



sonoridades

sonorities

Sandra Madureira (org.)

LIAAC - PUCSP

Apoio:  
Plano de Incentivo à Pesquisa (PIPEq)  
Pontifícia Universidade Católica de São Paulo

Sonoridades

Sonorities

*Sonoridades: a expressividade da fala, no  
canto e na declamação*

*Sonorities: expressivity in speech, singing,  
and reciting*

Capa e editoração: Jayme Preto

Edição: Beatriz Gabriel

Revisão: Aline C. Sacchi, Andréa S. Souza, Astrid M. M. Ferreira, Cristiane Magacho, Juliana A. Lomba, Márcia Polaczek, Piedade S. Cóstola, Marta F. Tanuri, Nathália Reis e Renata C. Vieira

#### Comitê Científico

Hernán Martínez (Universidade de Los Andes, Mérida, Venezuela)

Lenny Kyrillos (Pontifícia Universidade Católica de São Paulo)

Marcello Marcelino Rosa (Universidade Federal de São Paulo)

Maria Fabiana Bonfim de Lima Silva (Universidade Federal da Paraíba)

Nair Katia Nemr (Universidade de São Paulo)

Pablo Arantes (Universidade Federal de São Carlos)

Regina Yu Shon Chun (Universidade Estadual de Campinas)

Waldemar Ferreira Netto (Universidade de São Paulo)

S414

Sonoridades [recurso eletrônico]: a expressividade na fala, no canto e na declamação / Sonorities [eletronic device]: speech, singing and reciting expressivity / Sandra Madureira (Organizadora). - São Paulo: Pontifícia Universidade Católica de São Paulo, 2016.

181p. : il

1 recurso eletrônico: E-PUB

ISBN 978-85-60453-32-0

(Pontifícia Universidade Católica de São Paulo)

1. Fonética. 2. Fala. 3. Voz. 4. Fonética acústica. 5. Fonoaudiologia. I. Sandra Madureira. II. Título.

CDD – 414

Edição – Pontifícia Universidade Católica de São Paulo



Laboratório Integrado de Análise Acústica e Cognição  
PEPG em Linguística Aplicada e Estudo de Linguagem  
Pontifícia Universidade Católica de São Paulo

PUC-SP

[www.pucsp.br/liaac](http://www.pucsp.br/liaac)

# Sumário

Relação de Autores.....	1
Introdução.....	2
Resumos dos Capítulos.....	5
As Palavras Cantada, Falada e Declamada: Um Estudo Sobre a Relação Entre Canto, Fala e Declamação e seus Efeitos Impressivos.....	17
<i>Juliana Andreassa da Lomba, Sandra Madureira, Mario Augusto de Souza Fontes</i>	
Expressão Vocal de Emoções: Metáfora Sonora, Fala e Canto .....	31
<i>Glaucia Laís Salomão</i>	
Efeitos das Metáforas no Ensino do Canto: Dados de Imagem do Trato Vocal.....	44
<i>Cristina Canhetti, Zuleica Camargo</i>	
Metáforas no Ensino de Canto e seus Efeitos na Qualidade Vocal: um Estudo Acústico-Perceptivo.....	62
<i>Paulo Menegon, Sandra Madureira</i>	
Gestos Faciais e Corporais e os Tons Alto e Baixo: Qual a Relação? .....	90
<i>Vera Pacheco, Marian Oliveira</i>	
Os Papéis das Prosódias Vocal e Visual na Expressão de Emoções na Fala.....	99
<i>Mario Augusto de Souza Fontes</i>	
Aspectos Teóricos para Elaboração de Programa de Treinamento da Performance Comunicativa.....	111
<i>Leticia Correa Celeste, Aveliny Mantovan Lima-Gregio</i>	
Dez Tópicos Sobre Expressividade Oral para o Fonoaudiólogo: uma Proposta de Debate.....	119
<i>Izabel Viola, Léslie Piccolotto Ferreira</i>	
A Fala do Intérprete.....	137
<i>Layla Penha</i>	
On the Varying Reception of Speakers Expressivity Across Gender and Cultures, and Inference in their Personalities.....	149
<i>Albert Rilliard, Donna Erickson, João De Moraes, Takaaki Shochi</i>	
Scripts for the Acoustic Analysis of Speech Data.....	164
<i>Plínio A. Barbosa, Zuleica A. Camargo, Sandra Madureira</i>	

# *Relação de Autores*

**ALBERT RILLIARD**

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI), CNRS, Université Paris-SACLAY, F-91405, Orsay, France

**AVELINY MANTOVAN LIMA-GREGIO**

Universidade de Brasília

**CRISTINA CANHETTI**

Support Fonoaudiologia

**DONNA ERICKSON**

Kanazawa Medical University, Sophia University, Japan

**GLAUCIA LAÍS SALOMÃO**

Stockholm University

**IZABEL VIOLA**

Grupo de Pesquisa Voz: avaliação e intervenção fonoaudiológica - PUC/SP

**JOÃO ANTONIO DE MORAES**

Universidade Federal do Rio de Janeiro

**JULIANA ANDREASSA DA LOMBA**

Pontifícia Universidade Católica de São Paulo

**LAYLA PENHA**

Pontifícia Universidade Católica de São Paulo

**LÉSLIE PICCOLOTTO FERREIRA**

Pontifícia Universidade Católica de São Paulo

**LETÍCIA CORREA CELESTE**

Universidade de Brasília

**MARIAN OLIVEIRA**

Universidade Estadual do Sudoeste da Bahia

**MARIO AUGUSTO DE SOUZA FONTES**

Pontifícia Universidade Católica de São Paulo

**PAULO MENEGON**

Faculdades Metropolitanas Unidas

**PLÍNIO A. BARBOSA**

Universidade Estadual de Campinas

**SANDRA MADUREIRA**

Pontifícia Universidade Católica de São Paulo

**TAKA AKI SHOCHI**

CLLE-ERRSAB UMR 5263, France  
LABRI UMR 5800

**VERA PACHECO**

Universidade Estadual do Sudoeste da Bahia

**ZULEICA CAMARGO**

Pontifícia Universidade Católica de São Paulo

# INTRODUÇÃO

## *Sonoridades: a expressividade na fala, no canto e na declamação*

O título deste e-book “Sonoridades” remete a questões de expressividade da linguagem oral em três formas de expressão oral: a fala, o canto e a declamação. No cerne da expressividade oral figura o simbolismo sonoro que alude ao potencial que a matéria fônica tem de expressar sentidos, veicular características físicas, fisiológicas, psicológicas e sociais do falante e impressionar o ouvinte, por isso, os sons da fala podem ser qualificados como expressivos e impressivos (Madureira, 2016).

As relações som sentido remetem a múltiplas questões: o som tem sentido e o sentido tem som? (Albano, 1988); o som é sentido?; o som impressiona os sentidos? podemos atribuir características de forma, cor, textura, tamanho, comprimento e força ao som? O som constitui paisagens sonoras (Schaffer, 1992)? O som veicula emoções e atitudes?

Uma só palavra pode ser utilizada para responder às perguntas acima colocadas: “sim”. Tal assertividade nas respostas decorre do fato de que nossos ouvidos estão abertos ao som. Como se expressa Schaffer (1992:288) “Não existem pálpebras nos ouvidos. Os ouvidos de uma pessoa verdadeiramente sensível estão sempre abertos”. Abertos a que?, perguntamos. Abertos aos sons que vivenciamos - que estão nas paisagens sonoras que frequentamos; aos sons que imaginamos - que estão em nossas mentes pelas nossas experiências de som permeadas pelos nossos sentidos; aos sons que vemos nos espectros, espectrogramas, formas de onda e traçados de f0 e intensidade; aos silêncios que soam.

Segundo Tsur (1992) a potencialidade das características dos sons da fala impressionarem os ouvintes tem bases intersubjetivas nos níveis acústicos, fonéticos e fonológicos da estrutura sonora da língua. A associação de certos tipos de segmentos e prosódicos para expressar certos efeitos de sentido, tais como fricativas para movimentos rápidos de um objeto no ar e laringealização para expressar angústia, usos produtivos de sons que são comuns em outros contextos da língua e relaxamento de restrições que em outros contextos da língua são fortes, além de reduplicações evidenciam domínios de aplicação do simbolismo sonoro. É graças ao simbolismo sonoro que atribuímos efeitos de sentido ao que é falado.

Discutir as relações entre som e sentido implica em reconhecer a potencialidade do simbolismo sonoro, ou seja, do vínculo direto entre som e sentido, que é a base da expressividade sonora. Esse vínculo pode ser interpretado, segundo Madureira (2011), a partir da existência de códigos biológicos de base acústica, fonatória e articulatória: o código de frequência (Ohala, 1983; Ohala, 1984; Chuenwattanapranithi, 2008); os códigos de esforço e produção/respiratório (Gussenhoven, 2002, 2004). Acrescente-se a esses o código da sereia (Gussenhoven, 2016).

Na condução de trabalhos sobre a expressividade oral, no âmbito do Laboratório Integrado de Análise Acústica e Cognição da Pontifícia Universidade Católica de São Paulo (Liaac-Pucsp), destacamos os que levam em conta a análise dos parâmetros acústicos e perceptivos de ajustes da qualidade vocal (Madureira, Fontes E Fonseca, 2016); Fontes E Madureira, 2015; Madureira E Camargo, 2010; Viola E Madureira, 2007 A, B; Madureira, 1991, 1996, 2004, 2008, 2011, 2016).

Para a análise dos parâmetros acústicos o script desenvolvido por Barbosa (1999) tem sido utilizado e para a análise dos parâmetros perceptivos o VPAS, desenvolvido por Laver e Mackenzie-Beck (2007) em sua versão adaptada para o português por Camargo e Madureira (2008). Esses instrumentos mostraram-se de muita relevância para a realização das análises efetuadas. Referências a trabalhos sobre qualidade vocal com dados clínicos na vertente clínica são encontrados

no e-book intitulado “Fonética Clínica: vinte anos de LIAAC” organizado por Zuleica Camargo em 2016.

Este livro reúne 11 capítulos que tratam de questões sobre a expressividade vocal e visual e trazem contribuições para a reflexão sobre a viva voz (Fonagy, 1983) evocar sentidos, sobre como trabalhar a voz para expressar sentidos e como analisar a expressividade na fala, no canto e na declamação.

*Sandra Madureira*

## **Referências Bibliográficas**

ALBANO, E. C. Fazendo sentido do Som. **Ilha do Desterro**, Santa Catarina: Universidade Federal de Santa Catarina, 19: 1, 11-26, 1988.

BARBOSA, P. A. Detecting changes in speech expressiveness in participants of a radio program. **Proceedings of Interspeech**. Brighton, United Kingdom, 1, 2155-2158, 1999.

CAMARGO, Z.; MADUREIRA, S. Voice quality analysis from a phonetic perspective: voice profile analysis scheme profile for Brazilian Portuguese (BP-VPAS). **Proceedings of the Fourth Conference on Speech Prosody**, CAMPINAS: UNICAMP, CAPES, FAPESP, CNPq, 57-60, 2008.

CHUENWATTANAPRANITHI, S.; XU, Y.; THIPAKORN, B. & MANEEWONGGVATANA, S. Encoding emotions in speech with the size code: a perceptual investigation. **Phonetica**, 65, 210-230, 2008.

FONAGY, I. **La vive voix**. Paris: Payot, 1983.

FONTES, M. A. S.; MADUREIRA, S. Gestural prosody and the expression of emotions: a perceptual and acoustic experiment. **Proceedings of the 18th International Congress of Phonetic Sciences**, Glasgow: University of Glasgow, 1-5, 2015. GUSSENHOVEN, C. Foundations of Intonational Meaning: Anatomical and Physiological Factors. **Topics in Cognitive Science** 8, Wiley Periodicals, 425–434, 2016.

GUSSENHOVEN, C. **The Phonology of Tone and Intonation**, Cambridge: Cambridge University Press, 2004.

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

LAVER, J. & MACKENZIE-BECK J. **Vocal profile Analysis Scheme – VPAS**. Edinburgh: Queen Margareth University College- QMUC, Speech Science Research Centre, 2007.

MADUREIRA, S. Intonation and variation: the multiplicity of forms and senses. **Dialectologia**, Special Issue VI. 54-764, 2016.

\_\_\_\_\_ The Investigation of Speech Expressivity. In H. Mello; A. Panunzi; T. Raso (Eds) **Ilocution, modality, attitude, information patterning and speech annotation**. Firenze: Firenze University Press, 101-118, 2011.

\_\_\_\_\_ A expressão de atitudes e emoções na fala. In: Leny Kirillos. (Org.). **Expressividade**. São Paulo: Revinter, 15-25, 2004.

\_\_\_\_\_ Reciting a sonnet: production strategies and perceptual effects. In: **Proceedings of the Speech Prosody 2008 Conference**. São Paulo: Editora RG, 697-700, 2008.

\_\_\_\_\_ A materialidade fônica, os efeitos de sentido e os papéis do falante. **DELTA**, Sao Paulo, v. 12, n.1, 87-93, 1996.

\_\_\_\_\_ A qualidade de voz como recurso fonoestilístico. **Intercâmbio**. São Paulo, v. 1, p. 99-120, 1991.

- MADUREIRA, S.; FONTES, M. A. S. ; FONSECA, B. C. Voice quality and speaking styles. **Dialectologia**. Special Issue, v. VI, p. 171-190, 2016.
- MADUREIRA, S.; CAMARGO, Z. A. . Exploring sound symbolism in the investigation of speech expressivity. In **Proceedings of the Third ISCA Conference**, Athens, 105-108, 2010.
- OHALA, J. J. An ethological perspective on common cross-language utilization of F0 of voice. **Phonetica** 41: 1-16, 1984.
- \_\_\_\_\_The origin of sound patterns in vocal tract constraints. In MacNeilage, P. F. (ed), **The Production of Speech**, 189-216. New York: Springer, 1983.
- SCHAFER, M. **O ouvido pensante**. São Paulo: Unesp, 19-66, 1992.
- VIOLA, I. C.; MADUREIRA, S. A qualidade vocal como elemento expressivo. In: Aragão, M. S.S.; Silveira, R. C.. (Org.) **Diversidade de estudos em Fonética e Fonologia no Brasil**. João Pessoa - PB: ANPOLL 2007, 78-86, 2007a.
- VIOLA, I; MADUREIRA, S. Voice quality and Speech Expressiveness. **Proceedings of the 16th International Conference on Phonetic Sciences**. University of Maryland Working Papers in Linguistics. Saarbrucken: DFKI Publications, 65-70, 2007b.

# *Resumos dos Capítulos*

# ***AS PALAVRAS CANTADA, FALADA E DECLAMADA: UM ESTUDO SOBRE A RELAÇÃO ENTRE CANTO, FALA E DECLAMAÇÃO E SEUS EFEITOS IMPRESSIVOS***

JULIANA ANDREASSA DA LOMBA  
SANDRA MADUREIRA  
MARIO AUGUSTO DE SOUZA FONTES

## **RESUMO**

Este estudo tem como objetivo realizar um experimento de percepção para investigar, por meio de uma perspectiva contrastiva, o impacto impressivo causado nos ouvintes por trechos de fala, canto e declamação do espetáculo “Totatiando” estrelado por Zélia Duncan que interpreta composições de Luís Tatit. Para a aplicação do teste de percepção sobre os efeitos impressivos desses três estilos de oralidade foi elaborado um questionário semântico com os descritores “melodioso, vibrante, envolvente e comovente”. O teste foi aplicado a um grupo de trinta juízes. Os resultados mostram como as palavras cantada, falada e declamada atingem de maneira diferenciada os ouvintes, pois a avaliação dos descritores selecionados evidenciou efeitos impressivos diversos dependendo do estilo concernido.

## **ABSTRACT**

This work aims to carry out a perception experiment to investigate, from a contrastive perspective, the impressive effects on the listeners caused by excerpts of the sung, spoken and recited recordings from the musical show play “Totatiando”, starred by Zélia Duncan who interprets musical compositions by Luís Tatit. In order to apply the perceptual test about the impressive effects of these three oral styles a semantic questionnaire containing the descriptors “melodious, vibrant, involving and moving” was elaborated. The test was applied to a group of 30 judges. The results show that the sung, spoken and recited styles affect the listeners in different ways, since the selected descriptors were evaluated differently depending on the style concerned.

# *EXPRESSÃO VOCAL DE EMOÇÕES: METÁFORA SONORA, FALA E CANTO*

GLÁUCIA LAÍS SALOMÃO

## **RESUMO**

A comunicação de emoções é crucial para as relações sociais e cumpre papel fundamental na manutenção da ordem social entre as pessoas. Neste capítulo tratamos da comunicação de emoções por meio de duas modalidades expressivas que fazem uso do som como meio de comunicação: a fala e o canto. Ao longo do texto argumentamos a favor da ideia de que a expressão vocal de emoções espelha aspectos fisiológicos associados à própria emoção expressa; de que existem várias semelhanças entre os padrões expressivos observados na fala e no canto; e de que o canto é expressivo por conter lastros de padrões expressivos da fala.

## **ABSTRACT**

The communication of emotions is crucial to social relationships and plays a fundamental role in maintaining the social order between people. In this chapter we are looking at the communication of emotions through two expressive modalities that make use of sound as a mean of communication, i.e. speech and singing. Throughout the text we argue in favor of the idea that the vocal expression of emotions reflects physiological aspects associated to the emotion itself that is expressed; that there are many similarities between the expressive patterns found in speech and in singing; and that the singing is expressive because it has traces of expressive patterns of speech.

# ***EFEITOS DAS METÁFORAS NO ENSINO DO CANTO: DADOS DE IMAGEM DO TRATO VOCAL***

CRISTINA CANHETTI  
ZULEICA CAMARGO

## **RESUMO**

O objetivo deste estudo foi investigar os efeitos do uso das metáforas no ensino do canto, a partir de descrições de imagens do trato vocal (videofluoroscopia) em uma mulher não cantora, sem alterações vocais. Gravações da canção infantil “Sapo Jururu” em quatro tarefas: sem instrução e com instruções baseadas nas metáforas prisma, catedral e disco voador foram analisadas, a partir de medidas de abertura de lábios (abertura de mandíbula, distância dorso de língua-faringe e dorso de língua-palato por meio do software Osiris. Os achados foram submetidos a tratamento estatístico. As metáforas geraram mobilizações diferenciadas no trato vocal, especialmente para catedral e disco voador e reforçaram as relações entre som e sentido.

## **ABSTRACT**

Aiming at investigating the effects of the usage of metaphors in singing voice production, used by singing teachers, from images of the vocal tract (by means of videofluoroscopy) among women, non-singers, without voice disorder. For the experiment, the recordings of the folklore song Sapo Jururu in four different tasks: without and with instructions, the latter based on metaphors (prism, cathedral and flying saucer) were analyzed, from measures of labial opening (absolute and normalized), jaw opening, distance from the back of tongue-pharynx and back of tongue-palate were extracted by means of Osiris software. The findings of the experiment was submitted to statistical treatment. The phonetic approach of the data showed that metaphors lead to differentiated vocal tract adjustments, mainly for cathedral and flying saucer methaphors, and reinforced the relation between sound and meaning.

# ***METÁFORAS NO ENSINO DE CANTO E SEUS EFEITOS NA QUALIDADE VOCAL: UM ESTUDO ACÚSTICO-PERCEPTIVO***

PAULO MENEGON  
SANDRA MADUREIRA

## **RESUMO**

O objetivo foi investigar, com base em análises perceptiva e fonético-acústica, os efeitos nos ajustes de qualidade de voz de emissões de canto motivados pela apresentação de metáforas. Os dados do ponto de vista acústico, foram analisados no PRAAT com auxílio da aplicação do *script – ExpressionEvaluatorbySG* desenvolvido por Barbosa (2009). Os ajustes de qualidade de voz foram analisados pelo VPAS – PB. A sugestão da metáfora da catedral levou os 5 sujeitos analisados a utilizarem estratégias de ampliação de cavidades de ressonância com o uso de ajuste de mandíbula aberta e do abaixamento da laringe e corpo de língua recuada.

## **ABSTRACT**

The objective is to use perceptive and phonetic-acoustic analysis to investigate the effects in voice quality adjustments made in singing emissions motivated by the introduction of metaphors. Recordings were acoustically analyzed in PRAAT, applying the script - *ExpressionEvaluatorbySG* developed by Barbosa (2009). Voice quality adjustments were analyzed by VPAS – PB. The cathedral metaphor led the subjects to use strategies to enlarge their resonance cavities, mainly by opening the mouth (Open Jaw setting), but also by means of lowering the larynx (Lowered Larynx setting) retracting the tongue (Retracted Tongue Body setting).

# ***GESTOS FACIAIS E CORPORAIS E OS TONS ALTO E BAIXO: QUAL A RELAÇÃO?***

VERA PACHECO  
MARIAN OLIVEIRA

## **RESUMO**

Durante uma conversa, nós realizamos a todo tempo movimentos com o corpo. Esses movimentos informam intenções, interesses, sentimentos de quem fala. Com base nesse pressuposto, neste artigo, discutimos a definição de gestos e defendemos que os gestos e a fala pertencem ao mesmo sistema cognitivo. Além disso, avaliamos a relação entre os gestos faciais e corporais e a realização de tons altos e baixos. Nossos resultados mostram que tons altos tendem a ser acompanhados por gestos de elevação da cabeça, dos olhos e da mão, ao passo que os tons baixos são acompanhados por movimentos de abaixamento de cabeça, dos olhos.

## **ABSTRACT**

When speaking we make movements with the body all the time. These movements inform intentions, interests, feelings of the speaker. In this article we discuss the definition of gestures and defend the gestures and speech belong to the same cognitive system. In addition, we evaluate the relationship between facial and body gestures and the high and low tones. Our results show that high tones are accompanied by lifting of the head movements, eye and hand, while low tones are accompanied by lowering of head movements and eyes.

# *OS PAPÉIS DAS PROSÓDIAS VOCAL E VISUAL NA EXPRESSÃO DE EMOÇÕES NA FALA*

MARIO AUGUSTO DE SOUZA FONTES

## **RESUMO**

Este capítulo aborda questões das prosódias vocal e visual, considerando aspectos perceptivos e acústicos e relacionando-as com a expressão de emoções na fala. Entre os elementos prosódicos na fala, a qualidade de voz em sua conceituação enquanto ajustes fonatórios, articulatórios e de tensão é destacada. Para a análise perceptiva desses ajustes o VPAS, protocolo desenvolvido a partir do modelo fonético descritivo de qualidades de voz é abordado e para a análise acústica o script ExpressionEvaluator. No enfoque da prosódia visual, as unidades de ação, AUs, do protocolo denominado Facial Action Code System (FACS), são usadas para descrever os movimentos dos músculos da face que estão associados na expressão de emoções. Trabalhos relevantes da literatura sobre a expressão vocal e facial da emoção são discutidos.

## **ABSTRACT**

This chapter discusses issues related to the vocal and the visual prosodic elements, taking into consideration perceptual and acoustic aspects and relating them to the expression of emotions in speech. Among the speech prosodic elements, voice quality, defined as phonatory, articulatory and tension settings, is privileged. For the perceptual analysis of the voice quality settings, the VPAS, a protocol based on the phonetic model of voice quality description is presented and for the purpose of acoustic analysis the script ExpressionEvaluator is introduced. As far as visual prosody is concerned, the action unities, AUs, of the Faction Action Code System (FACS) are taken into account. They are used to describe the movements of the muscles of the face which are associated in the expression of emotions. Relevant works on the literature on vocal and facial expression of emotions are discussed.

# ***ASPECTOS TEÓRICOS PARA ELABORAÇÃO DE PROGRAMA DE TREINAMENTO DA PERFORMANCE COMUNICATIVA***

LETÍCIA CORREA CELESTE  
AVELINY MANTOVAN LIMA-GREGIO

## **RESUMO**

O presente trabalho propõe uma discussão teórica sobre os aspectos da comunicação humana que estão envolvidos no discurso do tipo conferência. A discussão parte de um conceito mais específico de comunicação, que remete ao desempenho da transmissão de informações entre indivíduos de maneira mais formal e em torno do desenvolvimento pessoal da oratória. Nesse contexto, assumimos o pressuposto de que a comunicação abrange, necessariamente, estados mentais, intenção e elementos comunicativos. Destacamos os recursos verbais e não verbais que incluem os aspectos vestuário, expressão facial e corporal, fala e voz. Cada um desses aspectos é discutido separadamente, considerando a perspectiva de elaboração de um treinamento das habilidades comunicativas. Sugerimos que o treinamento seja elaborado para grupos pequenos, exclusivamente prático e de curta duração, a fim de que seja de fácil compreensão e aproveitamento para os participantes. Ao final, apontamos a importância da avaliação da performance comunicativa dos participantes e uma sequência de tópicos que deve compor o treinamento, a saber: conhecer o grupo, levar o grupo a se conhecer, executar o treinamento e realizar *feedback*.

## **ABSTRACT**

This work discusses from a theoretical point of view some of the aspects concerning human communication which are involved in the conference- type of discourse. The discussion departs from a very specific communication concept, which refers to the performance related to the exchanging of information between individuals in a more formal way and to the personal development of the oratory. In this context, the presupposition that communication includes, necessarily, mental states, intentions and communicative elements is taken into account. Special emphasis is given to verbal and non-verbal resources, which include clothing, facial and body expression, speech and voice. Each of these aspects is discussed separately considering the perspective of the elaboration of a training program developed to foster communicative abilities. The training is suggested to be applied to small groups, wholly practical and short-term, so that it is easily understood and efficiently used by the participants. At the end of the chapter, we highlight the importance of evaluating the communicative performance of the participants and a sequence of topics which are integrant part of the training, that is , knowing the group, helping the members of the group to know each other, doing the training and giving feedback.

# ***DEZ TÓPICOS SOBRE EXPRESSIVIDADE ORAL PARA O FONOAUDIÓLOGO: UMA PROPOSTA DE DEBATE***

IZABEL VIOLA  
LÉSLIE PICCOLOTTO FERREIRA

## **RESUMO**

A elaboração deste capítulo partiu da evidência de que o estudo e a pesquisa sobre a expressividade oral, assim como sobre os gestos corporais, subsidiam o trabalho realizado pelo fonoaudiólogo com os profissionais da voz. Dessa forma, dez perguntas que frequentemente são feitas por fonoaudiólogos foram respondidas para permitir maior reflexão sobre o tema: o que é expressividade? Qual termo seria mais indicado para definir o trabalho do fonoaudiólogo na área: expressividade oral, vocal ou verbal? Pode-se dizer que um sujeito é mais ou menos expressivo? O que vem a ser ênfase? A qualidade de voz pode ser considerada um parâmetro para marcar a expressividade? Qual a diferença entre ritmo e taxa de elocução? Como compreender as pausas a serviço da expressividade? A articulação pode ser vista como marcador de expressividade? O que observar e como proceder na avaliação da expressividade oral? e finalmente Qual é a diferença entre oratória e expressividade oral?

## **ABSTRACT**

The idea of writing of this chapter departed from the evidence that the study and the research about oral expressivity, as well as about the body gestuality, are considered relevant to the work carried out by the speech therapists concerning voice professionals. That way, ten questions, which are frequently made by speech therapists, were answered in order to allow a wider discussion about the topic: "What is expressivity? What kind of term would be more adequate to define to the speech therapists' work related to vocal and verbal expressivity? Is it adequate to say that the subject is more or less expressive? What is the meaning of emphasis? Is voice quality a parameter to be taken into account when identifying expressivity? What is the difference between rhythm and speech rate? How can pauses be used to add to speech expressivity? How can oral expressivity be evaluated? And as a last question "What is the difference between oratory and speech expressivity?"

# *A FALA DO INTÉRPRETE*

LAYLA PENHA

## **RESUMO**

Este capítulo tem como objetivo considerar a investigação dos elementos prosódicos pausa, entoação e acentuação na fala do intérprete de conferências e os fatores que podem vir a atrapalhar a avaliação da qualidade e da compreensibilidade dessa fala. Considera-se o papel desses elementos prosódicos na comunicação falada e apresenta-se o resultado de trabalhos que investigam tais elementos na produção da fala em geral e especificamente na fala do intérprete de conferências. Argumenta-se sobre a necessidade de se investigar os aspectos prosódicos da fala do intérprete que, embora pouco explorada na literatura fonética, tem um papel muito importante na qualidade e compreensibilidade da fala do intérprete, dado que os parâmetros prosódicos desempenham um papel de relevância para o processamento da fala pelos ouvintes.

## **ABSTRACT**

This chapter aims at investigating the speech prosodic elements, pause, intonation and stress patterning in the context of conference interpreting and the intervening factors which may cause disturbances in the quality and comprehensibility of such a kind of speech. The role of the prosodic elements in the speech production and the results of the works which investigate these elements in the speech production in general and specifically in relation to the conference interpreting are discussed. The necessity of investigating the prosodic aspects of the interpreters' speech, although little explored in the phonetic literature, has a very important role in the quality and comprehensibility of the interpreters' speech, given that the prosodic parameters play a relevant role in the processing of speech by listeners.

