since no resetting was observed in the beginning of these constituents and f0 contour rose or lowered continuously towards a target word. Pitch accents were placed on highly relevant words from the information point of view.

The placement of f0 peak alignment can also be considered an important cue in signaling expressiveness [12]. Some sound meaning correlations can be made, contrasting the trajectories of the f0 contours in the productions of the word "infinite" (stress on the penultimate syllable), a key word in the sonnet. In S2's production f0 peak is aligned with "in" and with "ni" in S1's.

The continuous rising to the stressed syllable as in S1's production can be interpreted as emphasizing the semantic content of the word "infinito" and the continuous falling from the first syllable of the word up to the last one in S2's production can be interpreted metaphorically as something that fades away.

S1 produced the word "infinito" with an L+ H* pitch accent while S2 used an H pitch accent associated to the first syllable of "infinito" followed by an L-L% boundary tone. The impressive effects are quite different. In S1's production the meaning of the word "infinite" is reinforced.

3.4 Analysis of duration contours

The contours based on smoothed z-score values show clearly outlined peaks. In Figure 5, contours based on absolute values of f0 and duration are displayed.
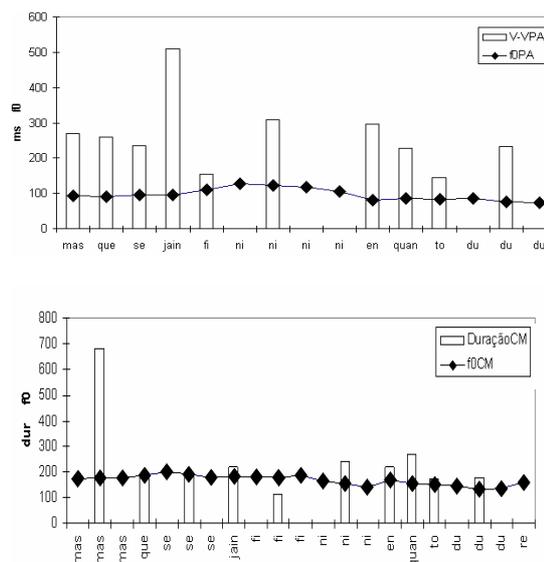

Figure 5 - f0 and duration contour of S1's production of the verse "mas que seja infinito enquanto dure" (upper graphic) and f0 and duration contour of S2's production of the same verse (lower graphic).

In Figures 6 and 7, the contours based on smoothed z-score values are displayed. Both refer to the productions of the verse "mas que seja infinito enquanto dure" (but should love be infinite while it lasts), as mentioned in the preceding paragraph.

In S1's production (Figure 6) of "mas que seja infinito enquanto dure", duration values culminate on the first syllable of the word "infinito" and the f0 peak is aligned with the stressed syllable "ni".

In S2's production (Figure 7) duration values culminate on the first V-V unity and f0 peak in the first syllable of the word "infinito".
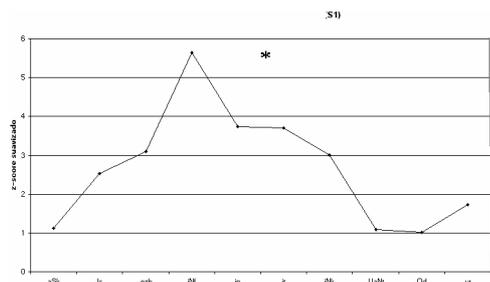
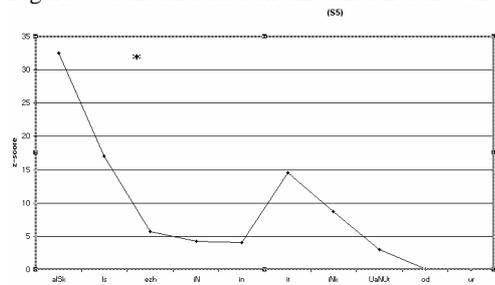Figure 6- Contour based on smoothed z-score values.

Figure 7- Contour based on smoothed z-score values.

3.5 Perceptual evaluation test

In answering question one, which concerned overall impression caused by the speakers' oral interpretation, judges employed qualifiers such as "splendorous and gorgeous" when they listened to S1's performance and "anxiety and grief" when they listened to S2's performance.