***ON THE VARYING RECEPTION OF SPEAKERS EXPRESSIVITY  
ACROSS GENDER AND CULTURES, AND INFERENCE  
IN THEIR PERSONALITIES***

ALBERT RILLIARD  
DONNA ERICKSON  
JOÃO DE MORAES  
TAKAAKI SHOCHI

**RESUMO**

Neste trabalho dois experimentos de percepção são apresentados e comparados com as medidas acústicas de variação nas vozes de falantes. O primeiro experimento relata a percepção da qualidade do desempenho em comportamento expressivo; o segundo relata como os ouvintes inferem a personalidade dos falantes a partir desse comportamento expressivo. Os falantes mais expressivos, aqueles que produzem uma gama de variação maior de alterações acústicas, são julgados como mais extrovertidos e conscienciosos - e apresentam melhores desempenhos na expressão de irritação, uma forma de expressão que remete explicitamente ao sentimento do ouvinte. Falantes com níveis altos de ruído na voz são julgados com traço de personalidade de baixa agreeabilidade e apresentam melhores desempenhos em expressões envolvendo atitudes impositivas sobre o interlocutor. Finalmente, os falantes julgados como possuidores de alto grau de neuroticismo e baixos traços de imparcialidade também são tidos como os que melhor expressam fortes restrições sociais, tais como, no caso dos japoneses, as situações convencionais de sinceridade e o “pisando em ovos”. Indivíduos interpretam a personalidade do outro e o papel social notavelmente por meio do comportamento expressivo contextualizado e podem inferir de tais interpretações tendências de ação potenciais, prováveis instâncias de atos de fala.

**ABSTRACT**

Two perception experiments are presented, and compared with acoustic measures of the variation in speakers' voices. The first experiment relates perception of speaker's performance quality in expressive behaviour; the second one relates how listeners infer the speaker's personality from such expressive behaviour. The most expressive speakers, those who produce the widest range in acoustic changes, are perceived as more extraverted and conscientious – and produce the best performances for irritation, an expression that renders explicitly the speaker's feeling. Speakers with high noise levels in the voice are judged with low agreeable personality trait, and produce the best performances for expressions involving an imposition onto the interlocutor. Finally, speakers judged as having high neuroticism and low openness traits are also perceived as the best performers for expressions with strong social constraints, such as the Japanese conventionalized situations of sincerity and “walking-on-eggs”. Individuals interpret the other's personality and/or social role notably through their contextualized expressive behaviour, and may infer from such interpretations potential action tendencies, probable claims or speech acts.

# *SCRIPTS FOR THE ACOUSTIC ANALYSIS OF SPEECH DATA*

PLÍNIO A. BARBOSA  
ZULEICA A. CAMARGO  
SANDRA MADUREIRA

## **RESUMO**

Este trabalho apresenta cinco scripts do PRAAT, desenvolvidos para analisar os parâmetros acústicos voltados à pesquisa científica sobre o modelamento prosódico e à investigação do vínculo entre a produção e a percepção da prosódia na fala normal e patológica. O script Beat-Extractor faz a detecção automática e independente de língua de ataques silábicos; o SGdetector realiza a detecção automática e independente de língua de picos de duração do tamanho de sílabas e normalizados para fins de estudo da proeminência e da marcação de fronteiras; o SaliencyDetector realiza a detecção automática e independente de língua de picos de duração do tamanho de sílabas e normalizados para fins de estudo da proeminência e das fronteiras, enquanto o ProsodyDescriptor e o ExpressionEvaluator permitem a geração de 12 parâmetros acústicos relacionados à duração, F0, LTAS e ênfase espectral para fins de estudo do ritmo, da entoação e da qualidade de voz em fala expressiva e patológica. Todos os scripts estão disponibilizados livremente. Até o momento foram testados no português brasileiro, português europeu, alemão, francês, inglês e sueco.

## **ABSTRACT**

This paper presents five Praat scripts developed for analysing prosodic acoustic parameters as an aid for scientific research on prosody modelling and for investigating the link between prosody production and perception in normal and pathological speech. The BeatExtractor script does automatic, language-independent detection of vowel onsets; the SGdetector script does language-dependent, semi-automatic detection of syllable-sized normalised duration peaks for the study of prominence and boundary marking, the SaliencyDetector script does language-independent automatic detection of syllable-sized normalised duration peaks for the study of prominence and boundary, whereas the ProsodyDescriptor and the ExpressionEvaluator scripts allow to generate 12 prosodic parameters related to duration, F0, LTAS and spectral emphasis to the study of rhythm, intonation and voice quality in expressive and pathological speech. All scripts are freely available and the languages tested so far were Brazilian Portuguese, European Portuguese, German, French, English and Swedish.

*As Palavras Cantada, Falada e  
Declamada: Um Estudo Sobre a Relação  
Entre Canto, Fala e Declamação e seus  
Efeitos Impressivos*

JULIANA ANDREASSA DA LOMBA  
SANDRA MADUREIRA  
MARIO AUGUSTO DE SOUZA FONTES

# 1. Introdução

Este estudo tem como objetivo realizar um experimento de percepção para investigar, por meio de uma perspectiva contrastiva, o impacto impressivo causado nos ouvintes por trechos de fala, canto e declamação do espetáculo “Totatiando” estrelado por Zélia Duncan que interpreta composições de Luís Tatit.

Tatit, juntamente com o grupo Rumo, que começou suas apresentações em 1974, apresentava canções cujas melodias se aproximavam da melodia da fala, o que levou Tatit a definir a palavra cantada, ou o canto falado, como a entoação que faz parte da melodia, assim como a melodia faz parte da fala.

Na peça “Totatiando”, Duncan usa a palavra cantada para que determinadas partes possam ser diferenciadas de outras, despertando, assim, a curiosidade de investigar o impacto que a palavra cantada, falada e declamada podem causar.

Os três estilos orais são construídos de forma a dar vida aos personagens. Não há mudança de figurino, apenas alguns acessórios são trocados durante a peça e a transformação de personagens se dá única e exclusivamente pela voz. A maneira como a melodia é empregada durante as canções faz com que as palavras cantada, falada e declamada sejam priorizadas em momentos específicos da peça e, por vezes, [alternada](#).

Segundo Tatit, no enleio entre a melodia da música e os elementos linguísticos, o cancionista deve produzir uma oralidade que se aproxima ao máximo da fala para que soe o mais natural possível, levando em consideração que seu maior instrumento é a entoação, uma vez que, é a partir desse elemento prosódico que a fala pode se estender ao canto.

As canções, para Tatit, são consideradas uma forma de linguagem, assim como a fala, e acrescentamos a declamação, uma vez que apenas a maneira como a prosódia é utilizada as diferenciam. Assim, a canção, a fala e a declamação são formas de linguagem oral que conjugam aspectos verbais e não verbais. Por ser a linguagem oral a principal ferramenta que o homem possui para a comunicação e a expressão de suas ideias e emoções (Fontes e Madureira, 2015), interessa investigar o seu papel comunicativo.

Se a linguagem oral tem o poder de transmitir ao outro informações e emoções, consequentemente, ela comunica, portanto, o canto, a fala e declamação são formas de expressão social.

A metodologia deste estudo foi construída a partir de um teste de percepção que continha descritores e escala para fins de atribuição de julgamentos das impressões causadas pelos estímulos sonoros. Os ouvintes, os juízes da pesquisa, avaliaram trechos de palavra cantada, falada e declamada selecionados e extraídos da peça “Totatiando”. Os procedimentos do teste envolveram 9 recortes sem alteração e os mesmos 9 com manipulação realizada no programa de análise acústica PRAAT, criando objetos de *pitch* de tal modo a apresentar apenas o aspecto melódico, eliminando os segmentos fônicos.

## 2. Fundamentos

### 2.1 Os componentes da sonoridade

Voz, ruído e silêncio constituem a sonoridade das palavras cantada, falada e declamada. A teoria acústica da produção da fala, desenvolvida por Fant (1970), explica como se geram no aparelho fonador as fontes de voz e ruído. A fonte de voz é gerada pela vibração das pregas vocais e as fontes de ruído contínuo pela passagem da corrente de ar entre articuladores muito aproximados, tornando-a turbulenta. As fontes de ruído transiente, pela liberação abrupta da corrente de ar que estava interrompida pelo contato entre articuladores. As fontes de voz e ruído constituem os sons

da fala ao sofrerem a ressonância no trato vocal (que compreende desde a glote fechada até os lábios e as narinas). A produção da fala depende dos órgãos que formam o aparelho fonador e que compreendem a parte respiratória, que fornece a corrente de ar, a fonatória e a articulatória que a modificam.

O elemento chave para o estudo de acústica da fala é a noção de espectro, que especifica os valores de intensidade em relação aos componentes de frequência que caracterizam os sons. Os sons são analisados acusticamente quanto aos parâmetros de frequência, duração e intensidade.

Em nível de produção o vozeamento está relacionado à taxa de vibrações das pregas vocais. Em nível acústico o correlato dessa taxa é a frequência fundamental e em nível auditivo o correlato é o *pitch* que corresponde à sensação auditiva que permite classificar o som como grave ou agudo.

## **2.2. Os aspectos simbólicos da sonoridade**

Investigar as características das palavras cantada, falada e declamada implica em estudar melodias e seus efeitos impressionantes sobre o ouvinte, o que remete às relações entre som e sentido. Ohala (1983) utiliza o termo simbolismo sonoro para denominar as relações entre som e sentido. Desta forma, discutir as relações entre som e sentido implica em reconhecer a potencialidade do simbolismo sonoro, ou seja, do vínculo direto entre som e sentido, que é a base da expressividade sonora (Madureira, 2011).

De acordo com Madureira (2011), o simbolismo sonoro pode ser interpretado com base em três códigos biológicos: o código de frequência (Ohala, 1983; Ohala, 1984; Chuenwattanapranithi, 2008), o código de esforço e o código de produção (Gussenhoven, 2002, 2004).

O código de frequência remete às variações de frequência fundamental ( $f_0$ ) que correspondem em nível de produção ao número de vibrações das pregas vocais por segundo e em nível auditivo ao *pitch*, a sensação auditiva de som grave e agudo. Sons graves são associados ao que é grande, forte, pesado, enquanto sons agudos ao que é pequeno, frágil e leve. Estabelecem-se, assim, relações entre os planos acústico e de sentido.

O código de esforço relaciona-se com a articulação. Na produção da fala, quanto maior o esforço da articulação, maior a proeminência alcançada e maior a duração dos segmentos fônicos. E o código de produção está relacionado com a pressão de ar subglótica: no início do discurso a pressão de ar subglótica aumenta e no final ela abaixa.

Esses três códigos biológicos evidenciam os efeitos de sentido que podem ser alcançados pela voz e são de interesse para a investigação dos efeitos impressionantes das palavras cantada, falada e declamada sobre os ouvintes.

## **2.3. Os aspectos impressionantes e expressivos da sonoridade**

A partir da percepção das características da voz percebidas, os ouvintes atribuem características físicas, sociais e psicológicas ao falante. Desse fato, pode-se depreender a relevância comunicativa da voz.

A voz que fala, que canta e que declama é usada para expressar opiniões, ideias, estados afetivos, ou seja, aponta para uma intenção comunicativa. E na música, segundo Schafer (1991), a intencionalidade comunicativa é a mesma compreendida em uma conversa oral.

Na fala, a expressividade é caracterizada pelas combinações entre variantes segmentais (vogais e consoantes) e prosódicas (ritmo, entoação, qualidade de voz, taxa de elocução, pausas e padrões de acento) e causa impressões nos ouvintes independentemente de como ela é produzida, pois toda fala é expressiva, ou seja, veicula um sentido (Madureira, 2004).

Um importante trabalho sobre expressividade da fala pelo conjunto das medidas acústicas utilizadas e as correlações entre essas medidas e os resultados de teste de avaliação perceptiva de primitivos emocionais efetuado é o de Barbosa (2009) que realizou um estudo experimental para detectar mudanças na expressividade da fala em interações entre um radialista e entrevistados anônimos. Utilizou medidas acústicas extraídas automaticamente pelo *script* ExpressionEvaluator de sua autoria e aplicou em juízes um teste perceptivo para avaliar quatro dimensões de primitivos emotivos: ativação, valência, dominância e envolvimento. Para correlacionar os resultados do teste de percepção com as medidas acústicas utilizou dois métodos de análise estatística: a análise de componentes principais e a análise de regressão múltipla. Os resultados indicaram as dimensões de ativação, valência e ativação eram capazes de explicar 97% da variância. Denominou esse fator “prontidão” e considerou-o promissor para a identificação de fenômenos da fala expressiva.

A construção da expressividade da fala se dá por meio de relações diretas que se estabelecem entre som e sentido, ou seja, são atribuídos sentidos a partir da avaliação das características da sonoridade. A materialidade fônica impressiona os ouvintes e estes atribuem sentidos com base nas características acústicas que avaliam.

A construção da expressividade oral em um espetáculo musical, como é o caso da peça teatral musical “Totatiando”, visa despertar nos espectadores um tipo de emoção que é designada por Scherer (2005) como emoção estética. As emoções estéticas, tais como admiração, êxtase, fascinação e entusiasmo são produzidas pela apreciação de elementos da natureza, obras de arte ou uma apresentação artística. Scherer menciona, também, as emoções utilitárias, as quais correspondem às emoções básicas, tais como raiva, medo, alegria. São consideradas utilitárias no sentido de facilitar as adaptações a eventos que possuem consequências importantes para o bem-estar do indivíduo.

Emoções são diferenciadas de sentimentos e atitudes. Segundo Scherer (*opus cit*), as emoções são descargas psíquicas de duração relativamente curta, pois os eventos emotivos mudam rapidamente, na maioria das vezes por causa de novas informações ou novos estímulos. Essas descargas psíquicas ocorrem no organismo humano e são fruto de um estímulo que desperta uma reação. Isso tem impacto na comunicação e pode ter, também, consequências para a interação social.

### **3. Características das Palavras Cantada, Falada e Declamada**

Os elementos verbais e não verbais, são utilizados na construção das palavras falada e cantada e declamada. Os não verbais são, por exemplo, os padrões respiratórios, os risos, os gritos, que, segundo Beller (2009) tem a função de marcadores semânticos da expressividade.

A canção bem como a fala e a declamação são caracterizadas por padrões melódicos e rítmicos. Na fala e na declamação esses padrões são alterados livremente, mas na canção não há essa flexibilidade.

Na voz que fala, canta e declama periodicidades, aperiodicidades e silêncios alternam-se. Schafer (1991) aponta que não apenas as periodicidades são usadas para produzir música, mas também as aperiodicidades, ou seja, os ruídos que o referido autor considera como um som musical. Refere-se aos ruídos como a parte negativa do que existe no som musical. Nas palavras falada, cantada e declamada os ruídos compõem nos sons obstruintes, ou seja, as consoantes oclusivas, fricativas e africadas, e nas vozes produzidas com escape de ar.

Os silêncios também colaboram para constituir a trama sonora. E, em relação aos silêncios, Wisnik (1989, p.18) aponta que “O som é presença e ausência e está, por menos que isso apareça, permeado de silêncio.”.

### 3.1. A palavra cantada

Segundo Tatit (1996, p. 9), “Cantar é uma gestualidade oral, ao mesmo tempo contínua, articulada, tensa e natural”, exigindo um equilíbrio entre o musical e o linguístico.

Na canção, para construir a expressividade musical, o compositor pode se limitar a determinados elementos, o que Schafer chama de “paleta limitada”, criando, assim, certa atmosfera. Como exemplo disto, Schafer dá o exemplo de duas peças distintas: “A tarde de um fauno”, de Claude Debussy e a “Noite numa montanha deserta”, de Modeste Moussorgsky. A primeira apresenta elementos musicais suaves, graves e lentos, enquanto a segunda se apropria de valores fortes, agudos e rápidos. Nesse sentido são opostas, entretanto, as duas peças podem ser consideradas composições “líricas”, de acordo com Schafer, uma vez que, possuem como finalidade criar e sustentar uma atmosfera determinada. Se fossem composições “dramáticas” precisariam de valores opostos com mudanças inesperadas.

Outro elemento formal da música é o timbre, segundo Schafer (1991, p.76), “O timbre é a qualidade característica de um som que distingue um instrumento do outro, dada a mesma frequência e amplitude”, ou seja, o timbre é caracterizado pelos componentes de frequências que compõem a onda sonora emitida pelos instrumentos. Na música a mudança de timbre não ocorre rapidamente, porque o compositor precisa de um tempo maior para que a mudança do timbre do instrumento aconteça.

Schafer (1991, p.81) afirma que para a construção de uma melodia é preciso “movimentar o som em diferentes altitudes”. As melodias musicais possuem limitações de movimento. Diferentemente da melodia que é a combinação de sons sucessivos a harmonia se dá por uma ou mais notas tocadas juntamente.

Além da melodia, um elemento de relevância para compor as texturas sonoras musicais é o ritmo. Schafer discute dois tipos de ritmos musicais, os regulares e os irregulares. Ritmos regulares propõem divisões cronológicas do tempo real, ou seja, ritmos mecânicos que acompanham a sequência cronológica dos atos em uma melodia. Todavia, os ritmos irregulares comprimem ou expandem o tempo real, criando, assim, um tempo virtual.

### 3.2 A palavra falada

Na fala, os aspectos melódicos correspondem à entoação, elemento prosódico quem tem como correlato acústico principal a frequência fundamental e como correlato perceptivo o *pitch*. A modulação da frequência fundamental é utilizada na fala para expressar sentidos. Ela exerce diferentes funções: demarcativa (delimitando os constituintes sintáticos), informativa (informação sobre proeminência, foco e informação nova e velha) expressiva (veiculação de efeitos de sentido, atitudes e emoções), interativa (a sinalização de turnos) e comunicativa (veiculação de intenções e de aspectos sociais e regionais).

Os aspectos rítmicos na fala implicam em uma alternância da sucessão vocálica e em uma alternância acentual resultante do deslocamento mais acentuado da mandíbula no fluxo da fala (Barbosa, 2000). Ritmos implica em sequenciação e estruturação.

O correlato principal acústico do ritmo é a duração. A sensação perceptiva de isocronismo de sílabas e acentos levou à postulação da existência de ritmos acentuais e silábicos nas línguas. Barbosa (2006, p.193) “a tendência de uma língua para um ritmo acentual ou silábico pode ser explicada respectivamente pela maior ou menor força de acoplamento do oscilador acentual que está relacionado com a determinação das sílabas fonéticas, entendidas como unidades que se estendem de vogal a vogal e o ritmo acentual com os locais de acentos frasal”.

### 3.3 A palavra declamada

Para a realização de uma declamação é necessária uma interpretação prévia do texto poético pelos locutores, para que essa forma de expressão artística cumpra o seu papel de transmitir sentimentos e impressionar os ouvintes. A escolha dos elementos prosódicos e das variantes segmentais pelos locutores reflete a interpretação dada ao texto, pois revela o trabalho de adequação das formas sonoras aos sentidos produzidos a partir dessa interpretação e nesse trabalho o recurso a formas de simbolismo sonoro é produtivo.

Resultados do estudo realizado por Madureira (2008), levando em consideração a declamação do poema “Soneto da Fidelidade” de Vinicius de Moraes realizada por uma atriz e um ator, explora o uso do simbolismo sonoro na construção da expressividade na fala. Esse estudo considera o uso de elementos prosódicos (qualidade de voz e entoação) e de variantes segmentais na declamação do poema.

Um exemplo da utilização de estratégias da fala baseadas no simbolismo sonoro foi usado pela atriz na produção da palavra “chama”. A atriz pronunciou essa palavra, com ajuste de qualidade de voz com escape de ar (Camargo e Madureira, 2008), forma sonora que se coaduna com o sentido do movimento e da instabilidade da chama da vela.

Um outro exemplo de uso de simbolismo sonoro em Madureira (2008), foi o alongamento vocálico produzido pelo ator na sílaba inicial da palavra “infinito” e a subida de  $f_0$  a culminar na tônica dessa palavra, o que contribuiu para enfatizar o significado da palavra.

Resultados do estudo realizado por Madureira e Camargo (2010), levando em consideração a declamação do poema “A Valsa” de Casimiro de Abreu feita por um ator, revelou que o ator utilizou o simbolismo sonoro para sinalizar a dinâmica da dança e a dinâmica do estado afetivo do observador da dança. Os elementos prosódicos na declamação desse poema foram utilizados de forma a revelarem uma relação entre os compassos e os movimentos da dança. O fim da dança no poema é sinalizado na declamação por mudanças na taxa de elocução, no ritmo e na entoação, o que mostra o uso do simbolismo sonoro para expressar efeitos de sentido.

## 4. Metodologia

Para a realização desta pesquisa, que tem como base a peça “Totatiando” de Zélia Duncan, primeiramente, foi realizada a seleção de trechos das palavras cantada, falada e declamada.

Após a seleção dos trechos da peça, um questionário de diferencial semântico foi montado para uma análise de percepção. O questionário de diferencial semântico consiste em um conjunto de descritores, que variam em número (quantos descritores se queiram avaliar) e grau de intensidade de ocorrência (em geral de 5 a 7 graus) a ser avaliado por juízes em teste de avaliação perceptiva. Foi criado por Osgood, Suci e Tannenbaum (1957) que propôs a utilização de pares dicotômicos de descritores (bom/mau, por exemplo) para avaliação de estímulos.

Para o estudo da peça “Totatiando” foram utilizados 4 descritores, variando em 5 graus. O questionário de diferencial semântico foi aplicado a um grupo de 30 juízes que julgaram 18 trechos extraídos da referida peça, sendo 9 para cada tipo de estilo (cantado, falado e declamado), os mesmos 9, porém, [manipulados para objeto pitch](#) por meio do programa PRAAT, ou seja, os segmentos fônicos são eliminados e só se escuta a melodia.

Nesta pesquisa, os descritores selecionados levaram em consideração quatro aspectos: a melodia, o envolvimento, a vibração e a comoção, cabendo aos juízes, em sua avaliação auditiva dos enunciados, considerar o quão esses aspectos os afetavam.

Os juízes ouviram os estímulos e, a partir da impressão que a sonoridade desses enunciados lhes causava, avaliaram os descritores em uma escala de graus de 1 a 5, sendo o grau 1

correspondente a pouco e o 5 a muito. Os estímulos foram apresentados em duas sessões. Na primeira foram apresentados os estímulos manipulados para objeto *pitch* e na segunda os sem manipulação. A apresentação dos estímulos foi feita de maneira aleatória e as respostas dos juízes foram anotadas por eles em um formulário fornecido pelo pesquisador.

Para analisar os resultados, foi utilizado o teste de Spearman. O coeficiente de correlação de postos de Spearman é usado na estatística para medir a correlação não paramétrica, ou seja, avaliar uma função monótona arbitrária, que pode ser a relação entre duas variáveis. Foi calculada a média que cada descritor obteve em cada trecho para depois ser possível calcular o  $\rho$  (rho). Para o cálculo do  $\rho$ , foi realizada a correlação entre cada par de descritores com o intuito de descobrir qual a possível relação existente entre cada deles.

Também foi realizada uma análise de correlação, usando o método de Análise de Fatores Múltiplos para verificar se haveria correlação entre as variáveis.

#### **4.1 Corpus**

Para a constituição do *corpus* foram feitos recortes da peça “Totatiando”, contemplando-se os três estilos de palavra: fala, canção e declamação. Foram selecionados três trechos do texto “Haicai” para análise da palavra falada, três do “Sem destino” para análise da palavra cantada e três do poema “Quando morrer quero ficar”, de Mário de Andrade para a análise da palavra declamada. A seguir, transcrevemos os trechos selecionados das palavras cantada e falada, os quais são de autoria de Luís Tatit e os da palavra declamada de autoria de Mário de Andrade.

##### **I - Trechos com a palavra cantada**

1. Era pra eu já ter encontrado um amor  
Era pra eu já ter esquecido o anterior
2. Ser assim tão sem destino  
Me preocupa muito, me deixa infeliz
3. Quero ouvir uma vidente  
Que me conte tudo, só esconda a morte

##### **II - Trechos com a palavra falada**

1. Imagine se ela vai ouvir poesia  
Sofia, escuta essa, vai  
É pequena, é um haicai
2. Fico aqui pensando  
Onde é que estou errando  
Será que é, ou é defeito de Sofia
3. Imagine agora a chance que me resta  
Pra pegá-la acordada  
Só encontrando numa festa

##### **III - Trechos com a palavra declamada**

1. Meus pés enterrem na rua Aurora  
No Paissandu deixem meu sexo
2. A língua no alto do Ipiranga  
Para cantar a liberdade  
Saudade
3. As tripas atirem pro Diabo

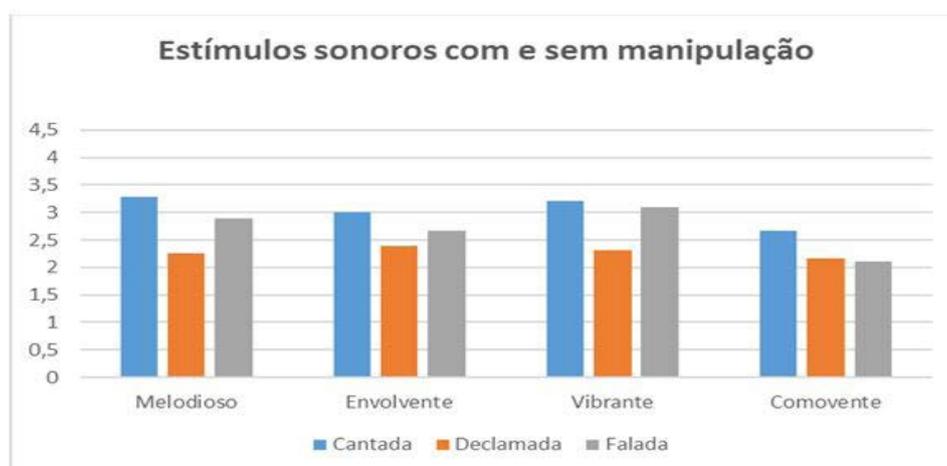
## 5. Resultados

Os quatro descritores avaliados pelo teste de percepção nesta pesquisa, melodioso, envolvente, vibrante e comovente, foram avaliados em relação aos estímulos de voz cantada, falada e declamada.

Os resultados dessa avaliação serão apresentados em três subseções. Na primeira é contemplada a avaliação dos trechos sem alteração e os modificados para objeto *pitch* conjuntamente, na segunda somente os trechos sem alteração e na terceira os com modificação para objeto *pitch*.

### 5.1 O resultado da avaliação do conjunto de estímulos

O gráfico da Figura 1, que apresentamos em seguida, mostra os resultados da avaliação dos quatro descritores do teste de percepção. A pontuação indicada no gráfico refere-se à média obtida a partir dos valores atribuídos pelos juízes para cada trecho dos três estilos orais, as palavras cantada, falada e declamada. A média calculada mescla os dois tipos de trechos apresentados no teste, os recortes de gravações do espetáculo e os mesmos recortes, manipulados para objeto *pitch*.

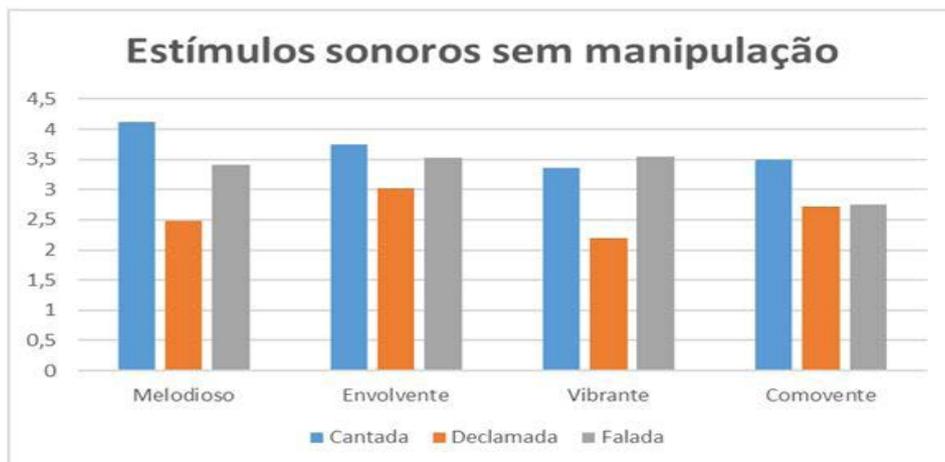


**Figura 1** – Pontuação (valores médios) referente à avaliação ao conjunto dos estímulos (sem e com manipulação para objeto das palavras declamada, cantada e falada em relação aos descritores “melodioso, envolvente, vibrante e comovente”.

Pelo exame do gráfico da Figura 1 é possível constatar que a palavra cantada apresentou valores maiores para todos os descritores. O descritor “melodioso” obteve a maior pontuação na avaliação da palavra falada, o descritor “vibrante” na avaliação da palavra falada e o descritor “envolvente” na avaliação da palavra declamada. As palavras cantada, falada e declamada foram avaliadas com menor pontuação em relação ao descritor “comovente” do que em relação aos demais descritores.

## 5.2 O resultado da avaliação dos estímulos sem modificação

No gráfico da Figura 2, a seguir, são apresentados os resultados da avaliação dos juízes referentes aos estímulos sem modificação.



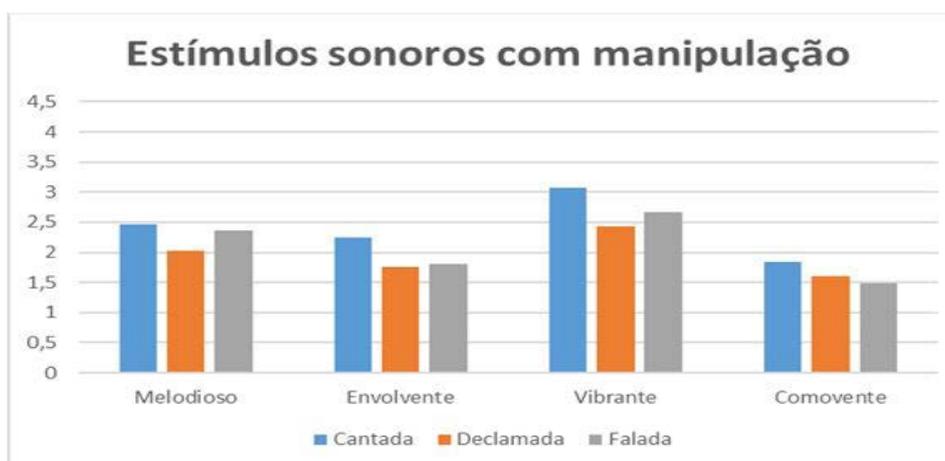
**Figura 2**– Pontuação (valores médios) referente à avaliação dos estímulos sem manipulação referentes às palavras declamada, cantada e falada em relação aos descritores “melodioso, envolvente, vibrante e comovente”.

No gráfico da Figura 2, fica evidenciado o fato de que as palavras cantada, foi melhor avaliada em relação ao descritor “melodioso”, a palavra falada em relação aos descritores “vibrante” e “envolvente” e a palavra declamada em relação ao descritor “envolvente”. Se compararmos os resultados apresentados na Figura 2 com os exibidos na Figura 1, verificamos que os descritores, com exceção do “vibrante” que não apresentou alteração, receberam maior pontuação quando considerados os estímulos sem manipulação.

Na palavra cantada, em nível de significância de 0,05, foram significativas as seguintes diferenças: entre o descritor “melodioso” e os descritores “vibrante” e “comovente” ( $p=0,0004$ ); na palavra declamada entre o “melodioso” e o “envolvente”, entre o “envolvente” e o “vibrante” e entre o “vibrante” e o “comovente” ( $p=0,0005$ ); na palavra falada, o comovente dos demais descritores ( $p=0,0001$ ).

## 5.3 O resultado da avaliação dos estímulos com modificação

No gráfico da Figura 3 são apresentados os resultados obtidos em relação aos estímulos manipulados para objeto *pitch*, levando em consideração, novamente, o quão melodioso, envolvente, vibrante e comovente os estímulos foram avaliados pelos juízes.



**Figura 3** – Pontuação (valores médios) referente à avaliação dos estímulos com manipulação para objeto *pitch* referentes às palavras declamada, cantada e falada em relação aos descritores “melodioso, envolvente, vibrante e comovente”.

Os resultados da avaliação dos estímulos com manipulação para objeto *pitch* evidenciam que diferentemente da dos estímulos sem manipulação o descritor melhor avaliado foi o vibrante em relação aos estilos de palavra cantada, falada e declamada. Observa-se também que a pontuação em relação aos resultados da avaliação dos estímulos com manipulação para objeto *pitch* apresentou valores menores dos que os sem manipulação.

Na palavra cantada, em nível de significância de 0,05, foram significativas ( $p=0.0000$ ) as seguintes diferenças: entre o descritor “vibrante” e os demais descritores e entre o “melodioso” e o “comovente”; na palavra declamada o “vibrante” do “envolvente” e do “comovente”; na palavra falada, o “melodioso” do “envolvente” e do “comovente”, o “envolvente” do “vibrante” e o “vibrante” do “comovente”.

#### 5.4 Correlações estatísticas entre pares de descritores

Para a realização da análise da força das correlações entre pares de descritores obtidas pelo coeficiente  $\rho$  (rho), os estilos de palavras falada, cantada e declamada foram considerados separadamente para que fosse possível posteriormente compará-los em termos das diferenças e semelhanças obtidas. Valores de 0 a 0,19 indicam uma correlação muito baixa, de 0,20 a 0,39 uma correlação fraca, de 0,4 a 0,59 uma correlação moderada e acima desses valores uma correlação forte.

Os coeficientes  $\rho$  obtidos para os trechos de palavra cantada são apresentados na tabela 1, os da palavra falada na Tabela 2 e os da palavra declamada na Tabela 3.

**Tabela 1** - Medidas relativas ao coeficiente  $\rho$  (Canto)

	Melodioso	Envolvente	Vibrante	Comovente
Melodioso		0,20	0,21	0,16
Envolvente	0,20		0,03	0,01
Vibrante	0,21	0,03		0,03
Comovente	0,16	0,01	0,03	

**Tabela 2** - Medidas relativas ao coeficiente  $\rho$  (Fala)

	Melodioso	Envolvente	Vibrante	Comovente
Melodioso		0,10	0,41	0,07
Envolvente	0,10		0,25	0,02
Vibrante	0,41	0,25		0,30
Comovente	0,07	0,02	0,30	

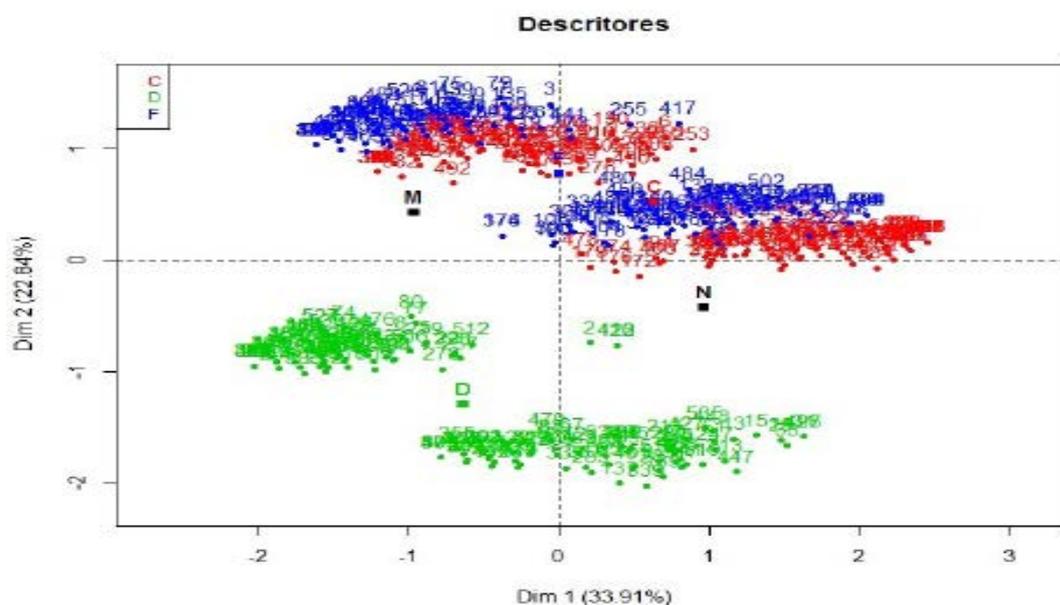
**Tabela 3** - Medidas relativas ao coeficiente  $\rho$  (Declamação)

	Melodioso	Envolvente	Vibrante	Comovente
Melodioso		0,21	0,32	0,22
Envolvente	0,21		0,43	0,07
Vibrante	0,32	0,43		0,32
Comovente	0,22	0,07	0,32	

Na palavra cantada as correlações entre os descritores vibrante e envolvente foram fracas e as demais muito baixas, mas na palavra falada ocorreu uma correlação moderada entre vibrante e melodioso e na palavra declamada entre vibrante e envolvente. As demais correlações foram fracas: entre vibrante e comovente na palavra falada e entre vibrante e melodioso e entre vibrante e comovente na palavra declamada. As correlações entre comovente e envolvente foram muito baixas nos três tipos de palavra.

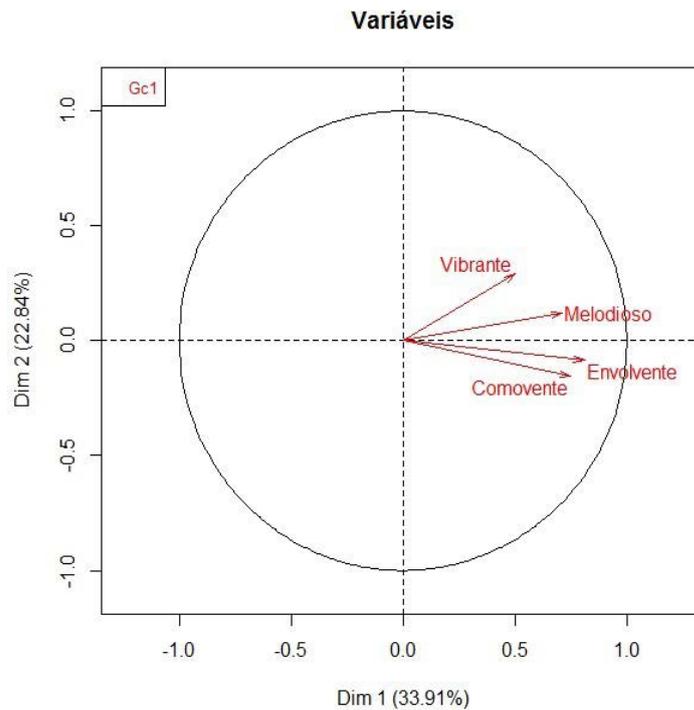
## 5.5 Análise multidimensional

A Análise de Múltiplos Fatores revelou, nas dimensões 1 e 2 do espaço vetorial referentes às medidas relativas aos trechos avaliados pelos juízes, que os três estilos orais, palavra cantada, falada e declamada diferenciam-se, formando *clusters* separados. O mesmo se deu com os dados manipulados e não manipulados. Formaram-se 6 *clusters*: palavra falada manipulada, palavra falada não manipulada, palavra cantada manipulada, palavra cantada não manipulada, palavra declamada manipulada, palavra declamada não manipulada. Na Figura 4, a seguir, apresentamos as nuvens de pontos correspondentes aos trechos avaliados pelos juízes. Em vermelho, a palavra cantada, em azul, a palavra falada e em verde, a palavra declamada. Os trechos modificados (M) aparecem concentrados nos quadrantes à esquerda e os não manipulados (N) nos quadrantes à direita. O posicionamento das letras “C”, “D”, “F”, “M” e “N” na Figura 4 indica o centróide de todos os pontos relativos aos trechos julgados pelos juízes.

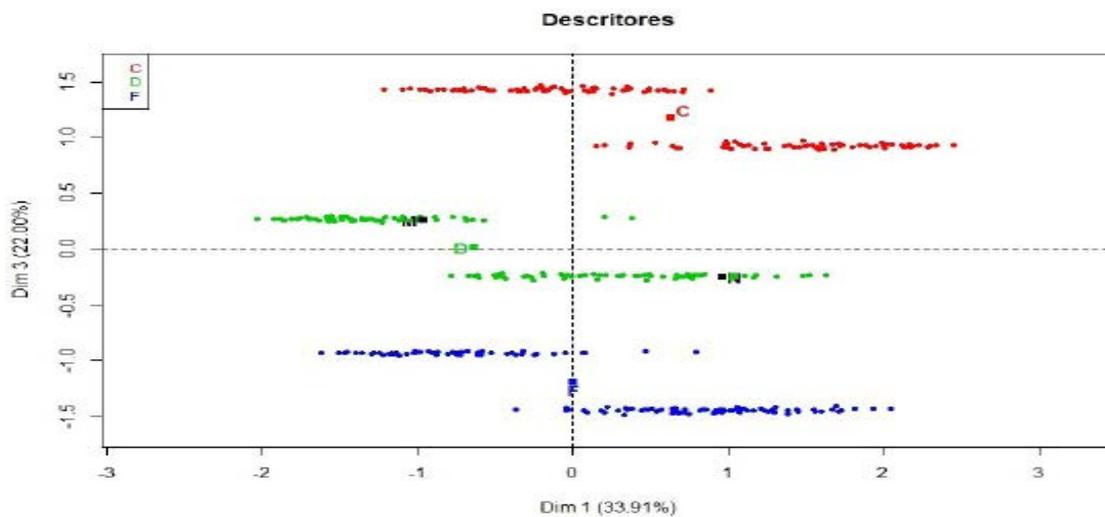


**Figura 4** – Clusters de nuvens de pontos referentes aos trechos avaliados pelos juízes. Em vermelho, a palavra cantada, em azul, a palavra falada e em verde, a palavra declamada. As letras “C”, “D”, “F”, “M” e “N” indicam a posição dos seus respectivos centróides.

A seguir, apresentamos, na Figura 5, os descritores que foram avaliados pelos juízes e que determinaram a distribuição dos pontos referentes aos trechos do corpus segundo a análise de correlação cujos resultados foram mostrados na Figura 4. No eixo horizontal da Figura 5, os descritores evoluem na mesma direção, mas no eixo vertical, vibrante e melodioso opõem-se à envolvente e comovente, ou seja, quando os valores de vibrante e melodioso crescem os de envolvente e comovente decrescem e vice-versa.



**Figura 5** – Gráfico da distribuição das variáveis referentes aos descritores nas dimensões 1 e 2 do espaço vetorial



**Figura 6** – Clusters de nuvens de pontos referentes aos trechos avaliados pelos juizes. Em vermelho, a palavra cantada, em azul, a palavra falada e em verde, a palavra declamada. As letras “M” (manipulado), “N” (não-manipulado), “C” (cantado), “D” (declamado) e “F” (falado) indicam a posição dos seus respectivos centróides.

Com base na distribuição dos pontos nos quadrantes referentes ao espaço vetorial formado pelas dimensões 1 e 3, configura-se um contínuo entre palavra falada (nuvem de pontos em azul), declamada (nuvem de pontos em verde) e cantada (nuvem de pontos em vermelho). A palavra falada aproxima-se mais da declamada e esta da cantada.

## 6. Conclusão

As avaliações das palavras cantada, falada e declamada manipuladas para objeto *pitch* foram melhor avaliadas quanto ao descritor “vibrante” do que em relação aos demais descritores, enquanto que as não manipuladas diferenciaram-se em relação aos descritores melhor avaliados: para a palavra falada, vibrante, envolvente e melodioso; para a palavra cantada melodioso e envolvente; para a palavra declamada, envolvente e comovente.

Ao passar do não manipulado para o manipulado, nas palavras cantada e falada o componente que sofreu menor alteração foi o vibrante e o que sofreu maior alteração foi no caso da palavra falada o envolvente e no caso da palavra cantada o melodioso e o envolvente, para a palavra declamada o envolvente e o comovente. A vinculação com o sentido tornou as palavras falada, cantada e declamada mais envolventes, a cantada mais melodiosa e a declamada mais comovente.

A manipulação prejudicou o envolvimento dos ouvintes, mas não a percepção do descritor vibrante. Esse foi o efeito de sentido atribuído pelos juízes para a sonoridade que ouviram, independentemente do estilo de palavra. O fato das palavras falada aproximarem-se mais da declamada do que entre si pode ser interpretado como influenciado pelos aspectos melódicos, ou seja pela musicalidade. Configuram-se nesses exemplos instâncias de simbolismo sonoro.

## Referências Bibliográficas

- BARBOSA, P. A. Detecting changes in speech expressiveness in participants of a radio program. In: **Proceedings of Interspeech 2009**, 6–10 September, Brighton: 2155–2158, 2009
- BARBOSA, P. A. A dynamical model for generating prosodic structure. In: **Speech Prosody 2006, Dresden. Proceedings of the Speech Prosody Conference**, 2006. p. 366-369.
- BARBOSA, P. A. Syllable-timing in Brazilian Portuguese: uma crítica a Roy Major. **D. E. L. T. A.** no. 2, 16, 369-402. São Paulo, 2000.
- CAMARGO, Z.; MADUREIRA, S. 2008. Voice quality analysis from a phonetic perspective: Voice Profile Analysis Scheme Profile for Brazilian Portuguese (BP-VPAS). In: **Proceedings of the 4th Conference on Speech Prosody v.1**, 57–60. Campinas, Brasil.
- BELLER, G. Transformation of expressivity in speech. In: **Linguistic Insights**, vol. 97, 2009. p. 259-284.
- CHUENWATTANAPRANITHI, S. et al. Encoding emotions in speech with the size code: a perceptual investigation. **Phonetica** 65, 2008. p. 210-230.
- FANT, G. **Acoustic theory of speech production**. Paris, Mouton, 1970.
- FONTES, M. A. S.; MADUREIRA, S. Gestural prosody and the expression of emotions: a perceptual and acoustic experiment. In: **Proceedings of the 18th International Congress of Phonetic Sciences**, Glasgow, Scotland, 2015. No prelo.
- GUSSENHOVEN, C. **The phonology of tone and intonation**. Cambridge University, 2004.
- GUSSENHOVEN, C. Intonation and interpretation: phonetics and phonology. **Speech Prosody 2002: Proceedings of the First International Conference on Speech Prosody. Aix-en-Provence, ProSig and Université de Provence Laboratoire Parole et Language**, 2002. p. 47-57.
- MADUREIRA, Sandra. **The Investigation of speech expressivity**. In: MELLO, H.; PANUNZI, A.; RASO, T. (Eds.). (Org.). *Illocution, modality, attitude, information patterning and speech annotation*. Firenze: Firenze University Press, v. 01, 2011. p. 101-118.
- MADUREIRA, Sandra; CAMARGO, Z. A. **Exploring sound symbolism in the investigation of speech expressivity**. In: ISCA, 2010, Athens. *Exploring sound symbolism in the investigation of speech expressivity*. Athens: Edited by Antonis Botinis, v. 01, 2010. p. 105-108.
- MADUREIRA, S. Reciting a sonnet: production strategies and perceptual effects. In: **Speech Pro-**

**sody 2008, Campinas. Proceedings of the Speech Prosody Conference, 2008.** São Paulo: Editora RG, v. 01, 2008. p. 697-700.

OHALA, J. J. An ethological perspective on common cross-language utilization of f0 of voice. **Phonetica** **41** (1), 1984. p. 1-16.

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

OSGOOD, C.E.; SUCI, G.; TANNENBAUM, P. **The measurement of meaning.** Urbana, IL: University of Illinois Press, 1957.

SCHERER, K.R. et al. Comparing the expression of emotion in the speaking and the singing voice. **Computer Speech and Language**, v. 29, 2015. P. 218–235.

SCHERER, K.R. What are emotions? And how can they be measured? **Social Science Information**, v. 44, 2005. p. 695-729.

SCHAFER, R. Murray. **O ouvido pensante.** Tradução de Marisa Trench de O. Fonterrada, Magda R. Gomes da Silva, Maria Lúcia Pascoal. São Paulo: Fundação Editora da UNESP, 1991.

TATIT, Luiz. **O cancionista: composição de canções no Brasil.** São Paulo: Editora da Universidade de São Paulo, 1996.

WISNIK, José Miguel. **O som e o sentido: Uma outra história das músicas.** São Paulo: Companhia das Letras, 1989.

*Expressão vocal de emoções: metáfora  
sonora, fala e canto*

GLÁUCIA LAÍS SALOMÃO

## 1. Introdução

A comunicação de emoções é crucial para as relações sociais e para a sobrevivência das espécies (Ekman, 1992), e cumpre papel fundamental na manutenção da ordem social entre as pessoas (Buck, 1984).

O cenário do conhecimento sobre a comunicação de emoções modificou-se consideravelmente nas últimas décadas. Estudos tradicionalmente mais especulativos ganharam o suporte de dados provenientes de pesquisas experimentais; pesquisas focadas principalmente na compreensão da expressão facial de emoções passaram a abarcar modalidades expressivas até então pouco estudadas como, por exemplo, a voz, o corpo, a música, o cheiro; investigações sobre a expressividade musical ganharam maior projeção neste cenário, permitindo melhores comparações entre a expressão da emoção pela fala e pela música.

Mesmo com focos e interesses distintos, linguistas, psicólogos, biólogos, musicólogos, antropólogos, fonoaudiólogos, engenheiros, neurocientistas, cientistas da computação interessados em compreender os processos envolvidos na comunicação de emoções vêm sendo motivados por questões de pesquisa muito semelhantes, quais sejam: como as emoções são expressas? Como elas são percebidas?

Neste capítulo, tratamos de duas modalidades expressivas que apresentam em comum o fato de fazerem uso da matéria sonora como meio fundamental de comunicação. Tais modalidades são a fala e a música, em particular, o canto. Na realidade, a semelhança entre o modo como as emoções são expressas pela música e pelos componentes musicais (não verbais) da fala parece se estender para além do fato de fazerem ambas uso do sinal acústico como meio de comunicação. É provável que as duas modalidades sejam semelhantes já ao moldarem a matéria sonora que veicula a emoção.

Ao longo do texto, argumentaremos a favor da ideia de que a expressão vocal de emoções espelha aspectos fisiológicos associados à própria emoção expressa; de que fala e canto compartilham padrões de expressão muito semelhantes; e de que o canto é expressivo por conter lastros de padrões expressivos da fala.

## 2. Metáfora Sonora, Fala e Canto

A ideia de que a voz seja um espelho das emoções, na fala ou no canto, é bastante antiga. Em um artigo clássico, “A origem e a função da música”, Spencer (1857) já havia levantado a hipótese de que diferentes emoções estivessem associadas a manifestações fisiológicas também diferentes, e de que estas, por sua vez, provocassem efeitos distintos na produção vocal, falada ou cantada. Deste modo, a raiva, por exemplo, estaria associada a um aumento da tensão da musculatura laríngea e da pressão sob a glote, deixando marcas de tensão nos mecanismos da produção vocal e nas características acústicas da voz.

Em 1983, Iván Fónagy publicou um “ensaio psico-fonético”, *La Vive Voix*, escrito em linguagem marcadamente simbólica e bastante livre do formalismo dos textos científicos, no qual o autor sugere o termo “metáfora sonora” para fazer alusão à relação de semelhança entre os ajustes motores (fonatórios e articulatórios) subjacentes ao som vocal produzido na expressão de diferentes emoções e a própria emoção expressa. Seria como se uma zona de intersecção contendo traços comuns às dimensões sonora, motora e emotiva fosse estabelecida a cada gesto expressivo vocal. Assim, ao fazer considerações sobre a “mímica vocal”, o autor aponta, por exemplo, para uma relação de semelhança entre a tensão psíquica, a tensão fisiológica e a tensão expressiva. Ou ainda, ao tratar das “mímicas faríngea e bucal”, o autor compara o movimento da língua na “fala ameaçadora” ao “disparo de uma flecha” de uma posição articulatória para outra, referindo-se ao mesmo

como uma verdadeira materialização da própria ameaça em si.

A quantidade de estudos experimentais sobre a comunicação de emoções pela voz e pela música aumentou muito desde então. Há pouco mais de uma década, Juslin & Laukka (2003) publicaram uma extensa revisão de literatura visando averiguar possíveis paralelos entre a comunicação de emoções por meio da expressão vocal e da performance musical. O objetivo principal desta revisão foi o de investigar se a comunicação de certas emoções era igualmente efetiva nas duas modalidades expressivas e se existiam semelhanças entre os padrões acústicos utilizados na fala e na música na expressão de uma mesma emoção. O canto foi tratado como um instrumento musical e os resultados sobre a expressividade da voz cantada foram analisados conjuntamente com resultados sobre a expressividade em outros instrumentos musicais (ex: a flauta, o violino, o piano, etc). A meta-análise dos dados de literatura mostrou que, de fato, diferentes emoções puderam ser comunicadas de maneira igualmente efetiva por meio de ambas as modalidades, e sugeriu que a comunicação de emoções pela voz e pela música deva ocorrer por meio de um código acústico bastante similar. Os autores concluíram que os padrões acústicos encontrados em ambas as modalidades podiam ser consistentemente interpretados à luz do modelo teórico proposto por Klaus Scherer (1986), de base neurofisiológica.

O modelo de Scherer, em acordo com os preceitos da psicologia evolutiva, entende que as emoções constituam processos de adaptação ao ambiente compostos por componentes de naturezas distintas, i.e., componentes cognitivo, neurofisiológico, motivacional, relativo à expressão motora, relativo à sensação subjetiva. O modelo descreve as emoções como episódios relativamente breves, compostos por respostas sincronizadas e inter-relacionadas provenientes de vários subsistemas, desencadeadas a partir do julgamento pelo indivíduo do grau de relevância de eventos externos ou internos. Sendo assim, emoção não se confundiria com estado de humor, “postura” na interação, atitude, ou personalidade.

Abordagens desta natureza, que assumem que a emoção reflete, em última instância, o ambiente na qual se desenvolve, em geral pressupõem a existência de categorias de emoções inatas e reconhecidas universalmente, a partir das quais todas as outras seriam derivadas (Ekman, 1992; Izard, 1992; Johnson-Laird & Oatley, 1992). Tais abordagens encontraram suporte em estudos no campo da neurofisiologia, que mostraram que certas emoções de fato estavam associadas a padrões fisiológicos distintos (Fridlund, Schwartz, & Fowler, 1984; Bloch, Orthous, & Santibanez, 1987; Levenson, 1994) e ao aumento de atividade em áreas diferentes do cérebro (Damasio, Grabowski, Bechara, Damasio, Ponto, Parvizi, & Hichwa, 2000; Phan, Wager, Taylor, & Liberzon, 2002). A felicidade, a raiva, a tristeza, o medo, o amor (ou a ternura) constituiriam exemplos deste tipo de emoção, chamadas de “emoções básicas”, as quais, concatenadas e entrelaçadas, comporiam diferentes “estados emocionais” (Scherer, 1986).

Assumir que existam emoções básicas não implica em assumir que as mesmas definam um número limitado de padrões neuromotores envolvidos na expressão de emoções (Scherer, 2001). Pelo contrário, diferentes emoções, ou membros de uma mesma família de emoção, poderiam se combinar na definição de variados estados emotivos, de menor ou maior complexidade, e associar-se a mecanismos neurofisiológicos subjacentes também variados. Assim, o modelo de Scherer não apenas realiza predições a respeito do impacto sobre a voz de diferentes ajustes respiratórios, fonatórios e articulatórios associados a diferentes emoções (para uma visão detalhada, ver Scherer, 1986; 2001), mas também se mantém aberto para abarcar novas hipóteses sobre possíveis constelações de ajustes vocais subjacentes à expressão vocal, a serem devidamente testadas por meio de pesquisas experimentais (Scherer, 2001).

Pois bem, a expressão vocal, neste modelo, é entendida como um dos componentes da expressão motora da emoção (o outro é a expressão facial), uma espécie de espelho dos mecanismos neurofisiológicos associados aos diferentes estados emocionais e, neste sentido, um compo-

nente da própria emoção.

Sendo assim, ainda que se possa argumentar que as ideias de Iván Fonagy tenham sido apresentadas sem um rigor científico maior – de fato, elas foram publicadas em um ensaio – a essência da concepção de expressividade vocal como um reflexo sonoro da emoção condiz com o modelo teórico de base neurofisiológica proposto por Klaus Scherer, e parece ter conseguido antecipar aspectos importantes apontados em estudos experimentais que emergiram nas últimas décadas, sejam eles sobre a expressividade na fala ou sobre a expressividade no canto. Principalmente, o termo “metáfora sonora” parece sumarizar de maneira bastante elegante elementos fundamentais da expressividade humana.

### **3. A Face Sonora da Metáfora**

Mas, como, então, descrever as características da faceta sonora da “metáfora de Iván Fonagy”?

A matéria sonora vocal ou musical que reflete a emoção expressa pode ser descrita em termos de configurações acústicas em um espaço que inclui as dimensões de frequência, intensidade, duração e também relações entre frequência e intensidade na distribuição de energia no espectro sonoro. Configurações acústicas associadas a uma ou outra emoção são, em geral, consideradas como representativas das respectivas emoções se puderem ser decodificadas consistentemente e sistematicamente por ouvintes, e em concordância com a intenção expressiva do falante ou musicista. Além disso, em se tratando da comunicação de emoções básicas, ouvintes de diferentes culturas supostamente também poderão, em grande medida, compreender a emoção veiculada (Elfenbein & Ambady, 2002).

#### **3.1. Semelhanças entre a fala e o canto**

Na análise de possíveis semelhanças entre características sonoras de modalidades expressivas diferentes, como a fala e o canto, é importante que se leve em conta que algumas configurações acústicas podem se apresentar bastante próprias a um ou outro código expressivo. O vibrato, por exemplo, é recurso muito frequentemente utilizado no canto, porém não na fala. Outras variáveis, como: progressão harmônica, escalas e modos, que compõem o quadro sonoro que veiculará diferentes emoções pela música, não encontram correspondência na expressão vocal. Por outro lado, existem diversas configurações acústicas que se situam numa zona expressiva comum a ambas as modalidades e parece razoável supor que possíveis semelhanças expressivas possam ser encontradas justamente nesta zona sonora. Tais configurações, em geral, envolvem parâmetros que podem ser modificados com maior ou menor liberdade pelo falante ou musicista, independentemente do material verbal ou musical envolvido. Este é o caso, por exemplo, de parâmetros como: velocidade, intensidade, qualidade vocal (ou timbre sonoro, no caso dos instrumentos musicais).