The results of the listeners' evaluation of question two of the protocol can be summarized as: - average values for S1: sadness (2.87); anguish (3.07); softness (3.13) and enthusiasm (4.67); average values for S2: sadness (4.67); anguish (3.97); softness (4.57) and enthusiasm (2.27)

Results indicate S1's interpretation was considered mainly enthusiastic while S2's was judged mainly as sad and soft. These judgments are influenced by the speaking strategies used productively by the subjects.

S1 varied speaking strategies (salient pause distribution, pitch range, speech rate and kinds of speech patterns) a lot and often produced a climax effect (Bolinger,1986) while S2 used silent pauses, settings of *whispered voice and whispery creaky voice* and a narrow pitch range (usually from 140 Hz to 220 Hz) productively, which may have influenced the judges to report feelings of sadness and softness in a higher degree.

In spite of the content of the text being the same, the subjects' speaking strategies differed a lot. Their prosodic and segmental choices reflect their interpretation of the text, and as sound and sense are functionally related, the impressive effects on the listeners are quite different.

4. Conclusion

Reciting a poem requires a previous activity: interpreting the meaning of the text, or in other words, producing meanings. The process of reading is a collaboration between the reader and the author and meanings are actively produced by the reader and reflect his background (knowledge, life experiences, among others).

In choosing prosodic and segmental elements to express the interpreted meanings, the speaker makes correlations between sound and meaning and these can be analyzed as a source of building up knowledge about the way speech expressiveness is constructed.

In the present work, S1 and S2, two gifted, prized and famous stage professionals adopted quite different stylistic speaking strategies. Their performances are based on their interpretation of the sonnet and reflect their professional history, their sociolinguistic backgrounds, their personal life and their knowledge of the poet and his life.

In order to impress the listeners, the speaker in his role as animator, uses prosodic and segmental variations to make his speech effective from the communication point of view. The kind of animation the speaker performs reveals his productions of meanings. His speaking strategies are intended to impress the listeners and they are used to emphasize certain semantic features and to create sound metaphors. In this way, sound and sense are related by the work of the speaker and this work constitutes his style.

Comparing strategies used by professional actors in reciting poems, examining their effects on listeners and checking their text meaning production and performance intents is thought to be a promising undertaking since correlations between sound and sense can be made and productivity of strategies evaluated.

In this work, the speakers' prosodic choices (use of voice qualities and intonation patterns mainly) were found to be quite divergent and consequently impressed the listener in dissimilar ways.

5. References

[1] Goffman, E. Forms of talk, University of Pennsylvania Press, 1981.
[2] Granger, G. G. Filosofia do Estilo, EDUSP,1974.
[3] Laver, J. "Phonetic evaluation of voice quality", Kent R. D., Ball M. J. Voice quality measurement, Singular Thomson Learning, 37-48, 2000.
[4] Bolinger, D. Intonation and Its Parts. Stanford University Press, 1986.
[5] Hinton, L.; Nichols, J. Ohala, J. (eds) Sound Symbolism, Cambridge University Press, 1995.
[6] Fonagy, I. La Vive Voix: Essais de psycho-phonétique, Payot, 1983.
[7] Browman C. and Goldstein L. Towards an articulatory phonology. Yearbook, 3, 219-252, 1989.
[8] Albano, E. O gesto e suas bordas: esboço de uma teoria acústico-articulatória do português brasileiro. Editora Mercado das Letras, 2000.
[9] Berthoz, A. The Brain Sense of Movement, Harvard University Press, 2000.
[10] Pompino-Marschall, B. "On the psychoacoustic nature of P-center phenomenon", Journal of Phonetics, 17,175-192, 1989.
[11] Barbosa, P. A. Incursões em torno do Ritmo da Fala, Pontes Editores, FAPESP, Campinas, 2006.
[12] Kohler, K. J. "Paradigms of experimental prosodic analysis: from measurement to function", in: Sudhoff, S. et al. (eds) Methods in Empirical Prosody Research Language and Cognition v. 3), Walter de Gruyter, 2006.