De fato, as semelhanças entre a expressividade pela voz e pela música encontradas na revisão de literatura de Juslin & Lauka (2003) estavam associadas aos seguintes parâmetros:

- Taxa de elocução na fala e tempo na execução musical;
- Nível sonoro, em ambas as modalidades;
- Distribuição da energia no espectro acústico, em ambas as modalidades;
- Ataque vocal na fala e ataque tonal na execução musical;
- Frequência fundamental ( $f_0$ ), em ambas as modalidades;
- Contorno de  $f_0$ , em ambas as modalidades;
- Regularidade microestrutural, em ambas as modalidades.

É natural imaginar que, enquanto os quatro primeiros parâmetros possam realmente ser controlados com maior ou menor liberdade pelo falante ou musicista, o mesmo não se aplicaria a priori aos parâmetros que, numa performance musical, envolvam  $f_0$ . Isto se deve ao fato de que  $f_0$  é definido de antemão pelo compositor ao compor sua obra. No entanto, é comum que músicos/intérpretes de vários instrumentos e cantores, façam, durante a execução de uma obra, pequenos desvios de  $f_0$  em relação às notas da partitura ou que atinjam estas notas de maneiras diferentes, ascendente ou descendente, por exemplo. Foram, assim, estes os tipos de recursos (de micro-variação de  $f_0$ ) que os autores pesquisaram na literatura. Além destes, também a regularidade de variações de  $f_0$ , de intensidade e/ou de duração num nível microestrutural pôde ser comparada em ambas as modalidades expressivas.

Os resultados encontrados mostraram que as configurações acústicas da expressão vocal e da execução musical apresentaram características bastante semelhantes ao veicularem uma mesma emoção, e que tais configurações variaram sistematicamente de acordo com a emoção expressa, de tal modo que:

- Na raiva, a taxa de elocução da fala foi alta e o tempo da execução musical acelerado; o nível sonoro tendeu a ser alto e a apresentar grande variabilidade; houve grande quantidade de energia na região das frequências altas do espectro; houve grande variabilidade de  $f_0$  e seus contornos costumaram ser ascendentes; os ataques vocais/tonais tenderam a ser rápidos; houve considerável irregularidade micro estrutural;

- No medo, a taxa de elocução da fala foi também alta e o tempo da execução musical acelerado; o nível sonoro foi baixo (com exceção de no medo pânico, quando o nível sonoro apresentou-se bem mais elevado) e apresentou também grande variabilidade; a quantidade de energia na região das frequências altas do espectro foi pequena/moderada; houve pequena variabilidade de  $f_0$  e seus contornos foram em geral ascendentes; houve grande irregularidade microestrutural;

- Na felicidade, a taxa de elocução da fala foi alta e o tempo da execução musical acelerado; o nível sonoro apresentou-se médio-elevado; houve uma quantidade moderada/grande de energia na região de frequências altas do espectro; a variabilidade de  $f_0$  foi grande e seus contornos foram, em geral, ascendentes; os ataques vocais/tonais tenderam a ser rápidos; houve alguma irregularidade micro estrutural;

- Na tristeza, a taxa de elocução da fala foi baixa e o tempo da execução musical lento; houve pequena variabilidade do nível sonoro; a quantidade de energia na região de frequências altas do espectro foi pequena; a variabilidade de  $f_0$  foi também pequena e seus contornos tenderam a ser descendentes; os ataques vocais/tonais foram em geral lentos; houve bastante irregularidade microestrutural;

- Na ternura, a taxa de elocução da fala foi baixa e o tempo da execução musical lento; o nível sonoro foi baixo e apresentou pequena variabilidade; a quantidade de energia na região de frequências altas do espectro foi também pequena; a variabilidade de  $f_0$  foi reduzida e seus contornos em geral descendentes; os ataques vocais/tonais foram lentos; observou-se considerável regularidade microestrutural.

Recentemente tivemos a oportunidade de colaborar na realização de um estudo no qual comparamos configurações acústicas especificamente relacionadas à qualidade da voz de atores e cantores líricos profissionais interpretando diferentes emoções. (Scherer, Sundberg, Tamarit, & Salomão, 2015). Este estudo foi pioneiro ao utilizar o mesmo desenho metodológico e exatamente os mesmos parâmetros acústicos para comparar a expressividade vocal na fala e no canto. Mais uma vez, uma grande similaridade entre o modo como as emoções foram expressas nas duas modalidades pôde ser demonstrada. Emoções como a raiva e o medo-pânico refletiram-se na voz por meio de espectros mais planos, ou seja, espectros com maior quantidade de energia nas frequências altas, enquanto que emoções como a tristeza e o medo produziram espectros mais íngremes,

com pouca energia nas frequências altas.

Estudos experimentais vêm, portanto, apontando para semelhanças realmente instigadoras entre configurações acústicas na expressão de emoções na fala e na música/no canto, sugerindo uma relação muito próxima entre mecanismos expressivos em ambas as modalidades. A possibilidade de se definir um conjunto de descritores acústicos que consiga espelhar esta semelhança tem sido inclusive testada e em grande medida comprovada em estudos que mostram que algoritmos treinados para reconhecer emoções na música são bastante eficientes para também reconhecer emoções fala, e vice-versa (Weninger, Eyben, Mortillaro, Scherer, 2013; Eyben, Salomão, Sundberg, Scherer, & Schuller, 2015).

Tamanha semelhança na manifestação sonora de emoções pela voz e pela música nos faz indagar sobre o por quê de padrões vocais expressivos de base neurofisiológica, supostamente originados já nos primórdios da existência humana sob a forma de manifestações vocais espontâneas e de reações de adaptação ao ambiente, se assemelharem aos padrões de expressividade na execução musical, em seu status de manifestação artística e com níveis bem mais elaborados de criação estética.

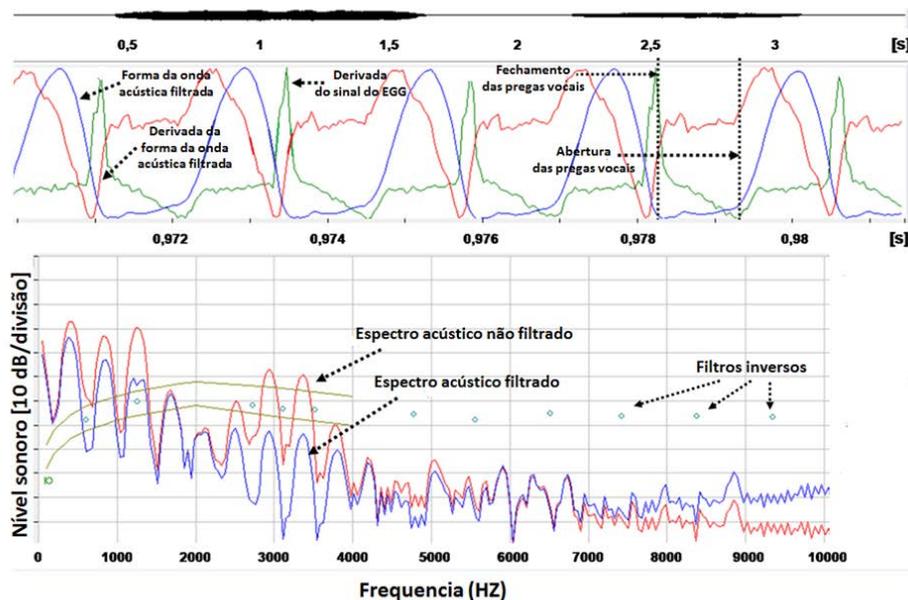
### **3.2. Semelhanças entre o som e o gesto vocal**

A voz que fala é a mesma voz que canta. Talvez seja possível alcançar melhor compreensão das similaridades expressivas entre a fala e o canto justamente pela investigação das semelhanças entre características do som vocal em ambas as modalidades e do gesto que o produz.

#### **3.2.1. O som da fonte glótica**

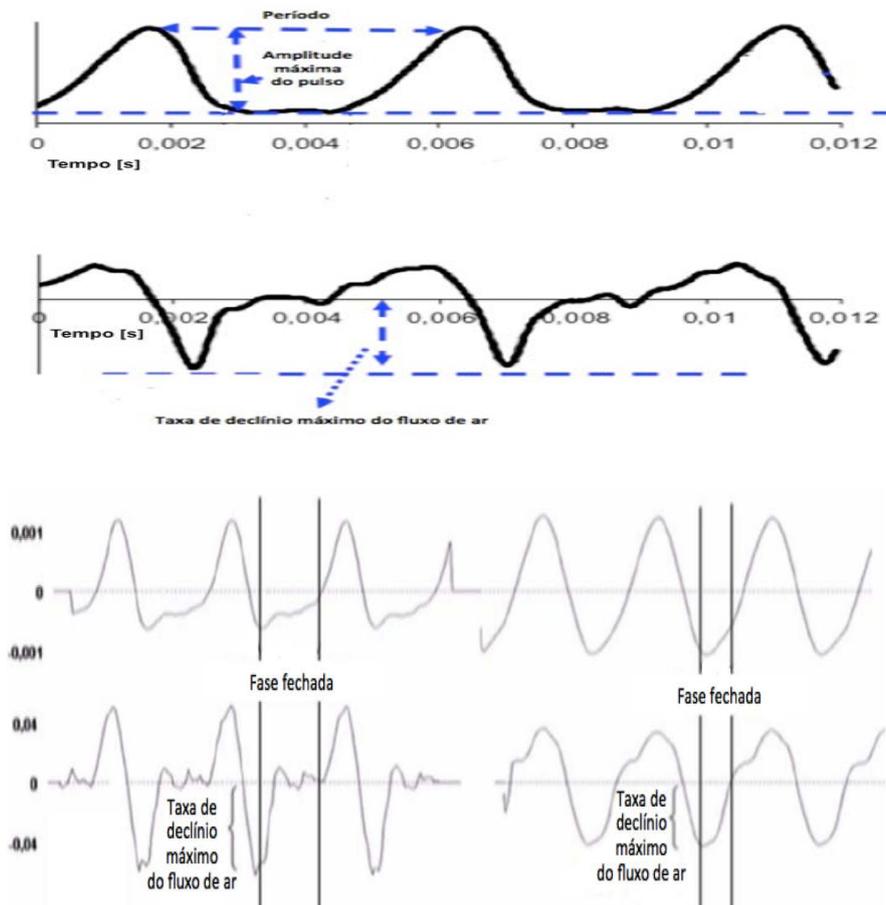
Embora relativamente pouco explorada, a possibilidade de se obter informações sobre as características acústicas do som produzido já no nível da laringe, ou seja, o som da fonte glótica, se mostra de grande relevância em estudos da expressividade vocal, especialmente no contexto de modelos que adotam uma perspectiva de base neurofisiológica. Analisar a fonte glótica significa aproximar-se da ponte que une componentes da expressão motora da emoção, neste caso, o gesto vocal que define os ajustes fonatórios e a matéria sonora tal qual gerada por estes ajustes, quando sob o efeito de uma ou de outra emoção. A interação entre o som gerado na glote, pela vibração das pregas vocais, e as ressonâncias do trato vocal, originadas pelas configurações articulatórias, definirá o espectro acústico do som irradiado pela boca, correlato do que percebemos como qualidade vocal.

A fonte glótica é representada por uma forma da onda do fluxo aéreo que atravessa a glote durante a fonação e pode ser estimada por meio da técnica da filtragem inversa. Esta técnica está baseada na teoria fonte-filtro que assume que, se a função de transferência do trato vocal for conhecida, é possível construir um filtro inverso de suas respectivas ressonâncias (Fant, 1970). Por meio da aplicação de filtros adequados no som da voz irradiado pela boca é possível estimar a forma de onda acústica produzida na fonte glótica. A filtragem inversa pode usada de maneira combinada à eletroglotografia (EGG), pois a derivada do sinal do EGG em geral permite uma visualização bastante clara de pequenas modificações relativas ao contato entre pregas vocais – representadas por picos (de abertura ou de fechamento das pregas vocais) – por vezes não tão nítidas no sinal estimado pela filtragem inversa (Henrich, D’Alexandro, Castellengo, & Doval, 2004; Salomão, Sundberg, 2008; Salomão, Sundberg, 2009). A sincronização dos dois sinais permite obter dados bastante robustos sobre a fonte glótica. A figura 1 mostra a forma de onda filtrada e o espectro sonoro antes e após a filtragem inversa.



**Figura 1** – Software Decap, desenvolvido pelo engenheiro acústico Svante Granqvist do Speech, Music and Hearing Department, Royal Institute of Technology.

A forma da onda da fonte glótica estimada pela filtragem inversa é comumente analisada por meio de parâmetros relacionados ao tempo (Bäckstöm, Alku, & Vilkmán, 2002), que indicam a duração de eventos no ciclo glótico (da fase de abertura e aberta, de fechamento e fechada, por exemplo), bem como a relação entre tais durações (quocientes fechado e aberto, por exemplo). A forma da onda glótica pode ser também analisada por meio de parâmetros relacionados à amplitude, os quais tem sido considerado mais robustos quando na captura de características relevantes da forma de onda estimada pela filtragem inversa (Alku&Vikman, 1996; Gobl, C. & Ní Chasaide, A., 2003). Um parâmetro de amplitude comumente utilizado é a “taxa de declínio máximo do fluxo de ar” (Fant, 1979; Alku e Vilkmán, 1996), facilmente visualizada como o pico negativo da derivada da forma da onda acústica, como mostra o gráfico inferior da figura 2. Esta medida representa a excitação máxima do trato vocal pelo fluxo de ar a qual é decisiva na determinação da intensidade da voz (Fant, 1979; Alku e Vilkmán, 1996a). O gráfico à esquerda da figura 3 mostra que picos negativos mais acentuados e definidos na derivada da forma da onda correspondem a formas de onda glótica mais angulosas, caracterizadas por mudanças mais bruscas no fluxo de ar que atravessa a glote e fases fechadas mais definidas; tais formas de onda tendem a produzir espectros de onda glótica mais planos, ou seja, com menor diferença entre quantidade de energia nos parciais graves e agudos. O gráfico à direita da mesma figura mostra, por outro lado, que picos negativos mais suaves e menos definidos na derivada correspondem a formas de onda mais arredondadas, caracterizadas por mudanças mais suaves no fluxo de ar que atravessa a glote e fases não fechadas de fato, mas “quase fechadas”; tais formas de onda produzirão espectros acústicos mais íngremes, ou seja, com maior diferença entre quantidade de energia nos parciais graves e agudos (Fant&Lin, 1988).



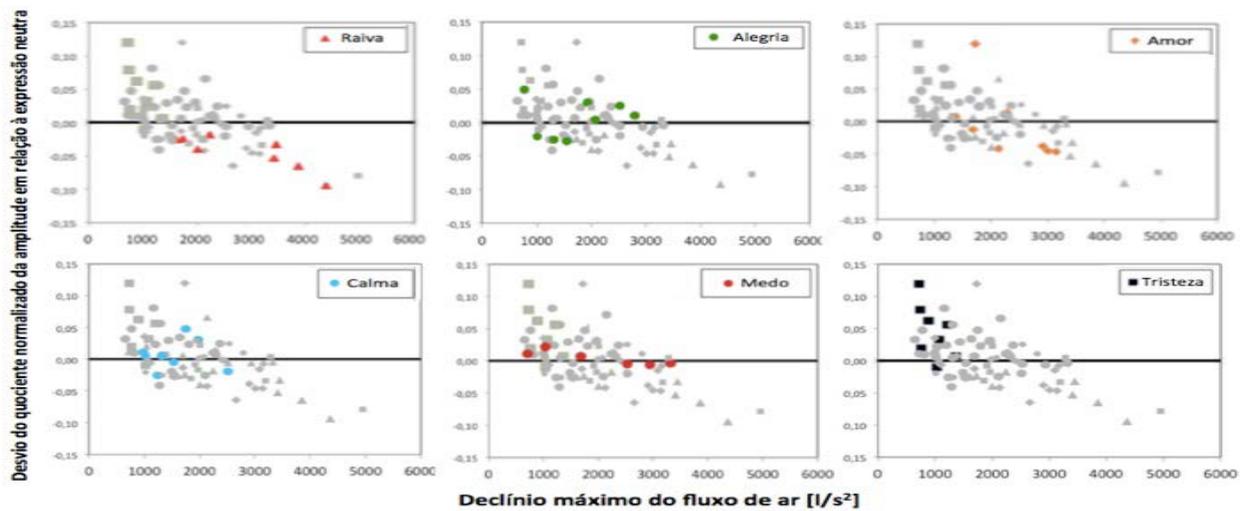
Outro parâmetro de amplitude que também vem sendo frequentemente utilizado é o quociente da amplitude. Este parâmetro fornece uma caracterização global do pulso glótico por meio de duas medidas de amplitude: a amplitude máxima do fluxo de ar (gráfico superior da figura 2) e a taxa do declínio máximo do fluxo de ar no ciclo glótico. Quando normalizado em relação à frequência fundamental da onda sonora, este parâmetro recebe o nome de “quociente de amplitude normalizado”. O quociente de amplitude normalizado tem se mostrado bastante sensível ao grau de adução das pregas vocais – maior o grau de adução, menor o valor do quociente da amplitude; e vice-versa – e à qualidade vocal, mais ou menos tensa, dentro de um contínuo que tem em seus extremos a voz tensa e a voz soprosa (Alku, P., & Vikman, E., 1996).

### 3.2.2. Som e ajustes vocais

O impacto de diferentes emoções sobre os ajustes laríngeos e sobre a fonte glótica durante o canto foi recentemente investigado em um cantor lírico profissional de renome (Salomão, Sundberg, & Scherer, 2015a; 2015b) ao expressar diferentes emoções por meio de emissões cantadas com vogal /a/. Por meio de uma avaliação perceptiva, ouvintes, com ou sem formação musical, foram solicitados a indicar a emoção que cada uma destas emissões veiculava. Ouvintes conseguiram identificar as emoções em acordo com a intenção do cantor com frequência bastante superior a que teria sido observada se as respostas tivessem ocorrido ao acaso, mostrando que as amostras de canto coletadas eram de fato representativas das emoções expressas. Os sinais acústicos foram captados por meio de um microfone posicionado a certa distância dos lábios do falante ou cantor e submetidos à filtragem inversa. As formas de onda glótica obtidas foram analisadas por meio de diversos parâmetros acústicos, de tempo e de amplitude.

A figura 4 mostra os resultados obtidos para os parâmetros de amplitude “taxa de declínio máximo do fluxo de ar” e “quociente de amplitude normalizado” na expressão de seis das emoções

analisadas. Os valores da taxa de declínio máximo do fluxo de ar foram normalizados em relação a possíveis dependências da frequência fundamental. No gráfico, os resultados estão apresentados em termos de desvios dos valores obtidos nas diferentes emoções em relação aos obtidos na emoção “neutra”, correspondente à linha horizontal que cruza o zero. Valores situados acima desta linha indicam valores maiores de quociente da amplitude normalizada do que os obtidos na emoção neutra e vice-versa.



A figura mostra que as emoções raiva e tristeza encontram-se nos extremos de um contínuo acústico-fisiológico-emotivo. De todas as emoções, a expressão vocal da raiva foi a que utilizou maiores graus de adução da glote (valores menores do quociente da amplitude) e taxas mais elevadas no declínio do fluxo de ar no ciclo glótico, sugerindo intensidades de voz comparativamente bastante mais elevadas. No outro extremo encontra-se a tristeza, expressa por meio de ajustes fonatórios associados a graus de adução bastante reduzidos em comparação aos encontrados na emoção neutra, e intensidade de voz baixa. A figura mostra também que este cantor expressou a alegria, a calma e o amor por meio de ajustes fonatórios comparativamente mais próximos aos da emissão neutra e relativamente semelhantes, ainda que pareça haver uma tendência a graus de adução um pouco maiores na alegria e na calma do que no amor. Além disso, para este cantor, a variação dos ajustes utilizados na expressão do amor foi maior do que na alegria. De todas as emoções, o medo foi a que apresentou ajustes laríngeos mais próximos aos utilizados na expressão neutra.

É preciso sempre tomar muita cautela a respeito de qualquer generalização de dados obtidos para apenas um cantor, um único estilo e/ou gênero musical, um contexto cultural específico. Por outro lado, é interessante notar que os resultados encontrados neste estudo encontram respaldo bastante sólido no modelo teórico proposto por Scherer, que prediz que emoções caracterizadas pelo aumento da atividade motora e do nível de tensão muscular, como a raiva, devem provocar efeitos tais como um aumento da pressão subglótica, da velocidade da passagem do ar através pela glote e das forças de adução das pregas vocais. Do ponto de vista perceptivo, a voz da raiva deste cantor soaria tensa (valores reduzidos do quociente da amplitude normalizada) e intensa (valores elevados da taxa do declínio do fluxo de ar), como a própria raiva expressa. No outro extremo, encontra-se a tristeza, caracterizada por uma diminuição geral da atividade motora e do nível de tensão muscular, e também por pressões subglóticas menores, velocidades mais baixas da passagem do fluxo de ar pela glote, menor força de adução das pregas vocais. A voz da tristeza, tal qual expressa por este cantor, soaria débil (valores elevados do quociente da amplitude normalizada, valores reduzidos da taxa do declínio do fluxo de ar).

É ainda interessante notar a proximidade dos ajustes laríngeos utilizados por este cantor para expressar tristeza e o amor, o que confirma dados da literatura que mostram que a diferencia-

ção entre a tristeza e o amor/a ternura tende a ser mais difícil, tanto na fala como no canto (Baars & Gabrielsson, 1997; Juslin, 1997c).

Está claro que as relações entre o componente da expressão motora de emoções do modelo de Scherer e os ajustes laríngeos captados pelos parâmetros de amplitude da fonte glótica são muito mais complexas do que o que descrevemos aqui. O modo como os diversos componentes interagem no processo da emoção e determinam a dinâmica dos padrões motores da expressão vocal em toda a sua gama de possibilidades (para uma visão detalhada sugerimos investigar diretamente Scherer 1986; 2001) parece ser tão complexo como as relações entre os próprios ajustes laríngeos na expressão de diferentes emoções e as características acústicas do som que eles produzem.

Definir um conjunto de parâmetros que consiga captar os aspectos acústicos mais relevantes na expressão e para o reconhecimento de todo um espectro de emoções constitui, ainda, um grande desafio. Talvez o fato de, por definição, a qualidade vocal apresentar-se como uma característica presente quase que permanentemente no som da voz (Abercrombie, 1967; Laver, 1968) dificulte, justamente, a captura e a sistematização de possíveis sutilezas nas gradações no contínuo acústico/sonoro vocal, aquelas que, no nível expressivo, possivelmente veiculariam as gradações dos estados emotivos.

#### **4. A Fala no Canto**

Ao tratar da expressividade da voz, Johan Sundberg, pesquisador de linha essencialmente experimental e com extensa produção científica sobre a produção e a percepção da voz cantada, argumenta que resultados de pesquisas quantitativas só podem ser bem compreendidos se a fala ou o canto expressivo forem entendidos como uma projeção sonora dos gestos fonatórios (Sundberg, 2001; 2015), pois, justamente, seria a conexão entre o gesto vocal expressivo, designado pelos ajustes laríngeos e articulatórios associados à emoção expressa, e a matéria sonora que ele produz, definiria, o código vocal expressivo de natureza mais contínua e icônica. A intensidade da voz, por exemplo, se modifica continuamente, e não discretamente, em geral de acordo com a “intensidade” da emoção.

A apropriação desse código, que conecta som e gesto, nos permitiria “compreender” a emoção expressa pelo falante ou cantor, e, sem ele, talvez simplesmente não pudessemos nos emocionar diante de vozes, cantos e músicas.

É importante novamente notar que a similaridade icônica entre a expressão vocal e musical de emoções muito possivelmente se aplique apenas a certos parâmetros acústicos, ou seja, àqueles que o musicista-intérprete possa controlar (mais ou menos livremente) durante a execução musical, como o tempo, a intensidade, o timbre. Até mesmo o parâmetro “frequência”, transcrito nas notas musicais que definem a estabilidade melódica da obra, estaria este também, sujeito a manipulações pelo musicista-intérprete, como mesmo apontado na literatura (Juslin&Laukka, 2003). Sugestivamente, o estudo de Scherer, Sundberg, Tamarit, & Salomão (2015), que mostrou várias similaridades entre configurações expressivas na fala e no canto, apontou também para divergências muito interessantes entre as duas modalidades. Tais divergências foram captadas pelos parâmetros de perturbação, em geral bem mais elevados no canto do que na fala para certas emoções, apontando claramente para a importância do vibrato no primeira modalidade. É interessante notar que tal divergência diminuiu consideravelmente quando vocalizes no canto, inicialmente comparados a sentenças faladas, foram então comparados a “explosões de emoção” na fala. Atores parecem ter feito uso consideravelmente maior de aperiodicidade como recurso expressivo nesta segunda situação, ou seja, quando na ausência das variações suprasegmentais próprias ao contexto linguístico. Na discussão dos resultados deste estudo, sugeriu-se que o vibrato viesse a constituir um recurso expressivo por meio do qual o cantor teria a possibilidade de se sobrepor às

restrições que a própria estrutura musical lhe impõe. Tais restrições se apresentariam muitíssimo mais diluídas na fala, bem mais irregular e instável em sua natureza. Neste sentido, o vibrato constituiria uma maneira de se fazer jus ao potencial expressivo de certas emoções sem desrespeitar a estrutura melódica que confere identidade à obra musical, uma estratégia para imprimir na melodia prevista marcas de maior ou menor irregularidade, de maior ou menor “tensão” expressiva. De fato, ainda que relativamente pouco estudado em suas diferentes formas de manifestação, o vibrato parece apresentar-se como um recurso expressivo importantíssimo no canto em geral, independentemente de gênero ou estilo musicais.

Finalmente, acreditamos que ao apossar-se de um código vocal expressivo que exhibe similaridades icônicas entre a expressão de emoções e a própria emoção expressa, o cantor se apossaria também da obra musical enquanto a executa, imprimindo nela suas próprias marcas expressivas. Prolongando ou acelerando a duração das vogais, ampliando ou encurtando a extensão dos saltos entre intervalos, diminuindo ou elevando a intensidade sonora, reduzindo ou enfatizando a energia nos componentes mais agudos do espectro sonoro, o artista traria para a estrutura do canto lastros de padrões expressivos provenientes do código expressivo da fala. Tais marcas expressivas, verdadeiros “fósseis vivos” presentes na expressão vocal, até certo ponto universais, teriam origem já nos primórdios da evolução humana pela necessidade de sinalizar as emoções em relações sociais e em situações básicas de sobrevivência (Scherer, 1986; 1991). As mesmas constituiriam, por assim dizer, a base de um código expressivo que se estenderia da fala ao canto. De fato, Helmholtz (1863/1954), um dos pioneiros da psicologia musical, já havia afirmado que teria sido a imitação de modulações involuntárias da voz falada o que possivelmente levou nossos ancestrais a descobrir as primeiras formas de expressão musical (p. 371). Trazer para o canto lastros de instabilidade da fala faria a realização musical soar verdadeira (Tatit, 1994).

## Referências Bibliográficas

- ABE ALKU, P.; VIKMAN, E. Amplitude domain quotient for characterization of the glottal volume velocity waveform estimated by inverse filtering. **Speech Communication**. 18, v.02, p. 131-138, 1996.
- BAARS.; GABRIELSSON. Emotional expression in singing: A case study. In A. Gabriellson (Ed.), **Proceedings of the third Triennial ESCOM Conference**, 1997. Uppsala. Uppsala, Sweden: Uppsala U, p. 479– 483, 1997.
- BÄCKSTÖM, T.; ALKU, P.; VILKMAN, E. Time-domain parameterization of the closing phase of glottal airflow waveform from voices over a large intensity range. **IEE Trans Speech Audio Process**, 10, v.03, p. 186-192, 2002.
- BLOCH, S.; ORTHOUS, P.; SANTIBANEZ, G. Effector patterns of basic emotions: A psychophysiological method for training actors. **J of Social Biology and Structure**, v.10, p.1–19, 1987.
- BUCK, R. **The Com of emotion**. New York, Guilford Press, 1984.
- DAMASIO, A. R. et al. Subcortical and cortical brain activity during the feeling of self-generated emotions. **Nature NeuroSci**, v. 03, p.1049–1056, 2000.
- EYBEN, F. et al. Emotion in The Singing Voice. A Deeper Look at Acoustic Features in the Light of Automatic Classification. **EURASIP J of Audio, Speech, and Music Processing**, 19, 2015.
- EKMAN, P. An argument for basic emotions. *Cognition & Emotion*, 6, 169–200.
- ELFENBEIN, H. A.; AMBADY, N. On the universality and cultural specificity of emotion recognition: A meta-analysis. **Psych Bulletin**, p.128, 203–235, 2002.
- FANT, G. **Glottal source and excitation analysis**. **Speech transmission Laboratory Quartely Progress and Status Report**. .v 01, p. 85-107, 1979.
- \_\_\_\_\_; LIN, Q. Frequency domain interpretation and derivation of glottal flow parameters. **Speech Transmission Laboratory Quarterly Progress Report**. 2- 3, v.85, 1-21, 1988.
- \_\_\_\_\_. *Acoustic theory of speech production*. 2 ed. Paris: Mouton, 1970.
- FONAGY, I. **La vive Voix**. Paris, Editions Payot, 1983.
- FRIDLUND, A. J.; SCHWARTZ, G. E.; FOWLER, S. C. Pattern recognition of self-reported emotional state from multiple-site facial EMG activity during affective imagery. **Psychophysiology**, v.21, p. 622–637, 1984.
- GOBL, C.; NÍ CHASAIDE, A. Amplitude-based source parameters for measuring voice quality. In: **Proceedings of the ISCA VOQUAL'03 Workshop on Voice Quality: Functions, Analysis and Synthesis**, 2003, Geneva. Geneva: p.151-15, 2003.
- \_\_\_\_\_; NÍ CHASAIDE, A. Amplitude-based source parameters for measuring voice quality. In: **Proceedings of the ISCA VOQUAL'03 Workshop on Voice**, 2003.
- HELMHOLTZ, H. L. F. von. **On the sensations of tone as a Psych Psych basis for the theory of music**. New York: Dover. (1863), 1954.
- HENRICH, N. et al. On the use of the derivative of electroglottographic signals for characterization of no pathological phonation. **J Acoustic Soc Am**. 115, v.03, p. 1321-1332, 2004.
- IZARD, C. E. Basic emotions, relations among emotions, and emotion– cognition relations. *Psych Review*, v.99, p. 561–565, 1992.
- \_\_\_\_\_. Vocal expression of emotion. In: **Hdbk of affective Scis**. New York: Oxford U, 2009.
- JOHNSON-LAIRD, P. N.; OATLEY, K. Basic emotions, rationality, and folk theory. **Cognition & Emotion**, v. 6, p. 201–223, 1992.
- JUSLIN, P. N.; LAUKKA, P. Com of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code? **Psych Bulletin**, v. 129, n. 5, p. 770–814, 2003.
- \_\_\_\_\_. Perceived emotional expression in synthesized performances of a short melody: Capturing the listener's judgment policy. **Musicae Scientiae**, v.01, p. 225– 256, 1997.

- LAVIER, J. Voice quality and indexical information. **British J of Disorders of Com.** v.03, p.43-54, 1968.
- LEVENSON, R. W. Human emotion: A functional view. In P. Ekman & R.J. Davidson (Eds). **The nature of emotion: Fundamental questions.** New York, Oxford U P, p.123–126, 1994.
- PHAN, K. L.; et. al. Functional neuroanatomy of emotion: A meta analysis of emotion activation studies in PET and fMRI. **NeuroImage**, v.16, p.331–348, 2002.
- RCROMBIE, D. Elements of General Phonetics. Edinburgh: Edinburgh U P, 1967.
- SALOMÃO, G.L.; SUNDBERG, J.; SCHERER, K. Emotional coloring of the singing voice. In: **Proceedings of the 11th Pan-European Voice Conf PEVOC**, Firenze: Firenze U P, p. 80, 2015.
- \_\_\_\_\_. Emotional coloring in singing: An in-depth spectral analysis of a tenor's voice. In: **Proceedings of the IV Intenational Conf on Music and Emotion**, Geneva: U of Geneva, G. U. o. M., p. 65, 2015.
- \_\_\_\_\_; SUNDBERG, J. Relation between perceived voice register and flow glottogram parameters in males. **The J of the Acoustical Society of America**, 124, v.01, 546-551, 2008.
- \_\_\_\_\_. What do male singers mean by modal and falsetto register? An investigation of the glottal voice source. **Logopedics Phoniatics Vocology**, 34, v.02, 73-83, 2009.
- SCHERER, K. R. Appraisal considered as a process of multi-level sequential checking. In: \_\_\_\_\_; SCHORR.; T. Johnstone (Eds.), Appraisal processes in emotion: **Theory, methods, Res.** New York: Oxford U P, p. 92–120, 2001.
- \_\_\_\_\_; et al. Comparing the acoustic expression of emotion in the speaking and the singing voice. **Computer Speech and Lang.**, 29, p. 281-235, 2015.
- \_\_\_\_\_. Vocal Affect Expression: A Review and a Model for Future Res. **Psych Bulletin**, v. 99, n. 2, p.143-165, 1986.
- SPENCER, H. **The origin and function of music.** Fraser's Magazine, 56, p.396–408, 1857.
- SUNDBERG, J. **Ciência da Voz: Fatos sobre a voz na fala e no canto.** Tradução Gláucia Lais Salomão. EDUSP, 2015.
- TATIT, L. **Semiótica da Canção: melodia e letra.** Ed. Escuta, 1994.
- WENINGER, F.; et. al. On the acoustics of emotion in audio: what speech, music, and sound have in common. **Front. Psychol**, 4, p.1-12, 2013.

***Efeitos das Metáforas no Ensino do  
Canto: Dados de Imagem do  
Trato Vocal***

CRISTINA CANHETTI  
ZULEICA CAMARGO

# 1. Introdução

A voz cantada fascina os mais diversos grupos de pesquisadores, justamente pelo uso altamente diferenciado e refinado do trato vocal, que congrega estruturas dos aparelhos respiratório e digestivo, as quais, no dia-a-dia, também desempenham as funções de fala<sup>1</sup>.

A origem do canto e da fala confundem-se e ambas estão relacionadas à necessidade de externalizar afeto e, portanto, carregam marcas fortes de emoção (Salomão, 2008). Durante o canto, quando há harmonia entre os aspectos explorados, é possível gerar diferentes sensações positivas no ouvinte, bem como inúmeras emoções durante sua interpretação (Andrada E Silva, Duprat, 2004).

Para o canto, utilizam-se as mesmas estruturas do trato vocal durante a fala, porém com diferentes mobilizações de acordo com as necessidades da produção da voz cantada (Pedroso, 1997). A emissão da voz cantada requer adaptações de pressão subglótica, de ajustes laríngeos e do trato vocal, visando uma produção estética, agradável e livre de mecanismos que possa gerar esforço vocal (Sundberg, 1990).

Da mesma forma que se admite que toda manifestação de fala é expressiva (Madureira, 2005), o mesmo valeria para as manifestações da voz cantada. A construção da expressividade pauta-se na combinação de elementos prosódicos (ritmo, entoação, qualidade vocal, entre outros) a elementos segmentais (vogais e consoantes), mediados pelas relações entre o som e o sentido (Madureira, 2005).

No que se refere ao enfoque das relações entre som e sentido, podemos apontar as metáforas como representações simbólicas (Hinton et al, 1994), dotadas de sensações proprioceptivas, motoras e táteis (Fonagy, 2000). Tais estudos são baseados em modelos de descrição fonética, com base em descrições perceptivas, acústicas e fisiológicas, no campo das Ciências Fonéticas.

Os estudos de voz cantada apontam que as metáforas, incorporadas inicialmente pelos professores de canto com sua base na prática de ensino. Revelaram-se, gradativamente, como recursos tanto para facilitar a memorização de sensações físicas (Rossbach, 2011), quanto para auxiliar a criação de correlações entre o simbólico e o fisiológico (Sousa et al, 2010).

Dessa forma, no universo do ensino do canto, as metáforas foram incorporadas enquanto estratégia para levar o aluno/cantor a vivenciar ajustes no trato vocal e sensações promovidas. Para vários destes alunos/cantores, as instruções baseadas somente na ação de grupos musculares, portanto na fisiologia da voz cantada, seriam de difícil compreensão, bem como de execução (Sousa et al, 2010).

Segundo Clements (2008), existem três tipos de métodos utilizados pelos professores de canto. O primeiro é embasado exclusivamente na anatomia e nos estudos científicos. O segundo utiliza-se de analogias e de imagens para realizar o aprendizado, tendo sido bastante utilizado em Portugal em meados do século XX, como ensino tradicional para o aprendizado (Aguiar, 2007). Por fim, o método que combina ambos os métodos supracitados, isto é, alia o ensino tradicional ao ensino científico (Aguiar, 2007).

Para Clements (2008), o terceiro método resultaria em melhores resultados de aprendizado, enquanto para Aguiar (2007) permitiria utilizar o aparelho fonador de maneira harmônica e estimularia um aprendizado global e estruturado.

Dessa forma, o professor de canto busca por estratégias que possam desencadear os ajustes de trato vocal (entendidos do ponto de vista fonético como tendências musculares recorrentes na produção vocal). Tais ajustes envolvem ações da musculatura respiratória, daquela da região laríngea (ajustes fonatórios) e da região supraglótica (ajustes articulatórios), além dos níveis gerais

---

<sup>1</sup> Essa pesquisa foi desenvolvida com auxílio bolsa FAPESP/SP

de tensão muscular (ajustes de tensão muscular) na concepção do modelo fonético de descrição da qualidade vocal (Laver, 1980).

As Ciências Fonéticas e, particularmente, a Fonética Acústica, têm se dedicado ao estudo dos variados efeitos sonoros que o aparelho fonador humano pode produzir ao modular a corrente de ar expiratória. Pode ser aplicada tanto para finalidades de produção da emissão da voz falada, quanto cantada (Kent, 1993; Sundberg, 1997).

Grande parte dos professores de canto refere que as metáforas geram efeitos positivos na “musicalidade do aluno e que podem auxiliar no desenvolvimento da imitação vocal, variações de timbre, extensão vocal, afinação, inteligibilidade e homogeneidade da voz, fraseado, dinâmica, flexibilidade da voz e intenção interpretativa” (Sousa et al, 2010, p. 323). Além disso, apontam que nem sempre uma mesma metáfora pode gerar bons resultados, em alguns casos, há necessidade de modificar ou adaptar o tipo de metáfora para cada aluno.

Segundo Medeiros (2002), existem diferenças entre a fala e o canto, principalmente no que diz respeito à inteligibilidade durante a produção do texto cantado e às modificações formânticas. Tais achados reforçam a demanda por enfoque de estratégias que visem estimular ao aluno/cantor a implementar tais diferenciações em suas produções de voz cantada.

Em relação aos estudos de imagens do trato vocal durante a produção vocal, as pesquisas são ainda escassas, apesar da técnica permitir uma detalhada descrição complementar dos vários níveis de atividade do trato vocal durante a emissão vocal.

Alguns estudos têm utilizado as imagens videofluoroscópicas para descrever mudanças fisiológicas durante as situações de fala ou canto (Gonçalves, 2002; Nuñez, 2006; Souza, 2008; Drahan et al, 2012).

Enquanto técnica de diagnóstico por imagem, a videofluoroscopia é um método que registra os eventos biológicos dinâmicos, gerados pela exposição de pacientes a baixos índices de radiação como fonte de imagens. Além disso, tais registros podem ser gravados e permitem traçar estimativas em tempo, dimensão e velocidade reais dos eventos observados, tais como a dinâmica da deglutição, da fala e canto (Firman et al, 2000; Costa et al, 2003).

Segundo Lopasso (2000), a videofluoroscopia agrega a medidas bidimensionais, modelos em que a densidade serve de parâmetro para a estimativa da avaliação por imagem em terceira dimensão. Tal modelagem permite a obtenção de valores volumétricos em áreas de interesse de estudo. A videofluoroscopia pode proporcionar subsídios importantes para a avaliação detalhada dos articuladores (lábios, língua, palato duro, palato mole) e cavidades ressoadoras (labial, oral, faríngea e nasal) durante a emissão da voz cantada.

A técnica possibilita a análise das estruturas do trato vocal por meio de gravação de imagens dinâmicas, de forma a permitir, rever e discutir seus dados e resultados com outros profissionais/especialistas utilizando recursos tecnológicos e sem a necessidade de manter o sujeito ou paciente exposto à radiação por mais tempo (Bilton, Lederman, 1998).

Sendo assim, este projeto focaliza a avaliação fisiológica (articulatória) por meio de imagens videofluoroscópicas no campo da voz cantada, tendo como fundamentação a Fonética Articulatória. Estudo desta natureza permite o detalhamento do refinamento de ações sob as quais o trato vocal é submetido nas situações de canto e pode colaborar para a compreensão de particularidades do processo de ensino e prática do canto.

Dessa forma, tais achados permitem inferir sobre os níveis de mobilização das estruturas do trato vocal, diante do estímulo das metáforas. O conhecimento gerado poderá, futuramente, auxiliar outras categorias de profissionais da voz que também demandam um uso altamente refinado do trato vocal, bem como colaborar para a discussão em torno das relações entre som e sentido e, finalmente, do ensino de canto em nosso meio.

Este trabalho tem por objetivo investigar, em amostras de voz cantada, os efeitos do uso das

metáforas utilizadas pelos professores de canto, a partir de descrições articulatórias (imagens do trato vocal por meio da videofluoroscopia) de qualidade vocal em uma mulher não cantora.

## 2. Métodos

Os procedimentos de coleta e de análise de dados foram conduzidos com a participação de um sujeito do sexo feminino, de 26 anos de idade, não cantora. Os critérios de inclusão para seleção referiram-se ao sexo e à faixa etária, bem como à manifestação de interesse em colaborar com a sessão de coleta de imagens. Os critérios de exclusão compreenderam a autorreferência de queixas vocais e de quadros inflamatórios/infecciosos de vias aéreas no momento da coleta, bem como de diagnóstico otorrinolaringológico e fonoaudiológico de alteração vocal no período. Além disso, foram considerados critérios de exclusão aqueles referentes às contraindicações para a realização de exames de imagem desta natureza: ter realizado esse exame recentemente, mais especificamente, no período de seis meses; e, no caso de mulheres, encontrar-se em período gestacional.

### 2.1. Procedimentos de coleta de amostras de imagem do trato vocal durante a emissão em voz cantada

A coleta de dados foi realizada em uma sessão, no serviço de Radiologia de um hospital privado da cidade de São Paulo.

Durante a sessão de coleta, o sujeito permaneceu sentado em sala do Serviço de Radiologia em que se utilizou equipamento *Philips, Tele Diagnost-Optimus*; LG monitor, LG DVD-HDD, monitor *Philips* com sistema de gravador/monitor para registro em DVD. Antes de se iniciar o procedimento de registro de imagens, foi apresentado ao sujeito o registro em áudio de um trecho da canção infantil “Sapo Jururu” gravado por uma professora de canto como referência. Segue a canção abaixo:

*Sapo Jururu, da beira do rio,  
quando o sapo canta: ó, maninha  
é que está com frio.  
A mulher do sapo deve estar lá dentro,  
fazendo rendinhas, maninha,  
para o casamento.  
(Palavra Cantada. 1998)*

Para a composição do *corpus* foram concebidas quatro tarefas. A primeira tarefa referiu-se a cantar o trecho da canção “Sapo Jururu, da beira do rio. Quando o sapo canta, ó maninha, é que está com frio” sem uso de metáforas. A segunda, terceira e quarta tarefas referiram-se a cantar o mesmo trecho da canção com o uso das metáforas prisma, catedral e disco voador respectivamente.

Para as tarefas 2, 3 e 4 foram utilizadas ilustrações para facilitar a compreensão das metáforas, de forma a auxiliar o sujeito para a emissão em voz cantada.

Nos estudos de Sousa et al (2008) encontramos a metáfora “cúpula de uma igreja dentro da boca”, utilizada nesse estudo como “catedral”; “ressonância como uma lâmpada”, utilizada nesse estudo como o “prisma”, uma vez que existe uma lâmpada que ao acender ilumina o ambiente ou em relação ao prisma espelha a voz como uma luz refletida, e, por fim, “ressonância como uma bolha de sabão”, utilizada nesse estudo como o “disco voador”, já que ambas remetem a algo que flutua com leveza.

O tempo total de exposição à radiação foi de 1 minuto e 46 segundos. Vale reforçar que durante os intervalos, entre uma emissão e outra, e durante as instruções, não houve exposição à radiação. Para que o tempo de exposição fosse breve, não houve solicitação de repetição de tarefas de emissão em voz cantada.

## 2.2. Procedimentos de edição, medidas de imagem e análise dos dados de imagem do trato vocal

Do ponto de vista de imagens de trato vocal, os registros videofluoroscópicos foram submetidos a processo de decupagem para análise das mobilizações com o auxílio do software de livre acesso VirtualDub. Após esse processo, foi possível efetuar a inspeção prévia do conjunto de movimentos de lábios, língua, palato mole, paredes de faringe e laringe. Para tanto, as imagens em formato .pap foram analisadas por meio do software de livre acesso Osiris (<http://en.softonic.com/s/osiris-medical-imaging:mac>) e as medidas de deslocamento de articuladores (lábios, mandíbula e língua) foram extraídas em milímetros (mm).

O conjunto de medidas definido para esta etapa de coleta de dados constou de: abertura de lábios absoluta; abertura de lábios normalizada; abertura de mandíbula; distância dorso de língua-faringe; distância dorso de língua-palato. Além disso, foi definido um elemento qualificador intitulado deslocamento ântero-posterior de língua. As medidas foram definidas a partir de estudos que utilizaram a técnica de ressonância magnética para investigações da voz cantada (Sundberg, 2008) pelo *software* Osiris e outro pela técnica de radiografia (Ericsson et al, 1998) com o uso do *software* APEX-model e são apresentadas abaixo:

- Abertura de lábios absoluta (figura 1): distância demarcada entre dente incisivo central superior e o dente incisivo central inferior



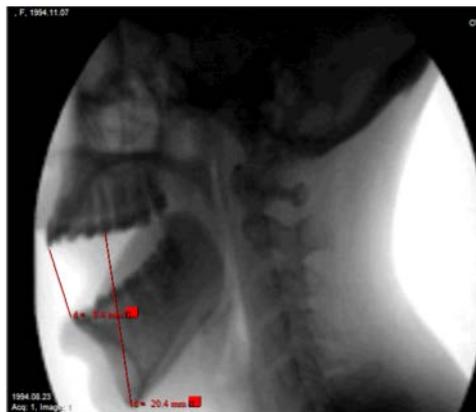
**Figura 1** – Medida de abertura de lábios absoluta (mm) gerada pelo software Osiris

- Abertura de mandíbula (figura 2): distância entre o centro da Glabela do osso frontal (parte óssea acima da órbita ocular – na testa) até o mento (queixo).



**Figura 2** – Medida de abertura de mandíbula (mm) gerada pelo software Osiris

- Abertura de lábios normalizada (figura 3): medida de abertura de mandíbula (figura 2) subtraída do valor da distância entre o dente segundo pré-molar superior e o queixo. Após o resultado dessa subtração, o valor gerado é subtraído do valor de abertura de lábios absoluta (figura 1) (Sundberg, 2008).



**Figura 3** – Medida de abertura de lábios normalizada (mm) gerada pelo software Osiris

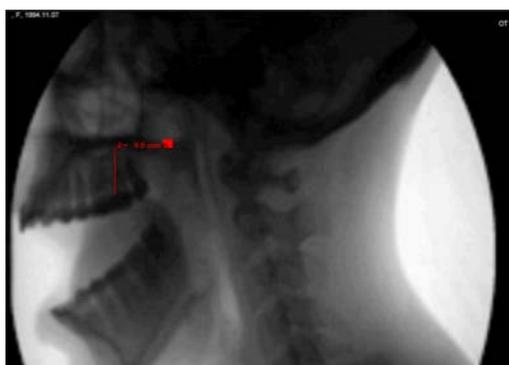
Vale ressaltar que o traçado referente ao cálculo da distância entre o dente segundo pré-molar superior e o mento (queixo) – medida necessária ao cálculo de abertura lábios normalizada – foi efetuado separadamente, pois, caso contrário, haveria sobreposição na imagem e excesso de informações na mesma figura, dificultando a observação da base de cálculos utilizada na presente pesquisa.

- Distância dorso de língua-faringe (figura 4): distância entre o início de base de língua até a segunda vértebra cervical (Echternach et al, 2008; Sundberg, 2008).



**Figura 4** – Medida da distância de língua-faringe (mm) gerada pelo software Osiris

- Distância dorso de língua-palato: localização do ponto mais alto de corpo de língua em relação ao palato duro. Em caso de vogais anteriores (figura 5), a medida era realizada no centro do palato duro. Para vogais posteriores (figura 6), a medida apontava o final do palato duro como referência.



**Figura 5** – Medida da distância de língua-palato de uma vogal anterior [ɛ] (mm) gerada pelo software Osiris



**Figura 6** – Medida da distância de língua-palato de uma vogal [a] posterior (mm) gerada pelo software Osiris

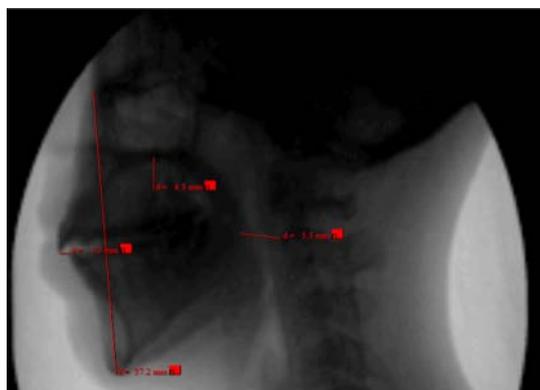
#### • Elemento qualificador

- Deslocamento ântero-posterior de língua: parâmetro de classificação de julgamento de deslocamento de dorso de língua em relação a região central do palato duro (1=anterior) ou região posterior do palato duro (2=posterior).

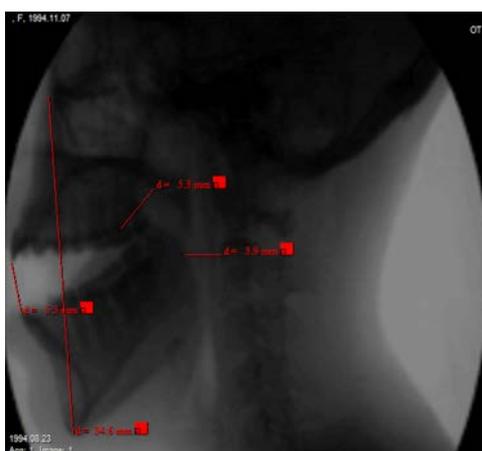
Os dados foram submetidos a análise *estatística por* meio de análise multivariada (análise discriminante das metáforas e das vogais emitidas, a partir das medidas de imagens do trato vocal, teste de correlação de Pearson ( $\alpha= 0,005$ ), análise fatorial, análise aglomerativa *hierarquia* de cluster, análise de correlação canônica, análise de regressão linear e ancova) .

### 3. Resultados

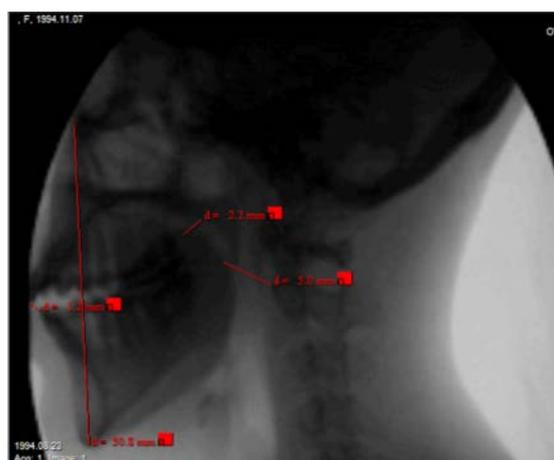
As imagens do trato vocal em situação de repouso, durante a emissão de voz cantada produzida sem instrução baseada em metáforas e durante a emissão em voz cantada produzida por meio de instruções baseadas em metáforas (prisma, catedral e disco voador) são apresentados nas Figuras 7 a 11.



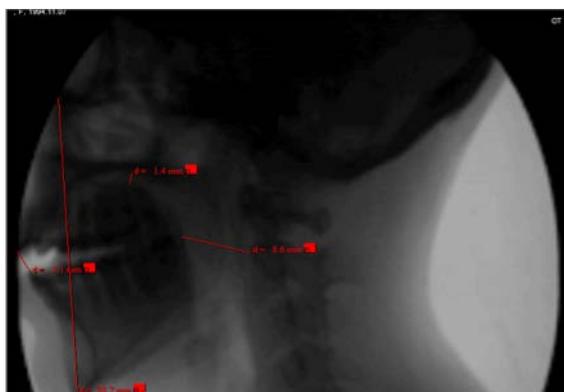
**Figura 7** – Dados de imagem de trato vocal ilustrativos das medidas (em mm) extraídas na situação de repouso



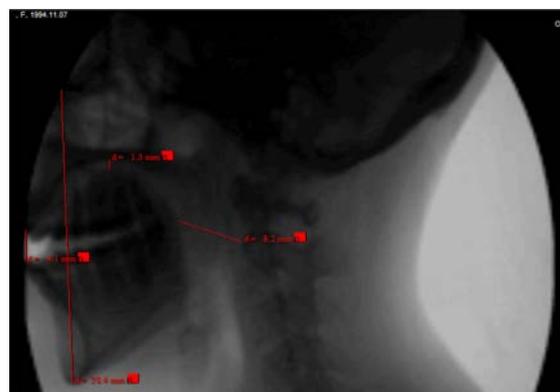
Emissão de [a] (sApo) – voz cantada sem instrução baseada em metáfora



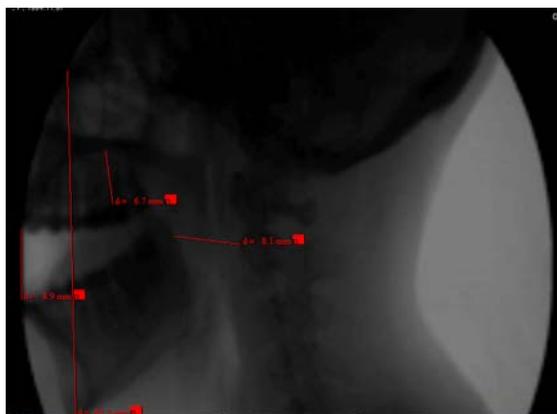
Emissão de [u] (jurUru) – voz cantada sem instrução baseada em metáfora



Emissão de [i] (rlo) – voz cantada sem instrução baseada em metáfora



Emissão de [i] (frlo) – voz cantada sem instrução baseada em metáfora



Emissão de [ɛ] (é que) – voz cantada sem instrução baseada em metáfora

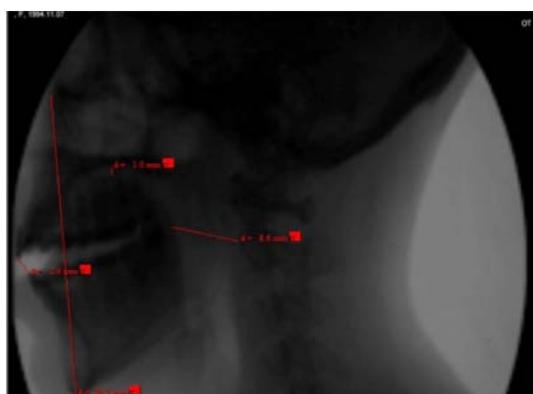
**Figura 8** – Dados de imagens de trato vocal ilustrativos das medidas (em mm) extraídas das emissões em voz cantada produzidas sem intruções baseadas em metáforas



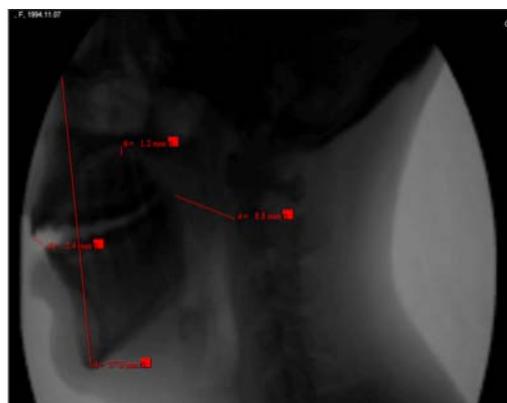
Emissão de [a] (sApo) – voz cantada produzida por meio de instrução baseada na metáfora prisma



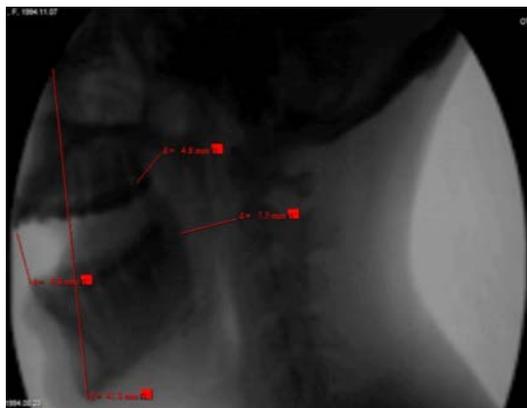
Emissão de [u] (jurUru) – voz cantada produzida por meio de instrução baseada na metáfora prisma



Emissão de [i] (rlo) – voz cantada produzida por meio de instrução baseada na metáfora prisma



Emissão de [i] (frlo) – voz cantada produzida por meio de instrução baseada na metáfora prisma



Emissão de [ɛ] (é que) – voz cantada produzida por meio de instrução baseada na metáfora prisma

**Figura 9** – Dados de imagens de trato vocal ilustrativos das medidas (em mm) de imagens extraídas das emissões em voz cantada produzidas por meio de instruções baseadas na metáfora prisma.



Emissão de [a] (sApo) – voz cantada produzida por meio de instrução baseada na metáfora catedral



Emissão de [i] (rlo) – voz cantada produzida por meio de instrução baseada na metáfora catedral



Emissão de [j] (frlo) – voz cantada produzida por meio de instrução baseada na metáfora catedral

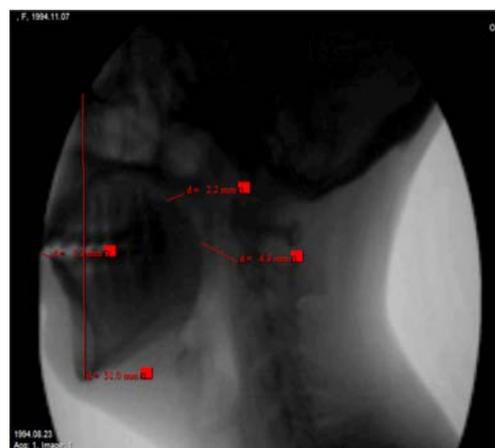


Emissão de [ɛ] (é que) – voz cantada produzida por meio de instrução baseada na metáfora catedral

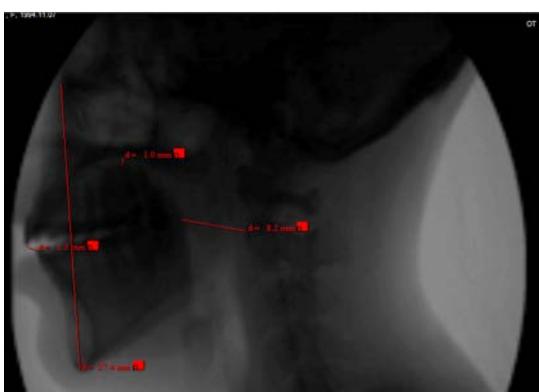
**Figura 10** – Dados de imagens de trato vocal ilustrativos das medidas (em mm) de imagens extraídas das emissões em voz cantada produzidas por meio de instruções baseadas na metáfora catedral.



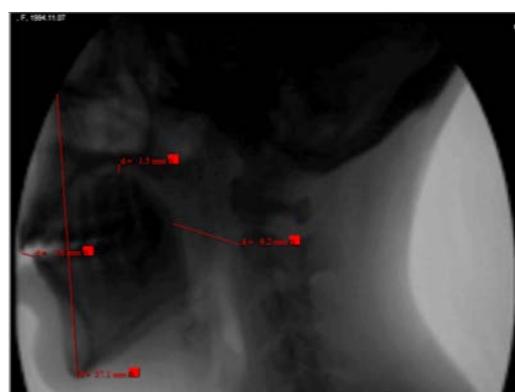
Emissão de [a] (sApo) – voz cantada produzida por meio de instrução baseada na metáfora disco voador



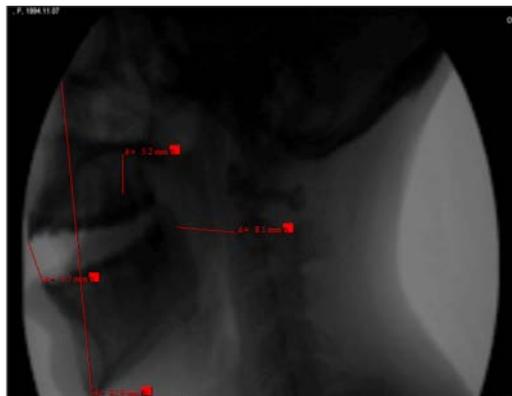
Emissão de [u] (jurUru) – voz cantada produzida por meio de instrução baseada na metáfora disco voador



Emissão de [i] (rlo) – voz cantada produzida por meio de instrução baseada na metáfora disco voador



Emissão de [i] (frlo) – voz cantada produzida por meio de instrução baseada na metáfora disco voador



Emissão de [ɛ] (é que) – voz cantada produzida por meio de instrução baseada na metáfora disco voador

**Figura 11** – Dados de imagens de trato vocal ilustrativos das medidas (em mm) de imagens extraídas das emissões em voz cantada produzidas por meio de instruções baseadas na metáfora disco voador.

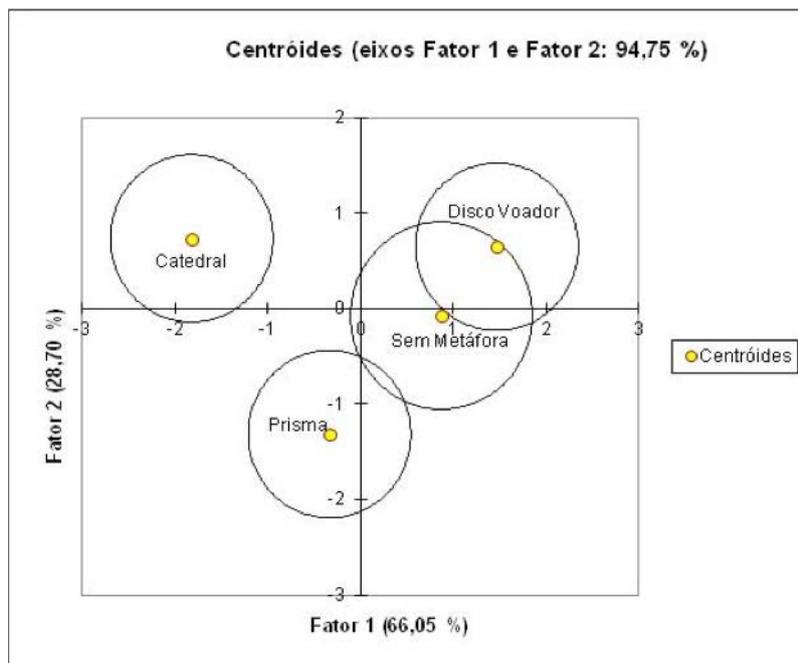
Para todas as emissões, as vogais [ɛ] e [i] foram identificadas como qualificador “anterior” em termos de deslocamento ântero-posterior de língua. Para todas as emissões, as vogais [a] e [u] foram identificadas como qualificador “posterior” em termos de deslocamento ântero-posterior de língua.

No caso da análise discriminante aplicada às medidas de imagem do trato vocal para detecção de metáforas, o melhor índice (60%) foi alcançado para a emissão produzida por meio de instrução baseada na metáfora catedral.

**Tabela 1** - Matriz de confusão para resultados de validação cruzada da análise discriminante para metáforas a partir das medidas de imagem do trato vocal.

de\a	Catedral	Disco Voador	Prisma	Metáfora	Total	% correto
Catedral	3	1	1	0	5	60,00%
Disco voador	0	1	2	2	5	20,00%
Prisma	2	0	1	2	5	0,00%
Sem Metáfora	1	2	1	0	4	0,00%
Total	6	4	5	4	19	26,32%

Para estimação das vogais a partir das medidas de imagem do trato vocal, os dados são apresentados no gráfico da distribuição de centróides.



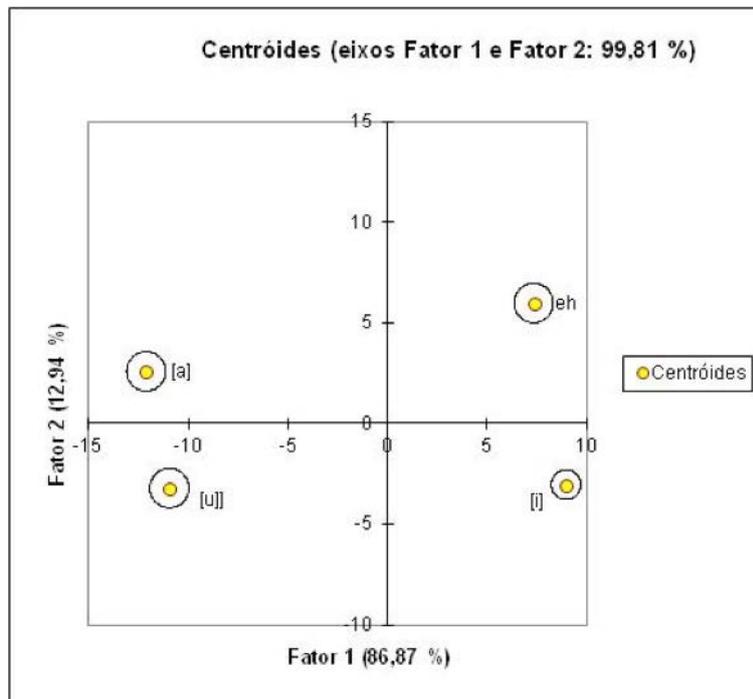
**Figura 12** – Gráfico de centróides da análise discriminante para estimação das metáforas a partir das medidas de imagem do trato vocal

Os fatores que influenciaram a segregação referiram-se às medidas de abertura de lábios normalizada (51,1%), abertura de lábios absoluta (42,9%) e de distância língua-palato (53,3%).

No caso da análise discriminante aplicada às medidas de imagem do trato vocal para detecção de vogais, o poder segregatório foi total (Tabela 2 e Figura 13).

**Tabela 2** - Matriz de confusão para resultados de validação cruzada da análise discriminante para vogais a partir das medidas de imagem do trato vocal.

de\a	[a]	[i]	[u]	[ɛ]	Total	% correto
[a]	4	0	0	0	4	100,00%
[i]	0	7	0	0	7	100,00%
[u]	0	0	4	0	4	100,00%
[ɛ]	0	0	0	4	4	100,00%
Total	4	7	4	4	19	100,00%



**Figura 13** – Gráfico de centróides da análise discriminante para estimação das vogais a partir das medidas de imagem do trato vocal

Os fatores que influenciaram a segregação das vogais referiram-se à qualificação de deslocamento ântero-posterior (99,3%), medidas de distância língua-faringe (98,3%), de distância língua-palato (97%), abertura de lábios absoluta (87,5%) e abertura de mandíbula (81%).

#### 4. Discussão

A descrição preliminar dos achados articulatórios gerados por registros de emissões em voz cantada produzidos por meio de instruções baseadas em metáforas, a partir da técnica da videofluoroscopia revelou que, num primeiro momento, maior tendência à abertura mandibular na metáfora catedral, especialmente para as vogais aberta [a] e semiaberta [ɛ] estudadas.

Do ponto de vista das medidas de mobilizações de articuladores extraídas das imagens videofluoroscópicas, algumas tendências foram reveladas para cada emissão produzida por meio de instruções baseadas em metáforas: catedral (maior grau de abertura de lábios e mandíbula, além de expansão faríngea), prisma (maior grau de abertura de lábios, sem correspondente em termos da abertura de mandíbula, língua elevada e menor amplitude de faringe) e disco voador (menor grau de abertura de lábios e mandíbula, língua elevada e expansão faríngea).

Dessa exposição, depreende-se que as instruções baseadas na metáfora catedral promoveram, no sujeito estudado, maiores efeitos sobre os ajustes de mandíbula e de faringe. As instruções baseadas na metáfora prisma promoveram ajustes de lábios, enquanto as instruções baseadas na metáfora disco voador centraram-se no estímulo à implementação de ajustes de língua e de mandíbula.

Destas observações, destacamos que o estudo de imagens do trato vocal permitiu revelar refinamentos e detalhes dos efeitos das mobilizações do trato vocal em resposta ao uso de instruções baseadas em metáforas.

Dados de mobilização laríngea não foram mensurados tanto pela ausência de referências de uso do *software Osiris* para tal medida, quanto pela dificuldade para delimitar a borda inferior da cartilagem cricóide (Williams, Eccles, 1990), especialmente diante da área delimitada nas imagens videofluoroscópicas coletadas.

De acordo com Sousa et al (2010), após análise de questionários aplicados a professores de canto, nota-se que há correspondências de ações musculares vinculadas às metáforas. No caso das metáforas “voz como uma lâmpada” e “bolha de sabão” estimam-se atividade fonatória e pressão subglótica equilibradas, além disso, sugerem maior atenção e percepção auditiva. Na metáfora “cúpula de uma igreja dentro da boca” há correspondências com a elevação da região do palato. Em comparação aos dados dessa pesquisa, a metáfora catedral apresentou a similaridade em termos de atividade muscular, uma vez que há efeitos de ajustes de mandíbula e de dimensão faríngea em sua produção.

Há indícios de que as instruções baseadas em metáforas tenham levado o sujeito estudado a promover ajustes diferenciados de trato vocal em função da exposição ao estímulo de instruções baseadas em metáforas para a produção de emissões em voz cantada (Clements, 2008; Sousa et al, 2010; Rossbach, 2011).

Sundberg (1970) conduziu um estudo comparativo entre fala e canto, com experimento baseado em coleta de dados radiográficos e acústicos. Na descrição de esquemas de imagens referentes ao canto, com a ressalva de que estudou apenas informantes do gênero masculino, a descrição assemelhou-se a dados de algumas das metáforas estudadas, especialmente catedral e disco voador, quanto à ampliação de faringe e ao abaixamento de laringe. Tais indicações reforçam que as metáforas colaboraram para implementação de ajustes de qualidade vocal característicos de emissões em voz cantada e essas representações simbólicas são capazes de gerar diferentes sensações proprioceptivas, motoras e táteis (Fonagy, 2000).

Os dados de outros sujeitos poderiam agregar maiores informações sobre a questão do uso das metáforas como elementos de instrução para que o cantor (e o aprendiz de canto) possa atingir ajustes no trato vocal que contemplem os efeitos estéticos desejáveis (Clements, 2008; Sousa Et Al, 2010; Rossbach, 2011).

Do ponto de vista dos dados de análise de imagem do trato vocal, a análise discriminante revelou poder segregatório parcial para estimação das metáforas estudadas (máximo de 60% para catedral), com influência das medidas de abertura de lábios (absoluta e normalizada) e distância língua-palato. Por outro lado, a análise discriminante de medidas de imagem de trato vocal para segregação das vogais, revelou a taxa de segregação total (100%), com maior influência do elemento qualificador de deslocamento anteroposterior de língua, medidas de distância língua-faringe, língua-palato, abertura de lábios e de mandíbula.

Além disso, a participante do estudo de caso, não era, de fato, cantora e, portanto, os ajustes implementados podem ser mais próximos àqueles usados na fala. Tais dados podem representar a demanda com que o professor de canto poderá se deparar ao iniciar o trabalho de instrução por meio de metáforas.

Medeiros (2002) aponta que as vogais produzidas durante o canto revelam ajustes de abaixamento da mandíbula (Sundberg, 1970; Sundberg, 1977) e de menor avultamento da língua (com exceção da vogal [a]) para manter afinação de  $f_0$ , portanto, compatíveis com achados do presente estudo, especialmente para a metáfora catedral, a qual revelou o maior grau de abertura mandibular, especialmente nas vogais [a] e [ɛ].

Sundberg (1970) também relatou efeitos articulatórios diferenciados no canto em função das vogais. No caso das vogais anteriores, maiores diferenças seriam encontradas na dimensão da cavidade faríngea e na abertura labial, efeitos estes que poderiam estar relacionados aos ajustes de abaixamento laríngeo e, portanto, a associação de tais mobilizações teria uma base fisiológica, uma vez que ajuste de laringe elevada pode levar a efeitos negativos na dinâmica de pregas vocais. Além disso, reforçou o cuidado em se considerar a dependência dos achados à classificação vocal, *pitch* habitual no canto e o formato do trato vocal do cantor estudado.

A esta altura da discussão, cabe, finalmente, pontuar algumas limitações experimentais em

função dos procedimentos de coleta de dados, tais como: o tempo de exposição à radiação, limitando o tempo de coleta (não foram registradas repetições de emissões, o que poderia permitir estudar uma mesma vogal em diferentes repetições); a representatividade das vogais orais do português brasileiro (não foi possível o registro de todas as vogais orais e, ainda, contamos com encontro vocálico, como nos casos de rio e frio) e, finalmente, não houve oportunidade de registro de outro sujeito de grupo de cantores.

Podemos destacar que os achados desse experimento sinalizam que algumas das metáforas estudadas propiciaram mobilizações diferenciadas do trato vocal, especialmente *catedral* e *disco voador*, que foram melhor detalhadas por meio de descrições fisiológicas (imagens videofluoroscópicas do trato vocal) das emissões de um sujeito não cantor.

Tais achados remetem a Fonagy (1976), que relatou associação de padrões de movimentação de língua e de abertura de boca a diferentes estados emocionais e atitudes na fala. Pautou tais descrições em investigações fisiológicas dos planos articulatório e fonatório.

Além disso, pensando em efeitos figurativos, é possível trazer à tona, as figuras utilizadas para a composição dessa pesquisa (*prisma*, *catedral* e *disco voador*). Todas elas apresentam uma imagem/significado pré-estabelecido, todavia, durante a produção de voz cantada as figuras apresentam um papel corporal, ou seja, propiciam diferentes ajustes de trato vocal. Portanto, criou-se uma nova representação simbólica que impactou a produção vocal do sujeito informante, visto que encontramos diferentes ajustes de lábios, mandíbula, postura de língua e constrição/expansão faríngea para cada tipo de imagem visualizada (diferentes sensações corporais, sinestésicas e metalinguísticas) (Hinton et al, 1994; Fonagy, 2000).

Os dados aqui relatados reforçam que a participante mobilizou o trato vocal para diferenciar ações que revelam o elo indissociável entre som e sentido e reforçam as questões de base dos estudos do simbolismo sonoro (Hinton et al, 1994), na busca por produzir emissões em voz cantada a partir de diferentes instruções.

O conceito da metáfora sonora (Fonagy, 1983) remete justamente às relações entre som e sentido. Tais relações pautam-se em cinco categorias de simbolismo; corporal, imitativo; sinestésico, metalinguístico e convencional (Hinton et al, 1994), sendo que para o uso das instruções para o canto, destacou-se o simbolismo sinestésico, quando consideramos o detalhamento de cada metáfora acima esboçado, em que as propriedades sonoras são empregadas para representar propriedades visuais, táteis e proprioceptivas de objetos.

As metáforas podem ajudar na implementação de ajustes do trato vocal do cantor, e do estudante de canto, levando-se em conta vários fatores que envolvem e permeiam a pedagogia vocal nesta situação particular e fascinante do canto.

## **5. Conclusão**

A análise dos dados revelou que as metáforas geraram mobilizações diferenciadas no trato vocal, especialmente para *“catedral”* e *“disco voador”*, reveladas por meio de descrições articulares- imagens videofluoroscópicas (medidas de abertura de lábios, abertura de mandíbula, distância língua-faringe, distância língua-palato e elemento qualificador de deslocamento anteroposterior de língua).

Os achados referentes a análise das imagens do trato vocal foram relevantes à investigação da produção de voz cantada e reforçaram a importância do enfoque das relações entre som e sentido, especialmente mediadas pelas questões do simbolismo sonoro sinestésico.

## Referências Bibliográficas

- AGUIAR, M.C.P. **O Ensino de Canto em Portugal: uma perspectiva analítico-reflexiva a partir de meados do Século XX**. Dissertação de Mestrado. Aveiro: Universidade de Aveiro, 2007.
- ANDRADA e SILVA, M. A.; DUPRAT, A. C. Voz Cantada. In: FERREIRA, L. P.; BEFI-LOPES, D.M.; LIMONGI, S.C. **Tratado de Fonoaudiologia**. São Paulo: Roca, 1 ed, 2004.
- BILTON, T.L; LEDERNMAN, H.M. Videodeglutuesofagograma: Principais aspectos por imagem. **Rev Imagem**, v.20, n. 1, p. 27-28, 1998.
- CLEMENTS, J.F. **The use of imagery in teaching voice to the twenty-first century student**. Dissertação de Mestrado. Florida: 2008.
- COSTA, M.M.B. et al. Valores típicos do “produto dose-área” (DAP) obtido durante o estudo videofluoroscópico da deglutição. **Revista Radiologia Brasileira**, v.36, n.1, p.17-20, 2003.
- DRAHAN, S.H. et al. Dimensão do trato vocal de cantoras profissionais e não profissionais durante o repouso e o canto. **Arquivos internacionais de otorrinolaringologia, 11º Congresso da Fundação Otorrinolaringologia**, v.16, n.1, 2012.
- ECHTERNACH, M. et al. Vocal Tract in Female Registers- A Dynamic Real-Time MRI Study. **J of Voice**, v.24, n.2, p.133-139, 2008.
- ERICSDOTTER, C. et al. The apex model - From articulatory position to sound. **Acoustical Society of America**, v.104, n.3, p.1820-1820, 1998.
- FIRMAN, C.M.G. et al. Avaliação quantitativa e qualitativa dos esvaziamento gástrico através do método videofluoroscópico. **Rev Arquivos de Gastroenterologia**, São Paulo: v.37, n.2, 2000.
- FONAGY, I. **La mimique buccale**. **Phonetica**, v.33, n,1, p.31-44, 1976.
- \_\_\_\_\_. **La Vive Voix**. Paris: Payot, 1983.
- \_\_\_\_\_. **Lang.s within Lang.s: an evolutive approach**. Amsterdam: John Benjamins Pub Co, 2000.
- GONÇALVES, C.G.A.B. **Análise perceptivo-auditiva da fala e avaliação videofluoroscópica da posição da língua em fissurados de palato**. Dissertação de Mestrado São Paulo: USP, 2002.
- HINTON, L.; NICHOLS, J.; OHALA, J. J. **Sound Symbolism**. Cambridge: Cambridge U P, 1994.
- KENT, R.D. **Speech Scis**. Rio de Janeiro: Revinter, 1993.
- LAVER, J. **The phonetic description of voice quality**. Cambridge textbook in Ling, Cambridge: Cambridge U P, 1980.
- LOPASSO, C. A videofluoroscopia e quantificação virtual do esvaziamento gástrico. **Rev Arquivos de Gastroenterologia**, v.37, n.2, 2000.
- MADUREIRA, S. Sobre a expressividade da fala. In: KYRILLOS, L. **Expressividade: da teoria à prática**., Rio de Janeiro: Revinter, 2005.
- MEDEIROS, B.R. **Descrição comparativa de aspectos fonético-acústicos selecionados da fala e do canto em português brasileiro**. Tese de Doutorado. Campinas: UNICAMP, 2002.
- NUÑES, K.L. **Configuração do trato vocal na fonação habitual e no canto popular**. Dissertação de Mestrado. Curitiba: Universidade Tuiuti do Paraná, 2006.
- PEDROSO, M.I.L. **Técnicas vocais para os profissionais da voz**. Monografia de Especialização. São Paulo: CEFAC, 1997.
- ROSSBACH, R.F. Aplicando metáforas físicas no repertório coral. **Rev de Letras, Artes e Comunicação**, v.5, n.2, p.153-166, 2011.
- SALOMÃO, G.L. **Registros vocais no canto: aspectos perceptivos, acústicos, aerodinâmicos e fisiológicos da voz modal e da voz de falsete**. Tese de Doutorado. São Paulo: PUCSP, 2008.
- SOUSA, J.M.; ANDRADA E SILVA, M.A.; FERREIRA, L.P. O uso de metáforas como recurso didático no ensino do canto: diferentes abordagens. **Rev Soc Brasileira de Fonoaudiologia**, São Paulo: v.15, n.3, p.317-328, 2010.

- SOUZA, D.P.D. **Estudo da fonação e de respiração, com apoio abdominal expandido e contraído, em cantores líricos**. Dissertação de Mestrado. Curitiba: Universidade Tuiti do Paraná, 2008.
- SUNDBERG, J. Formant structure and articulation of spoken and sung vowels. **Folia Phoniatic**, v.22, p.28-48, 1970.
- \_\_\_\_\_. **The acoustic of singing voice**, p.104-116, 1977.
- \_\_\_\_\_. What's so special about singers? **J of Voice**, , v.4, n.2, p.107-119,1990.
- \_\_\_\_\_. Human singing voice. In: CROCKER, M.J. **Encyclopedia of acoustics**, John Wiley & Sons, Nova Jersey, p.1687-1695, 1997.
- \_\_\_\_\_. Articulatory configuration and *pitch* in a classically trained soprano singer. **J of Voice**, v.23, n.5. p.546-551, 2008.
- WILIAMS, R.G.; ECCLES, R. A new clinical measure of external laryngeal size which predicts the fundamental frequency of the larynx. **Acta Otolaryngol**. Stockholm: v.110, p.141-148, 1990.

*Metáforas no Ensino de Canto e seus  
Efeitos na Qualidade Vocal: um Estudo  
Acústico-Perceptivo*

PAULO MENEGON  
SANDRA MADUREIRA

## 1. Introdução

O avanço nas ciências da fala e nas tecnologias de fala a partir de meados do século XX redundou no aprofundamento do conhecimento sobre a produção e a percepção da fala e do canto.

Salomão (2008) menciona que uma das primeiras preocupações nas ciências da voz cantada é com a questão da produção e percepção dos registros vocais, sendo um marco, o primeiro estudo de um professor de canto Manuel Garcia II (1805 – 1906) que conseguiu, como Salomão (2008:4) afirma, “observar as pregas vocais na situação do canto, pela utilização de um laringoscópio bastante rudimentar”.

Discípula de Garcia, Marchesi, influenciada por Garcia, defende o estudo científico da voz, argumentando que “após a experiência bem sucedida de muitos anos, estou convencida de que o conhecimento científico é indispensável para o professor de canto, porque lhe permite tratar o instrumento vocal de uma forma natural e racional e com maior certeza, corrigir as inúmeras falhas de emissão que cada aluno traz, resultante de maus hábitos” (Marchesi, 1970: xiv).

Na atualidade, entre as ciências da fala, a fonética, especialmente a fonética acústica, por enfocar o meio que integra a produção e a percepção da fala, possibilita, de maneira não invasiva, inferências sobre o posicionamento dos articuladores e a verificação de suas consequências acústicas e perceptivas.

Sob a perspectiva fonética acústica, de acordo com o modelo da teoria acústica de produção da fala (Fant, 1970), também conhecido como modelo da fonte e do filtro, a fala é entendida como o resultado do efeito de ressonância no trato vocal que incide sobre as fontes sonoras (de voz e de ruído) que se formam no aparelho fonador. As ressonâncias do trato vocal são os formantes.

Sundberg (1974), ao comparar as produções de vogais em cantores e não cantores, verificou na produção do canto uma ênfase espectral, uma região de frequências reforçadas, à qual denominou o formante do cantor. A existência desse formante em torno de 3 khz no canto lírico masculino faz com que a voz do cantor não seja sobrepujada pelo som da orquestra conforme relata Sundberg (1987). Essa ênfase espectral está relacionada, entre outros fatores, com o abaixamento da laringe e com o alargamento da cavidade faríngea (Sundberg e Gauffin, 1979).

É comum no ensino de estratégias de canto, especialmente no canto lírico, no qual a voz do cantor tem de se projetar em grandes salas de espetáculo, a utilização de metáforas, entre elas a do prisma, a do disco voador e a da catedral. A motivação para a utilização da metáfora do prisma é fazer com que o aluno iniciante entenda, na prática, como projetar a voz, utilizando da ressonância de canto “frontalizada”, enquanto que a do disco voador é evitar que o aluno tensione as pregas vocais, ajudando-o a produzir uma voz melodiosa e a da catedral é fazer com que o aluno realize plenamente a ressonância bucal. A metáfora da catedral é amplamente utilizada no treinamento em canto lírico para ajudar o aluno a projetar a voz em ambientes como grandes teatros de ópera.

Neste trabalho temos como objetivo investigar, com base em análises perceptiva e fonético-acústica, os efeitos nos ajustes de qualidade de voz de emissões de canto motivadas pela apresentação dessas três metáforas, estratégia de treinamento comumente utilizada no contexto de ensino de canto, especialmente do canto lírico.

Sob a perspectiva da fonética perceptiva, a qualidade de voz é descrita por meio de ajustes fonatórios e articulatórios de longo termo que caracterizam a fala de um indivíduo (Laver, 1980; 2000; Laver e Mackenzie-Beck, 2007).

## 2. A Metáfora e sua Conceituação

A metáfora pode ser entendida como um recurso que o ser humano utiliza na linguagem de tal modo que, ao serem atribuídas qualidades figurativas a determinado elemento, este passa a

designar algo novo.

A alta relevância comunicativa da metáfora a torna um recurso valioso nos processos educacionais e na definição de objetos e objetivos pedagógicos. Segundo Berber Sardinha (2007: 13), “as metáforas são recursos retóricos poderosos e são conscientemente usadas por políticos, advogados, jornalistas, escritores e poetas, entre outros, para dar mais ‘cor’ e ‘força’ a sua fala e a sua escrita.

Quanto as principais abordagens relacionadas à metáfora, consideramos alguns tipos, dando ênfase ao tipo de metáfora de maior relevância para este trabalho: a metáfora sonora (Fonagy, 1983). As metáforas sonoras revestem-se de dinamicidade e são expressas por meio de gestos vocais e visuais.

Por abordarem aspectos cognitivos, linguístico-pragmáticos e tangerem questões de melodicidade, outros tipos de metáforas que se mostram relevantes para a nossa investigação, são:

- a metáfora conceptual, concebida como um fenômeno cognitivo, é assim designada porque conceitua alguma coisa. Um exemplo disso, encontra-se na frase “Esporte é saúde”. A concepção cognitiva de metáfora é defendida por Lakoff e Johnson (1980). Para esses autores, “Ela também é corporificada, já que nessa visão o corpo humano é a base ou fonte de muitas metáforas cognitivas e, portanto aparece representada no corpo do falante também”.

- a metáfora sistemática foi desenvolvida a partir do trabalho seminal de Cameron (2003) e utiliza-se de pressupostos teóricos próprios e de conceitos influenciados pelas obras de Bakhtin, Vygotsky, Firth e Sinclair, conforme aponta Berber Sardinha (2007). Entre os conceitos estão a noção de alteridade (Bakhtin-1992), a qual implica colocar-se na posição do outro para compreensão de alguém, a noção de pensamento segundo Vygotsky (1934) como ação internalizada que parte de uma experiência individual e a noção de que a linguagem deve ser estudada em *corpora* autênticos, conforme indica Sinclair (1991).

- a metáfora gramatical que compreende um recurso linguístico que influencia determinada categoria a exercer uma função que não lhe é inerente. Como exemplo, podemos citar um verbo que é utilizado como substantivo: “Ele não veio para almoçar”. No momento em que o verbo “almoçar” corresponde ao substantivo “almoço”, cria-se uma ação metafórica que produz desdobramentos no sistema linguístico, pois uma categoria ocupa a posição de outra na expressão do sentido pretendido.

- a metáfora como forma de instrução no contexto musical, a qual evidencia uma maneira de apreender a teoria e facilitar a prática do ensino e, dessa forma, configura-se com um instrumento para a compreensão da realidade.

Alguns termos comumente empregados com sentido metafórico na música, os quais expressam qualificações como *tenso*, *alegre*, *azul*, *claro*, *doce* e *ríspido* são frequentemente utilizados por profissionais da música de forma geral para indicar o andamento e a harmonia de uma obra musical. Um qualificador muito usado pelos músicos para descrever o andamento de uma obra musical desde a época do classicismo musical é *allegro*, que inserido na frase “Allegro ma non troppo” (Alegre, mas não tanto), refere-se ao andamento que se pretende de um movimento em uma obra musical, remetendo a palavra “alegre” à velocidade rápida.

Para Zbikowski (2008: 503), a metáfora constitui “uma parte inevitável das descrições musicais que aspiram mais do que um ensaio de termos definidos”. Para esse autor, a metáfora está diretamente ligada à atividade musical em todas as suas formas. Nos últimos anos, os estudos sobre o uso da metáfora no canto têm se intensificado em diversas subáreas da comunicação.

O professor de canto busca, pelo emprego de metáforas, criar estratégias que possam eliciar ajustes dos órgãos envolvidos na produção dos sons da fala, sejam esses produzidos na região

glótica (ajustes fonatórios) ou na região supraglótica (ajustes articulatórios) ao fazer uso de metáforas sonoras para eliciar esses ajustes tem como objetivo estimular o aluno a vivenciar diferenciados mecanismos que resultem em qualidades vocais específicas.

A metáfora também é utilizada no ensino de instrumentos com propósitos didáticos para que os alunos possam entender e executar bem os instrumentos (violino, piano, flauta, fagote e trompete), mas é professor de canto que mais se utiliza dessa estratégia para explicar como o aparelho fonador molda-se para produzir os sons.

Sousa, Silva e Ferreira (2010) investigaram o uso de metáforas no ensino de canto e verificaram que 90% dos professores entrevistados consideravam o uso de metáforas no ensino de canto uma ferramenta didática eficaz por estimular a propriocepção e a musicalidade, além de entender que a instrução por meio de linguagem fisiológica tornava-se de difícil aplicação devido a sua complexidade.

As questões de som e sentido estão diretamente implicadas no conceito de metáfora sonora e de simbolismo sonoro. Trabalhando com as noções da naturalidade e da motivação, ou seja, com as relações diretas entre som e sentido, abre-se espaço para considerar como a matéria fônica tem potencialidade para expressar sentidos. Jakobson (1977) resgata nos estudos linguísticos pós-saussureanos, a discussão sobre as relações entre som e sentido, a qual, pelo que se tem registrado, remonta a quatro séculos antes de Cristo. Em Madureira (2002), encontramos uma resenha sobre a evolução dessa discussão desde o Crátilo, de Platão, até os estudos de natureza psicolinguística (Albano, 1988).

Com os avanços no desenvolvimento das ciências e das tecnologias da fala, aprofundou-se o conhecimento sobre a produção dos sons, seu produto acústico e sua percepção. Essa evolução teve impacto nos estudos sobre o simbolismo sonoro. Nesse contexto, a obra de Fónagy (1983) opõe “a viva voz” “à letra morta” e introduz o conceito de metáfora sonora, o qual se apoia no código natural, pré-linguístico e de caráter “motivado”, relacionando som e sentido, vínculo que emerge de forma consciente ou inconsciente e se incorpora pela cultura. O produto acústico de configurações fonatórias e articulatórias específicas, se consideramos o código de caráter motivado, carregam sentido, causando impressões no ouvinte.

Os efeitos de sentido veiculados pela metáfora sonora levam o ouvinte a perceber emoções como desgosto, aflição, alegria, angústia, raiva ou atitudes como prepotência, amizade e gentileza. Quando ficamos com raiva, por exemplo, nossos músculos se tensionam e, ao falarmos, essa tensão muscular influencia o resultado acústico, que passa a representar essa tensão. Desse modo, estabelecem-se entre configurações do trato vocal e dimensões de sentido, tais como tamanho, luminosidade, brilho e maciez.

A exploração das relações entre som e sentido concerne ao nível paralinguístico e é nesse nível que se insere a discussão sobre a metáfora sonora que envolve representações de estados afetivos, de gestos e de efeitos de sentido por meio de características acústicas com as quais esses estados, gestos ou efeitos são associados. No nível paralinguístico considera-se o valor indexical das características fonéticas de segmentos da fala e dos elementos prosódicos (acentuação, entoação, qualidade de voz, pausa, ritmo, taxa de elocução).

Para Fónagy (1983), a entoação é a “imagem da alma”, dado o valor expressivo das modulações melódicas. Esse efeito causado pela entoação é utilizado pelos cantores que, pelas modulações, criam tensões e as resolvem, revelando emoções em movimentos, os movimentos melódicos. Para desenvolvermos estudos relacionados à metáfora sonora e ao simbolismo sonoro, é preciso abordar aspectos que concernem à expressividade da fala e que remetem à potencialidade que a matéria fônica tem de veicular efeitos de sentido (Madureira, 2011). Segundo a autora, os gestos vocais e visuais são elementos constitutivos da expressividade na fala, entendida como informação na comunicação (Beller, 2009).

Alves (2013) explorou a gestualidade no canto, a partir de descrições acústicas e articulatórias de ajustes de qualidade de voz. A análise fonética dos dados revelou que as metáforas motivaram mobilizações diferenciadas no trato vocal, reveladas por descrições articulatórias, baseadas em medidas de imagens videofluoroscópicas e em análise acústica.

A fim de investigarmos as configurações articulatórias influenciadas pela sugestão de metáforas conceituais, recorreremos aqui, aos procedimentos, pressupostos teóricos e instrumentais de análise fonética e ao apresentarmos as metáforas para eliciar modificações nas configurações articulatórias, percorremos, segundo (Madureira, 2011), o trajeto inverso da constituição das metáforas sonoras na fala, pois enquanto a metáfora sonora passa a representar por analogia uma certa configuração articulatória que a produziu, as metáforas instrutivas que utilizamos são as propulsores das configurações articulatórias que se moldam para contemplar a associação elicitada pela metáfora.

### **3. A Qualidade de Voz**

Adotamos aqui, o conceito de qualidade de voz de acordo com Laver (1979, 1980, 2000). Para a descrição dos ajustes de qualidade de voz, Laver (1980) propõe um modelo fonético descritivo que tem como unidade básica o ajuste neutro. Esse ajuste neutro tem como base a ideia de que o trato vocal não se encontra modificado pela ação da musculatura, parte de um ponto de repouso.

O ajuste neutro compreende: lábios não projetados; laringe nem abaixada, nem elevada; diâmetro do trato supralaríngeo constante; trato vocal sem estreitamentos; mandíbula, nem acen-tuadamente aberta, nem fechada; base da língua nem avançada, nem recuada; articulações orais anteriores realizadas pela lâmina da língua; nasalidade presente apenas nos segmentos caracte-rísticos; uso da corrente de ar: eficiente, sem ruídos; vibração de pregas vocais: periódica, com tensão longitudinal e adutora moderada, compressão medial também moderada; tensão muscular no aparelho fonador: nem alta, nem baixa.

São considerados três tipos de ajustes no modelo fonético descritivo de qualidade de voz. supralaríngeos, laríngeos e de tensão. Os ajustes supralaríngeos englobam mobilizações dos ór-gãos do trato vocal, lábios, língua, mandíbula, véu palatino e laringe, modificando as configurações das cavidades oral, nasofaríngea, orofaríngea e laringofaríngea. Os ajustes de tensão concernem ao trato vocal e o laríngeo. Os ajustes fonatórios compreendem ajustes de extensão, aproximação e contato realizados pelas pregas vocais e com variados modos de fonação.

No módulo fonético descritivo de qualidade de voz proposto por Laver (1980), dois princí-pios são postulados: o da susceptibilidade e o da compatibilidade. Esses princípios são relevantes a serem considerados na identificação dos ajustes. Pelo princípio da susceptibilidade, temos que alguns segmentos fônicos são mais susceptíveis do que outros a determinados ajustes. Dessa ma-neira, os segmentos fônicos orais são mais susceptíveis ao ajuste de nasalidade do que os nasais, ou seja, a característica do ajuste da nasalidade será mais facilmente percebida se os segmentos fônicos forem orais. Essa relação entre ajustes e segmentos deve ser observada para fins de identi-ficação dos ajustes, pois quando os ajustes têm atributos não comuns ao segmento, esses passam a ser mais susceptíveis à sua influência.

Baseado nas concepções teóricas do modelo descritivo de qualidade de voz, Laver (1979, 1980, 2000) elaborou um roteiro denominado *Voice Profile Analysis Schemme* (VPAS), que permite avaliar os ajustes de qualidade de voz empregados pelos falantes. Ao longo dos anos, esse roteiro sofreu modificações efetuadas por Laver e colegas (2007).

O novo *Vocal Profile Analysis Scheme* (Laver, 2007) foi traduzido e adaptado para o portu-guês brasileiro por Camargo e Madureira (2008). O resultado é o que vemos a seguir na Figura 1.

QUALIDADE VOCAL	PRIMEIRA PASSADA		PRIMEIRA PASSADA							
	Neutro	Não neutro	AJUSTE	Moderado			Extremo			
				1	2	3	4	5	6	
<b>A. ELEMENTOS DO TRATO VOCAL</b>										
1. Lábios			Arredondados/ Protraídos							
			Estirados							
			Labiodentaliza- ção							
			Extensão diminuída							
			Extensão aumentada							
2. Mandíbula			Fechada							
			Aberta							
			Protraída							
			Extensão diminuída							
			Extensão aumentada							
3. Língua ponta/lâmina			Avançada							
			Recuada							
4. Corpo de língua			Avançada							
			Recuada							
			Elevado							
			Abaixado							
			Extensão diminuída							
Extensão aumentada										
5. Faringe			Constricção							
			Expansão							
6. Velofaringe			Escape nasal audível							
			Nasal							
			Denasal							
7. Altura de laringe			Elevada							
			Abaixada							
<b>B. TENSÃO MUSCULAR GERAL</b>										
8. Tensão do trato vocal			Hiperfunção							
			Hipofunção							

9. Tensão laríngea			Hiperfunção						
			Hipofunção						
<b>C. ELEMENTOS FONATÓRIOS</b>									
	<b>AJUSTE</b>	<b>Presente</b>		<b>Graus de escala</b>					
		<b>Neutro</b>	<b>Não Neutro</b>	<b>Moderado</b>			<b>Extremo</b>		
				<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
10. Modo de Fonação	Modal								
	Falsete								
	Crepitância/ vocal fry								
	Voz crepitante								
11. Fricção laríngea	Escape de ar								
	Voz soprosa								
12. Irregularidade laríngea	Voz áspera								

Ocorrências em curto termo ( ) quebras ( ) instabilidades ( ) diplofonia ( ) tremor. Para ajustes de ocorrência intermitente assinalar (i)

DINÂMICA VOCAL	Neutro	AJUSTE	Moderado			Extremo			
			1	2	3	4	5	6	
<b>D. ELEMENTOS PROSÓDICOS</b>									
13. Pitch (f0)	Habitual		Elevado						
			Abaixado						
	Extensão		Diminuída						
			Aumentada						
Variabilidade		Diminuída							
		Aumentada							
14. Loudness (intensidade)	Habitual		Aumentado						
			Diminuído						
	Extensão		Diminuída						
			Aumentada						
	Variabilidade		Diminuída						
			Aumentada						
<b>15. TEMPO</b>									
Continuidade			Interrompida						
Taxa de elocução			Rápida						
			Lenta						

16. OUTROS ELEMENTOS							
Suporte respiratório	Adequado						
	Inadequado						
	Presente						

**Figura 1** – Roteiro Vocal Profile Analysis Scheme (VPAS) traduzido e adaptado para o português brasileiro por Camargo e Madureira (2008).

Essa versão adaptada do roteiro foi aplicada em treinamento de juízes, e os resultados dessa aplicação foram relatados em Camargo e Madureira (2008). A versão do roteiro de Camargo e Madureira (2008) também foi aplicada em vários trabalhos desenvolvidos no Laboratório Integrado de Análise Acústica e Cognição (LIAAC), da Pontifícia Universidade Católica de São Paulo, entre eles: Pessoa, 2012; Madureira, Camargo e Rusilo (2011); Camargo e Madureira (2010) e Madureira (2011).

O modelo fonético da qualidade de voz contempla aspectos articulatórios, fisiológicos acústicos e auditivos relativos aos ajustes de qualidade de voz e é orientado para a descrição de ajustes na fala, não sendo concebido para o canto. Para a análise de alguns ajustes de voz específicos do canto profissional faz-se necessária uma adaptação do roteiro. Em nosso trabalho, que aborda ajustes das configurações do trato vocal implementados por associações disparadas por metáforas, utilizamos o modelo sem propor alterações, observando os efeitos nos ajustes de qualidade de voz que acarretam modificações nos segmentos fônicos por meio do princípio de susceptibilidade. Cabe aqui indicar uma menção para uma proposta regulada para o canto profissional.

## 4. Método

A metodologia compreende informações sobre a elaboração do *corpus* de pesquisa, o local e a forma de realização da coleta; os equipamentos e *softwares* utilizados para gravação e análise dos dados, os sujeitos da pesquisa, os procedimentos e as técnicas de análise, bem como os testes estatísticos utilizados. A seguir, especificamos o *corpus* e os procedimentos metodológicos utilizados em nossa pesquisa.

### 4.1. Corpus

O *corpus* de pesquisa compreende uma canção popular intitulada *Sapo Jururu*, a qual transcrevemos a seguir:

*Sapo Jururu, da beira do rio, quando o sapo canta, ó maninha, é que está com frio. A mulher do sapo, é que está lá dentro, fazendo rendinhas, maninha, para o casamento.*

Nossa escolha por essa canção decorreu do fato de sua melodia ser amplamente conhecida. *Sapo Jururu* é uma canção infantil proveniente do folclore brasileiro, assim, é de domínio público do Brasil. O termo “jururu” está relacionado com o sentimento desse sapo em relação ao frio que o desanima.

Para fins de análise, dividimos a canção em quatro enunciados:

1. Sapo jururu da beira do rio
2. Quando o sapo canta maninha é que está com frio

3. A mulher do sapo deve estar lá dentro
4. fazendo rendinha maninha para o casamento

## 4.2. Sujeitos da pesquisa

Participaram da pesquisa cinco sujeitos do sexo feminino, paulistanos, sem experiência no canto e com idade entre 18 e 35 anos.

## 4.3. Gravação do *corpus*

A coleta de dados foi realizada no Estúdio de Rádio e TV da PUC-SP, que dispunha dos equipamentos e *softwares* necessários para a coleta de dados da pesquisa: mesa de som, microfone unidirecional ML 70-D, *software* Praat (<http://www.fon.hum.uva.nl/praat>) e *software* SoundForgeEdit (versão 7).

O microfone *headset* ao qual nos referimos anteriormente foi acoplado tanto ao equipamento utilizado para calibração (decibelímetro) quanto à mesa de som citada.

A gravação teve início com emissão de fala semiespontânea (nome, data e breve relato sobre a cidade em que nasceu), o que permitiu o monitoramento dos níveis de gravação e a adequação do instrumental. Em seguida, foi apresentado aos sujeitos participantes da pesquisa o registro em áudio da canção infantil [Sapo Jururu](#), gravada por uma professora de canto como referência. A canção foi reproduzida três vezes no sistema de autofalantes da sala de gravação. Também foi entregue a cada sujeito uma versão impressa da letra da canção.

A coleta de dados de emissões em voz cantada compreendeu quatro etapas, as quais relacionamos a seguir.

Na primeira etapa, solicitamos aos sujeitos que cantassem da forma que desejassem. Nenhuma outra orientação foi dada a eles.

Na segunda etapa, apresentamos a imagem que segue (Figura 2). Após olhar a imagem, os sujeitos foram orientados a cantar de acordo com a seguinte instrução:

“Imagine que sua voz se propaga no ar como uma luz branca incidindo sobre um prisma”.



**Figura 2** – Imagem relativa à primeira metáfora: “Imagine que sua voz se propaga no ar como uma luz branca incidindo sobre um prisma”.

Na terceira etapa, apresentamos a imagem a seguir (Figura 3). Após olhar a imagem, os sujeitos foram orientados a cantar conforme a seguinte instrução: “Cantar a melodia proposta, fluindo com a voz dentro da ‘boca’”. Ao mesmo tempo, o sujeito, via a figura de um *disco voador*.



**Figura 3** – Imagem relativa à segunda metáfora: “Cantar a melodia proposta, fluindo com a voz dentro da ‘boca’”.

Já na quarta etapa, apresentamos a imagem que segue (Figura 4). Após olhar a imagem, os sujeitos foram orientados a cantar de acordo com a seguinte instrução: “Cantar a melodia proposta preenchendo de som por toda a cavidade bucal, transformando a cavidade bucal numa catedral”. Ao mesmo tempo, o sujeito, via a figura de uma *cúpula de catedral*.



**Figura 4** – Imagem relativa à terceira metáfora: “Cantar a melodia proposta preenchendo de som por toda a cavidade bucal, transformando a cavidade bucal numa catedral”.

#### **4.4. Procedimentos de análise de dados**

As gravações foram analisadas do ponto de vista acústico e perceptivo-auditivo. Os resultados dessas análises foram correlacionados com o auxílio de testes estatísticos: análise discriminante e Test t para duas amostras independentes.

##### **4.4.1. Análise acústica**

Para a análise acústica, as gravações da canção *Sapo Jururu* foram analisadas com a aplicação do SG – ExpressionEvaluator desenvolvido por Barbosa (2009) no ambiente do *software* de livre acesso PRAAT, versão 5212, quanto à frequência, medida em Hz, à intensidade medida em dB e à banda dos formantes, medida em Hz. Os parâmetros extraídos automaticamente pelo *script* consideraram medidas de frequência fundamental – f0 (mediana, semi-amplitude entre quar-

tis, assimetria e quantil 99,5%) e 1ª derivada de f0 (média, desvio padrão e assimetria), intensidade (assimetria), declínio espectral (média, desvio padrão e assimetria) e espectro de longo termo-ELT (desvio padrão).

As medidas extraídas do *script* são elucidadas a seguir:

- mediana de f0 (frequência fundamental): medida que pode ser considerada como a f0 habitual do falante. Sua extração prevê a suavização de valores. Trata-se de uma medida que evita erros de detecção do valor de f0 que podem ser frequentes, sobretudo os que mudam o valor para uma oitava abaixo ou acima;

- semi-amplitude entre quartis de f0: medida de variação dos valores de f0, excluindo-se valores espúrios (por motivos de normalização, foi dividida pela metade, justificando a atribuição do termo “semi-amplitude”). O termo amplitude refere-se à variação da medida;

- quantil 99,5% de f0: valores de frequência fundamental normalizados por meio do cálculo de Z score da mediana (dobro), acrescido do valor de desvio padrão. Essa medida busca, sem valores espúrios, detectar o limite superior de f0 do falante;

- média de derivada de f0: representa a taxa de variação de uma função. Medida que considera não somente a variação, mas também a “velocidade” dessa alteração/variação;

- desvio padrão de derivada de f0: medida que infere a dispersão estatística conforme a variância de f0 contida na amostra;

- assimetria de derivada de f0: infere relações de como aconteceram as mudanças de taxa de variação de f0 na função;

- assimetria de f0: medida da assimetria de distribuição de f0, baseada na razão de diferença entre média e mediana/semi-amplitude entre quartis de f0. Essa medida busca estimar a tendência a ter valores à esquerda (negativos) ou à direita da média (positivos);

- assimetria de intensidade: medida de intensidade normalizada, baseada na proporção de intensidade no intervalo de frequências de 0,0-1250 Hz/1250-4000 Hz.

Permite que as amostras possam ter suas intensidades estimadas, mesmo aquelas que não tenham sido captadas com fixação da distância entre boca e microfone;

- média de declínio espectral: média de valores da proporção de intensidade nos intervalos de 0-1,0 kHz/1,0-4,0 kHz. Considerada como importante medida do nível de tensão laríngea do estímulo aferido.

- desvio padrão de declínio espectral: medida do desvio padrão de declínio espectral (anteriormente explanada);

- assimetria de declínio espectral: medida de assimetria de distribuição das medidas de declínio espectral (anteriormente explanada);

- desvio padrão de espectro de longo termo (ELT): desvio padrão das medidas normalizadas de intensidade ao longo de intervalos de frequências do espectro sonoro.

#### **4.4.2. Análise perceptiva**

Neste trabalho, utilizamos VPAS-PB, traduzido e adaptado de Laver e Mackenzie-Beck (2007) por Camargo e Madureira (2008). Esse roteiro sofreu algumas alterações desde a sua primeira versão elaborada por Laver (1980). Outras versões compreendem Laver (1994, 2000) e Mackenzie-Beck (2005).

A aplicação do roteiro exige juiz qualificado. Neste trabalho, a caracterização dos ajustes foi feita por uma foneticista com experiência na aplicação do roteiro.

O roteiro, em sua versão 2007, contém os ajustes de qualidade de voz passíveis de serem graduados em uma escala de seis graus e também os aspectos da dinâmica da voz.

#### 4.4.3. Análise integrada de dados perceptivos e acústicos

Para a correlação entre informações de âmbito perceptivo-auditivo e medidas acústicas, foram utilizados os procedimentos desenvolvidos no Grupo de Pesquisas sobre fala do LIAAC. A apresentação dos indicadores dos resultados a partir de correlatos perceptivos e acústicos e qualidade e dinâmica vocal são descritos a seguir:

Dados perceptivo-auditivos:

- dados de qualidade e dinâmica vocal - VPAS-PB;
- dados dos julgamentos de naturalidade, precisão e compreensão da fala;

Dados das medidas acústicas:

- dados das medidas acústicas do *Script Expression Evaluator*;

Correlação dos dados perceptivo-auditivos e acústicos:

- correlação dos dados de naturalidade, precisão e compreensão da fala e dados do VPAS-PB.

Os resultados das extrações das medidas acústicas e da caracterização dos ajustes de qualidade de voz serão confrontados como forma de se traçar correlações entre as dimensões acústica e perceptiva e considerados em relação às metáforas apresentadas. As metáforas do prisma, do disco voador e da catedral aqui referidas são utilizadas no cotidiano do meu trabalho como professor de canto. Como as metáforas são comumente utilizadas em contextos do ensino de canto, o recurso à análise fonética, em suas dimensões acústica, articulatória e perceptiva, pode revelar os ajustes fonatórios e articulatórios que o cantor faz influenciado pelo efeito impressionante das metáforas. Os procedimentos metodológicos aqui desenvolvidos foram replicados por Alves (2013) e visam a possibilitar a aquisição de mais conhecimento sobre o efeito das metáforas nas configurações do trato vocal em emissões de canto.

## 5. Resultados

Retomamos, antes de apresentar os resultados da análise perceptiva e acústica os enunciados (Quadro 1) e as metáforas (Quadro 2) utilizadas para eliciar as repetições da canção "Sapo Jururu" pelos sujeitos da pesquisa.

**Quadro 1** – Quatro enunciados (E1, E2, E3 e E4) que compuseram o tema folclórico cantado pelos cinco sujeitos da pesquisa.

Enunciado	Frase melódica
E1	Sapo jururu da beira do rio
E2	Quando o sapo canta maninha é que está com frio
E3	A mulher do sapo deve estar lá dentro
E4	fazendo rendinha maninha para o casamento

No quadro 2, a seguir, as repetições realizadas na gravação conforme a orientação da metáfora:

**Quadro 2** – Metáforas desenvolvidas nesta pesquisa, aqui definidas como repetições (R1, R2, R3 e R4).

Repetição	Metáfora
R1	Sem metáfora, sujeito cantando da forma que deseja
R2	Prisma
R3	Disco Voador
R4	Catedral

### 5.1 Análise perceptiva quanto à qualidade vocal

Apresentamos, nos quadros a seguir (Quadros 3, 4, 5, 6 e 7) as avaliações dos ajustes de qualidade vocal utilizados pelos sujeitos em suas emissões dos enunciados da canção folclórica.

No que se refere ao sujeito 1, temos:

**Quadro 3** – Resultados do julgamento perceptivo-auditivo pelo roteiro VPAS-PB (qualidade vocal) para amostras do sujeito 1.

Enunciado	Repetição	Avaliação dos ajustes de qualidade vocal
E1	R1	Corpo de língua recuado
E1	R2	Mandíbula aberta
E1	R3	Corpo de língua recuado e abaixado
E1	R4	Mandíbula aberta
E2	R1	Laringe elevada
E2	R2	Projeção de ponta/lâmina de língua avançada
E2	R3	Laringe abaixada
E2	R4	Laringe abaixada
E3	R2	Laringe abaixada
E3	R3	Laringe abaixada
E3	R4	Mandíbula aberta
E4	R1	Laringe elevada e hiperfunção
E4	R2	Laringe abaixada
E4	R3	Laringe abaixada
E4	R4	Laringe abaixada

Em relação ao sujeito 2, temos:

**Quadro 4** – Resultados do julgamento perceptivo-auditivo pelo roteiro VPAS-PB (qualidade vocal) para amostras do sujeito 2.

Enunciado	Repetição	Avaliação dos ajustes de qualidade vocal
E1	R1	Voz soprosa
E1	R2	Lábios projetados, expansão faríngea e corpo de língua recuado

Enunciado	Repetição	Avaliação dos ajustes de qualidade vocal
E1	R3	Lábios projetados, expansão faríngea e corpo de língua recuado
E1	R4	Mandíbula aberta e expansão faríngea
E2	R1	Voz soprosa e laringe elevada
E2	R2	Corpo de língua recuado
E2	R3	Corpo de língua recuado
E2	R4	Mandíbula aberta e laringe elevada
E3	R1	Voz soprosa
E3	R2	Corpo de língua recuado
E3	R3	Expansão faríngea e corpo de língua recuado
E3	R4	Mandíbula aberta
E4	R1	Laringe elevada e voz soprosa
E4	R2	Corpo de língua recuado
E4	R3	Expansão faríngea
E4	R4	Mandíbula aberta

No que diz respeito ao sujeito 3, temos:

**Quadro 5** – Resultados do julgamento perceptivo-auditivo pelo roteiro VPAS-PB (qualidade vocal) para amostras do sujeito 3.

Enunciado	Repetição	Avaliação dos ajustes de qualidade vocal
E1	R1	Projeção de ponta/lâmina de língua avançada
E1	R2	Corpo de língua recuado
E1	R3	Corpo de língua recuado
E1	R4	Mandíbula aberta
E2	R1	Laringe elevada
E2	R2	Corpo de língua recuado
E2	R4	Corpo de língua avançado e nasalização
E3	R2	Corpo de língua recuado
E3	R4	Corpo de língua recuado e mandíbula aberta
E4	R1	Constricção faríngea
E4	R2	Corpo de língua recuado
E4	R4	Mandíbula aberta

Em relação ao sujeito 4, temos:

**Quadro 6** – Resultados do julgamento perceptivo-auditivo pelo roteiro VPAS-PB (qualidade vocal) para amostras do sujeito 4.

Enunciado	Repetição	Avaliação dos ajustes de qualidade vocal
E1	R1	Voz soprosa
E1	R2	Corpo de língua recuado
E1	R4	Mandíbula aberta
E2	R1	Projeção de ponta/lâmina de língua avançada
E2	R4	Mandíbula aberta
E3	R1	Corpo de língua recuado

Enunciado	Repetição	Avaliação dos ajustes de qualidade vocal
E3	R3	Corpo de língua recuado
E3	R4	Mandíbula aberta
E4	R2	Mandíbula aberta
E4	R3	Corpo de língua recuado
E4	R4	Mandíbula aberta

Quanto ao sujeito 5, temos:

**Quadro 7** – Resultados do julgamento perceptivo-auditivo pelo roteiro VPAS-PB (qualidade vocal) para amostras do sujeito 5.

Enunciado	Repetição	Avaliação dos ajustes de qualidade vocal
E1	R1	Corpo de língua recuado
E1	R2	Projeção de ponta/lâmina de língua avançada
E1	R3	Voz sopro
E1	R4	Mandíbula aberta
E2	R1	Lábios projetados e expansão faríngea
E2	R2	Projeção de ponta/lâmina de língua avançada e laringe abaixada
E2	R3	Voz sopro
E2	R4	Mandíbula aberta
E3	R1	Corpo de língua recuado e constrição faríngea
E3	R2	Projeção de ponta/lâmina de língua avançada
E3	R3	Corpo de língua recuado
E3	R4	Mandíbula aberta
E4	R1	Voz sopro
E4	R2	Projeção de ponta/lâmina de língua avançada e lábios projetados
E4	R3	Corpo de língua recuado
E4	R4	Mandíbula aberta

## 5.2. Análise perceptiva quanto à dinâmica vocal

Apresentamos, nos quadros a seguir (Quadros 8, 9, 10, 11 e 12), as avaliações da dinâmica vocal utilizados pelos sujeitos em suas emissões dos enunciados da canção folclórica.

Em relação ao sujeito 1, temos:

**Quadro 8** – Resultados do julgamento perceptivo-auditivo pelo roteiro VPAS-PB (dinâmica vocal) para amostras do sujeito 1.

Enunciado	Repetição	Avaliação da dinâmica vocal
E1	R1	<i>Loudness</i> diminuído
E1	R4	Taxa de elocução diminuída
E2	R1	<i>Loudness</i> diminuído
E3	R1	<i>Loudness</i> diminuído e <i>pitch</i> de variabilidade alta
E3	R2	<i>Pitch</i> de variabilidade alta
E4	R1	<i>Pitch</i> de variabilidade alta

Quanto ao sujeito 2, temos:

**Quadro 9** – Resultados do julgamento perceptivo-auditivo pelo roteiro VPAS-PB (dinâmica vocal) para amostras do sujeito 2.

Enunciado	Repetição	Avaliação da dinâmica vocal
E1	R1	Instabilidade
E1	R3	<i>Pitch</i> abaixado
E1	R4	<i>Pitch</i> abaixado
E2	R3	<i>Loudness</i> diminuído
E3	R2	<i>Loudness</i> diminuído
E3	R3	<i>Loudness</i> diminuído
E3	R4	<i>Loudness</i> aumentado
E4	R2	<i>Pitch</i> elevado
E4	R4	<i>Pitch</i> elevado e <i>loudness</i> aumentado

No que se refere ao sujeito 3, temos:

**Quadro 10** – Resultados do julgamento perceptivo-auditivo pelo roteiro VPAS-PB (dinâmica vocal) para amostras do sujeito 3.

Enunciado	Repetição	Avaliação da dinâmica vocal
E1	R3	Taxa de elocução diminuída
E1	R4	<i>Loudness</i> aumentado
E2	R1	<i>Pitch</i> de variabilidade alta e <i>pitch</i> elevado
E2	R2	<i>Pitch</i> elevado
E2	R3	Taxa de elocução diminuída
E2	R4	<i>Pitch</i> elevado e <i>loudness</i> aumentado
E3	R1	<i>Pitch</i> elevado
E3	R2	<i>Loudness</i> aumentado
E3	R3	Taxa de elocução diminuída
E3	R4	<i>Loudness</i> aumentado
E4	R2	<i>Pitch</i> elevado
E4	R3	Taxa de elocução diminuída
E4	R4	<i>Loudness</i> aumentado e <i>pitch</i> elevado

Em relação ao sujeito 4, temos:

**Quadro 11** – Resultados do julgamento perceptivo-auditivo pelo roteiro VPAS-PB (dinâmica vocal) para amostras do sujeito 4.

Enunciado	Repetição	Avaliação da dinâmica vocal
E1	R1	<i>Pitch</i> abaixado
E1	R3	<i>Pitch</i> abaixado e <i>loudness</i> diminuído
E1	R4	<i>Loudness</i> diminuído
E2	R2	<i>Loudness</i> diminuído
E2	R3	<i>Loudness</i> aumentado
E2	R4	<i>Loudness</i> aumentado
E3	R2	<i>Loudness</i> diminuído

Enunciado	Repetição	Avaliação da dinâmica vocal
E3	R3	<i>Pitch</i> abaixado
E3	R4	<i>Loudness</i> diminuído
E4	R1	<i>Loudness</i> diminuído
E4	R4	<i>Loudness</i> aumentado

No que se refere ao sujeito 5, temos:

**Quadro 12** – Resultados do julgamento perceptivo-auditivo pelo roteiro VPAS-PB (dinâmica vocal) para amostras do sujeito 5.

Enunciado	Repetição	Avaliação da dinâmica vocal
E1	R2	<i>Loudness</i> aumentado
E1	R3	<i>Pitch</i> abaixado
E2	R3	<i>Pitch</i> elevado
E2	R4	<i>Loudness</i> aumentado
E3	R3	<i>Pitch</i> abaixado
E4	R1	<i>Pitch</i> elevado
E4	R4	<i>Pitch</i> elevado

### 5.3. Análise acústica

As medidas extraídas por meio da aplicação do *SG-Expression evaluator* são apresentadas a seguir.

**Tabela 1** – Medidas extraídas por meio da aplicação do *SG-Expression Evaluator* indicando todos os valores para o Enunciado 1

Sujeito	Repetição	medf0	samp- qf0	q99, 5f0	skf0	mdf0	s-df0	skdf0	skint	mSpTilt	sSpTilt	skSpec- Tilt	Ltas
1	1	0,30	0,76	0,59	0,07	0,03	0,24	0,12	0,05	0,25	0,28	1,28	1,43
1	2	0,38	0,55	0,60	0,34	0,05	0,21	0,33	0,01	0,23	0,28	1,45	1,14
1	3	0,24	0,24	0,26	-0,18	0,01	0,07	-0,48	0,04	0,30	0,33	1,25	1,56
1	4	0,57	0,34	0,55	-0,13	0,06	0,10	-0,34	0,08	0,21	0,26	1,45	1,02
2	1	0,37	0,57	0,60	0,31	0,25	0,23	0,05	0,30	0,29	0,32	1,26	2,23
2	2	0,35	0,58	0,58	0,39	0,12	0,23	0,04	0,34	0,37	0,41	1,23	2,20
2	3	-0,02	0,51	0,41	0,38	0,21	0,19	0,26	0,09	0,41	0,44	1,18	2,37
2	4	0,56	0,63	0,74	0,36	0,10	0,25	0,11	0,20	0,31	0,34	1,27	2,25
3	1	0,25	0,62	0,55	0,30	0,12	0,23	0,29	0,18	0,27	0,31	1,33	1,90
3	2	0,29	0,62	0,56	0,28	0,28	0,23	0,15	0,24	0,27	0,31	1,35	1,64
3	3	0,29	0,60	0,55	0,25	0,27	0,23	0,08	0,22	0,30	0,37	1,41	1,78
3	4	0,28	0,60	0,56	0,35	0,08	0,23	0,15	0,05	0,20	0,26	1,51	1,01
4	1	-0,14	0,58	0,36	0,29	-0,02	0,20	0,48	0,11	0,31	0,35	1,28	2,23
4	2	0,06	0,56	0,39	0,14	0,02	0,19	0,32	0,07	0,30	0,33	1,24	2,17
4	3	0,00	0,54	0,32	0,15	0,19	0,19	0,25	0,07	0,28	0,32	1,27	1,97
4	4	0,04	0,57	0,39	0,21	0,21	0,20	0,36	0,05	0,24	0,28	1,37	1,85
5	1	-0,17	0,48	0,27	0,32	-0,04	0,19	0,31	0,10	0,34	0,37	1,26	1,57
5	2	0,02	0,44	0,27	0,16	0,08	0,16	0,09	0,18	0,30	0,33	1,26	1,79
5	3	-0,13	0,50	0,27	0,28	0,14	0,19	0,15	0,01	0,37	0,41	1,24	1,96
5	4	-0,14	0,40	0,18	0,17	0,18	0,15	0,17	0,02	0,36	0,40	1,25	1,72

**Tabela 2 –** Medidas extraídas por meio da aplicação do *SG-Expression Evaluator* indicando todos os valores para o Enunciado 2

Sujeito	Repetição	medf0	samp-qf0	q99,5f0	skf0	mdf0	s-df0	skdf0	skint	mSpTilt	sSpTilt	skSpec-Tilt	Ltas
1	1	1,05	0,89	0,93	-0,09	-0,06	0,27	-0,61	0,12	0,19	0,23	1,45	1,34
1	2	0,92	0,51	0,81	0,01	-0,07	0,18	-0,16	0,09	0,17	0,21	1,51	1,21
1	3	0,38	0,44	0,43	-0,13	-0,03	0,13	-0,18	0,09	0,23	0,28	1,42	1,56
1	4	0,56	0,48	0,58	-0,03	-0,11	0,15	-0,19	0,15	0,20	0,23	1,38	1,19
2	1	1,02	0,78	1,03	0,10	-0,04	0,25	0,05	0,25	0,21	0,24	1,30	1,26
2	2	1,04	0,80	1,03	0,13	-0,17	0,26	0,12	0,21	0,28	0,32	1,29	1,79
2	3	0,64	0,69	0,77	0,08	-0,21	0,22	0,14	0,22	0,34	0,38	1,28	1,96
2	4	1,43	0,91	1,26	0,03	0,69	0,28	0,05	0,12	0,18	0,22	1,37	1,12
3	1	1,04	0,80	1,03	0,05	0,08	0,27	-0,04	0,20	0,16	0,19	1,45	1,18
3	2	1,02	0,75	1,00	-0,04	0,11	0,28	-0,17	0,20	0,13	0,18	1,43	0,78
3	3	1,03	0,76	0,99	0,00	0,01	0,26	-0,17	0,12	0,13	0,18	1,46	0,70
3	4	1,01	0,75	0,99	0,05	-0,04	0,26	-0,14	0,09	0,10	0,15	1,56	0,61
4	1	0,22	0,56	0,65	-0,10	-0,06	0,18	0,01	0,09	0,29	0,32	1,27	2,01
4	2	0,25	0,51	0,43	-0,08	0,00	0,17	-0,36	0,08	0,28	0,31	1,25	2,02
4	3	0,21	0,51	0,43	-0,14	-0,03	0,18	-0,47	0,16	0,30	0,34	1,32	1,90
4	4	0,43	0,58	0,50	-0,13	-0,17	0,18	-0,66	0,21	0,21	0,25	1,34	1,61
5	1	0,17	0,51	0,37	0,08	-0,15	0,17	-0,16	0,07	0,25	0,29	1,31	1,90
5	2	0,23	0,53	0,89	-0,02	-0,06	0,19	0,70	0,15	0,27	0,31	1,29	1,72
5	3	1,80	1,25	1,75	0,07	-0,51	0,39	-0,24	0,02	0,30	0,35	1,28	2,09
5	4	0,04	0,42	0,26	-0,06	-0,02	0,14	-0,31	0,03	0,27	0,31	1,25	1,87

**Tabela 3 –** Medidas extraídas por meio da aplicação do *SG-Expression Evaluator* indicando todos os valores para o Enunciado 3

Sujeito	Repetição	medf0	samp-qf0	q99,5f0	skf0	mdf0	s-df0	skdf0	skint	mSpTilt	sSpTilt	skSpec-Tilt	Ltas
1	1	0,43	0,93	1,05	0,13	0,06	0,23	2,09	0,07	0,20	0,23	1,37	1,39
1	2	0,46	0,99	1,20	0,14	0,10	0,28	2,22	0,11	0,20	0,24	1,56	1,14
1	3	0,25	0,27	0,27	-0,14	0,68	0,08	-1,53	0,08	0,20	0,25	1,64	1,20
1	4	0,56	0,31	0,45	-0,15	0,11	0,09	-0,82	0,13	0,17	0,22	1,69	1,00
2	1	0,38	0,58	0,56	0,28	0,16	0,22	0,00	0,14	0,25	0,29	1,46	1,50
2	2	0,52	0,58	0,58	0,13	0,10	0,22	-0,10	0,24	0,29	0,32	1,31	2,04
2	3	0,12	0,52	0,42	0,20	0,05	0,19	0,03	0,14	0,32	0,36	1,31	2,04
2	4	0,71	0,66	0,75	0,19	0,42	0,25	-0,03	0,17	0,17	0,22	1,53	1,06
3	1	0,38	0,63	0,55	0,20	0,40	0,23	-0,09	0,04	0,18	0,22	1,50	1,27
3	2	0,53	0,58	0,55	0,05	0,46	0,23	-0,21	0,09	0,20	0,25	1,55	1,31
3	3	0,54	0,58	0,55	-0,01	0,35	0,22	-0,18	0,08	0,17	0,22	1,64	0,96
3	4	0,33	0,61	0,55	0,26	0,37	0,22	-0,02	0,09	0,14	0,19	1,80	0,86
4	1	-0,09	0,51	0,31	0,19	0,11	0,19	-0,02	0,11	0,27	0,31	1,35	1,99
4	2	-0,02	0,53	0,36	0,17	0,02	0,20	0,05	0,07	0,24	0,28	1,37	1,89
4	3	-0,04	0,49	0,34	0,16	0,11	0,20	0,07	0,06	0,28	0,31	1,31	2,02
4	4	1,70	1,06	1,57	0,12	0,29	0,35	-0,08	0,11	0,24	0,28	1,38	1,72
5	1	1,31	1,12	1,35	0,10	0,10	0,34	-0,20	0,06	0,25	0,29	1,40	1,62
5	2	-0,14	0,45	0,21	0,22	0,24	0,17	0,12	0,03	0,26	0,30	1,40	1,67
5	3	-0,10	0,62	0,34	0,19	0,10	0,24	0,11	0,07	0,27	0,31	1,41	1,79
5	4	-0,18	0,45	0,19	0,18	0,25	0,17	0,16	0,02	0,26	0,30	1,40	1,61

**Tabela 4 –** Medidas extraídas por meio da aplicação do *SG-Expression Evaluator* indicando todos os valores para o Enunciado 4

Sujeito	Repetição	medf0	samp-qf0	q99,5f0	skf0	mdf0	s-df0	skdf0	skint	mSpTilt	sSpTilt	skSpecTilt	Ltas
1	1	1,04	0,73	1,11	0,19	-0,20	0,26	0,15	0,07	0,22	0,25	1,35	1,56
1	2	0,65	0,50	0,67	0,08	-0,27	0,18	-0,10	0,22	0,19	0,23	1,45	1,04
1	3	0,34	0,48	0,54	0,15	-0,29	0,17	0,12	0,08	0,21	0,25	1,49	1,12
1	4	0,58	0,48	0,59	-0,01	-0,19	0,16	-0,34	0,17	0,19	0,23	1,46	1,07
2	1	1,01	0,81	1,03	0,12	-0,11	0,25	0,10	0,23	0,26	0,30	1,39	1,33
2	2	1,04	0,81	1,06	0,10	-0,17	0,26	0,14	0,24	0,30	0,33	1,28	1,76
2	3	0,63	0,68	0,74	0,10	-0,19	0,22	0,16	0,26	0,32	0,35	1,28	1,87
2	4	1,37	0,90	1,27	0,08	-0,04	0,28	0,13	0,21	0,18	0,22	1,58	0,94
3	1	1,06	0,75	0,98	0,08	0,06	0,24	-0,06	0,03	0,18	0,23	1,49	1,14
3	2	1,04	0,80	0,93	0,07	0,01	0,25	-0,15	0,09	0,16	0,22	1,70	0,89
3	3	1,02	0,72	0,94	0,04	0,06	0,23	-0,11	0,13	0,13	0,20	1,65	0,58
3	4	1,04	0,78	0,94	0,05	-0,07	0,24	-0,07	0,12	0,11	0,19	1,80	0,41
4	1	0,12	0,50	0,37	-0,08	-0,02	0,16	-0,42	0,05	0,32	0,35	1,27	2,15
4	2	0,16	0,45	0,36	-0,08	0,02	0,15	-0,39	0,12	0,30	0,32	1,25	1,96
4	3	0,26	0,51	0,36	-0,15	-0,13	0,16	-0,76	0,07	0,29	0,32	1,25	1,97
4	4	0,33	0,51	0,37	-0,16	0,11	0,15	-0,99	0,12	0,26	0,30	1,30	1,69
5	1	0,09	0,66	0,44	0,00	0,35	0,20	-0,15	0,05	0,30	0,35	1,30	2,14
5	2	0,20	0,49	0,41	-0,01	-0,07	0,16	-0,62	0,10	0,28	0,32	1,32	1,75
5	3	0,22	0,66	0,51	0,02	0,23	0,21	-0,25	0,07	0,30	0,35	1,31	1,80
5	4	1,63	1,23	1,66	0,03	-0,48	0,34	-0,08	0,05	0,31	0,35	1,31	2,04

A aplicação do Test t para duas amostras independentes revelou diferenças significativas somente em relação aos modos de cantar influenciados pela metáfora do disco voador e da catedral.

Em relação ao modo de cantar influenciado pela metáfora do disco voador, foram verificadas diferenças significativas no que diz respeito ao desvio padrão de declínio espectral (sSpTILT) e à média de declínio espectral (mSpTILT), como se pode observar nas tabelas que seguem (Tabelas 5 e 6):

**Tabela 5 –** Teste t para duas amostras independentes (mSpTilt)

Intervalo de confiança de 95% para a diferença entre as médias: ] -0,067 ; -0,001 [

Diferença	-0,034
t (valor observado)	-2,039
t  (valor crítico)	1,991
GL	78
p-valor (bilateral)	0,045
Alfa	0,05

**Tabela 6 –** Teste t para duas amostras independentes (sSpTilt):

Intervalo de confiança de 95% para a diferença entre as médias: ] -0,068 ; -0,007 [

Diferença	-0,037
t (valor observado)	-2,435
t  (valor crítico)	1,991
GL	78
p-valor (bilateral)	0,017
Alfa	0,05

Em relação ao modo de cantar influenciado pela metáfora da catedral, foram verificadas diferenças significativas em relação ao desvio padrão de declínio espectral (sSpTILT) e à média de declínio espectral (mSpTILT), desvio padrão de LTAS e assimetria de declínio espectral (skSTILT), como é possível observar nas tabelas a seguir (Tabelas 7, 8, 9 e 10).

**Tabela 7 – Teste t para duas amostras independentes (mSpTilt):**

Intervalo de confiança de 95% para a diferença entre as médias: ] 0,010; 0,075 [

Diferença	0,042
t (valor observado)	2,580
t  (valor crítico)	1,991
GL	78
p-valor (bilateral)	0,012
Alfa	0,05

**Tabela 8 – Teste t para duas amostras independentes/Teste bilateral (sSpTilt):**

Intervalo de confiança de 95% para a diferença entre as médias: ] 0,007 ; -0,068 [

Diferença	0,037
t (valor observado)	2,435
t  (valor crítico)	1,991
GL	78
p-valor (bilateral)	0,017
Alfa	0,05

**Tabela 9 – Teste t para duas amostras independentes (skSpecTilt):**

Intervalo de confiança de 95% para a diferença entre as médias: ] -0,152; -0,016 [

Diferença	-0,084
t (valor observado)	-2,461
t  (valor crítico)	1,991
GL	78
p-valor (bilateral)	0,016
Alfa	0,05

**Tabela 10 – Teste t para duas amostras independentes (Ltas):**

Intervalo de confiança de 95% para a diferença entre as médias: ] 0,075; 0,535 [

Diferença	0,305
t (valor observado)	2,640
t  (valor crítico)	1,991
GL	78
p-valor (bilateral)	0,010
Alfa	0,05

A aplicação do teste de análise discriminante revelou as variáveis que segregam os quatro modos de canto. O perfil de correlação entre as variáveis é apresentado na tabela a seguir (Tabela 11), na qual os valores destacados são iguais ou superiores a 0,25. Esses valores revelam o peso das variáveis na segregação dos modos do canto.

**Tabela 11** – Perfil de correlação entre as variáveis: dados de aplicação do teste de análise discriminante

	F1	F2	F3
medf0	-0,138	-0,148	0,118
sampqf0	0,091	-0,173	0,166
q99, 5f0	-0,028	-0,187	0,095
skf0	0,145	-0,134	-0,032
mdf0	-0,073	-0,018	0,070
s-df0	0,109	-0,161	0,083
skdf0	0,191	-0,181	-0,135
Skint	0,039	-0,126	<b>-0,233</b>
mSpTilt	<b>0,223</b>	<b>0,250</b>	-0,127
sSpTilt	0,189	<b>0,300</b>	-0,125
skSpecTilt	<b>-0,284</b>	-0,069	0,084
Ltas	<b>0,284</b>	0,128	-0,091
Qualid-CLR	0,171	0,145	<b>-0,411</b>
Qualid-MA	<b>-0,673</b>	<b>-0,287</b>	<b>0,249</b>
Qualid-CLR CLAb	0,013	<b>0,224</b>	0,010
Qualid-VS	<b>0,284</b>	0,038	<b>0,229</b>
Qualid-CLR EF LP	0,015	0,111	-0,168
Qualid-EF MA	-0,183	-0,067	0,105
Qualid-PLLA	0,161	-0,175	<b>-0,218</b>
Qualid-	0,124	<b>0,210</b>	0,018
Qualid-LE	<b>0,231</b>	-0,127	0,187
Qualid-LA	-0,123	0,159	-0,099
Qualid-LE VS	<b>0,231</b>	-0,127	0,187
Qualid-LE MA	-0,183	-0,067	0,105
Qualid-CLAv N	-0,183	-0,067	0,105
Qualid-EF LP	0,162	-0,090	0,131
Qualid-LA PLLA	0,009	-0,068	<b>-0,246</b>
Qualid-CLR EF	0,013	<b>0,224</b>	0,010
Qualid-CLR MA	-0,183	-0,067	0,105
Qualid-CF CLR	0,162	-0,090	0,131
Qualid-HF LE	0,162	-0,090	0,131
Qualid-EF	0,013	<b>0,224</b>	0,010
Qualid-CF	0,162	-0,090	0,131
Qualid-LP PLLA	0,009	-0,068	<b>-0,246</b>
Dinam-LD	0,057	-0,053	-0,038
Dinam-	0,158	-0,106	-0,156
Dinam-TED	-0,061	<b>0,381</b>	0,066
Dinam-I	0,162	-0,090	0,131
Dinam-PA	0,012	<b>0,312</b>	0,116
Dinam-LA	<b>-0,376</b>	-0,110	0,052

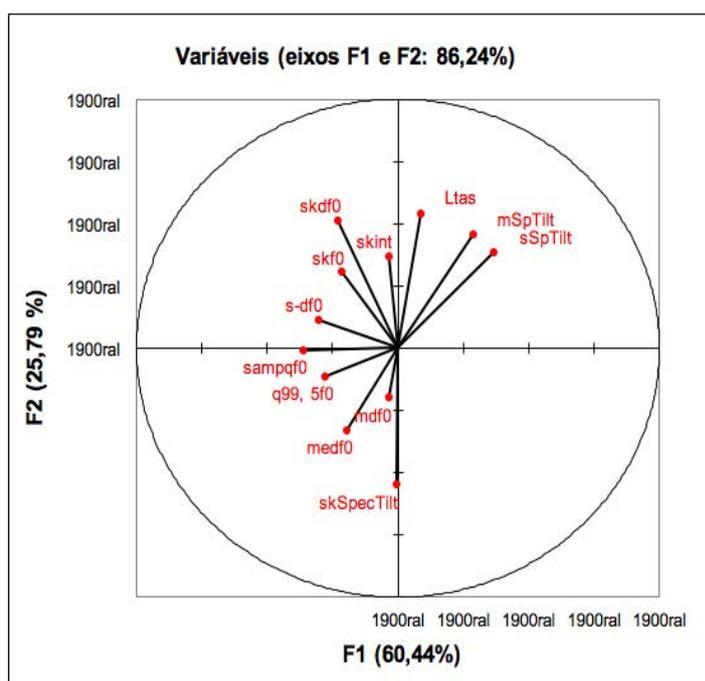
Dinam-LD PA	0,013	<b>0,224</b>	0,010
Dinam-PE PVA	0,162	-0,090	0,131
Dinam-PE	0,071	-0,088	-0,142
Dinam-LA PE	<b>-0,321</b>	-0,118	0,184
Dinam-LD PVA	0,162	-0,090	0,131
Dinam-PVA	0,121	-0,112	-0,082

A consideração das variáveis acústicas em relação aos fatores revelou que as medidas de desvio padrão de declínio espectral (sSpTILT) e de média de declínio espectral (mSpTILT) foram influentes nos três fatores (Tabela 12).

**Tabela 12 – Correlações Variáveis/Fatores**

	F1	F2	F3
medf0	-0,195	<b>-0,331</b>	<b>-0,292</b>
sampqf0	<b>-0,361</b>	-0,010	0,138
q99, 5f0	<b>-0,280</b>	-0,115	-0,190
skf0	-0,215	<b>0,306</b>	0,013
mdf0	-0,032	-0,198	-0,056
s-df0	<b>-0,303</b>	0,115	0,071
skdf0	-0,233	<b>0,515</b>	-0,120
Skint	-0,034	<b>0,371</b>	<b>-0,441</b>
mSpTilt	<b>0,290</b>	<b>0,457</b>	<b>0,578</b>
sSpTilt	<b>0,367</b>	<b>0,388</b>	<b>0,594</b>
skSpecTilt	-0,007	<b>-0,547</b>	<b>-0,478</b>
Ltas	0,086	<b>0,542</b>	<b>0,556</b>

O gráfico a seguir demonstra as variáveis nos eixos F1 e F2. Observamos que as medidas de desvio de padrão de declínio espectral (sSpTILT) e de média de declínio espectral (mSpTILT) ocorrem no mesmo quadrante que a LTas (Figura 5).



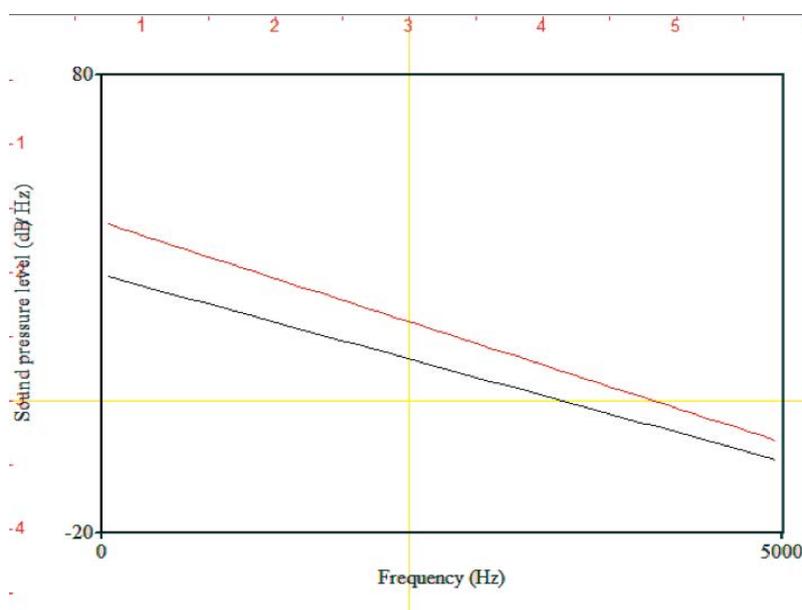
**Figura 5 – Gráfico demonstrativo das variáveis nos eixos F1 e F2**

Os resultados do experimento apontam a influência das metáforas na adoção de configurações do trato vocal pelos sujeitos da pesquisa. Essas configurações são resultados de ajustes de qualidade de voz, implementados pelos sujeitos com base em analogias que estabeleceram entre formas: a suscitada pela metáfora e a moldada pelo trato vocal.

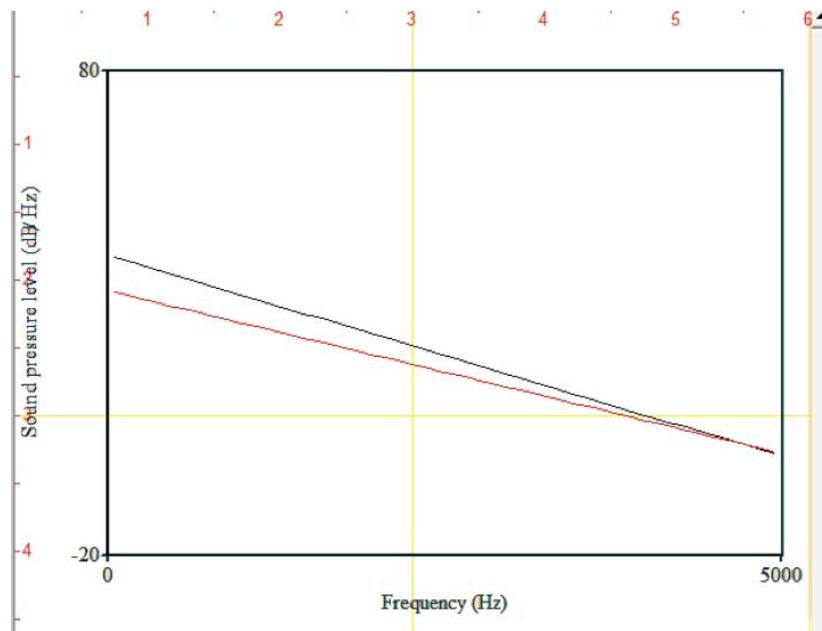
Com base nos resultados obtidos na avaliação perceptiva dos ajustes e da qualidade da dinâmica vocal e da aplicação dos testes estatísticos, é possível verificar que as medidas de longo termo relacionadas ao declínio espectral mostraram-se relevantes.

Essas medidas refletem modificações nos ajustes da qualidade vocal e por isso são destacadas na literatura de natureza fonético-acústica sobre qualidade vocal (Sundberg, 1985).

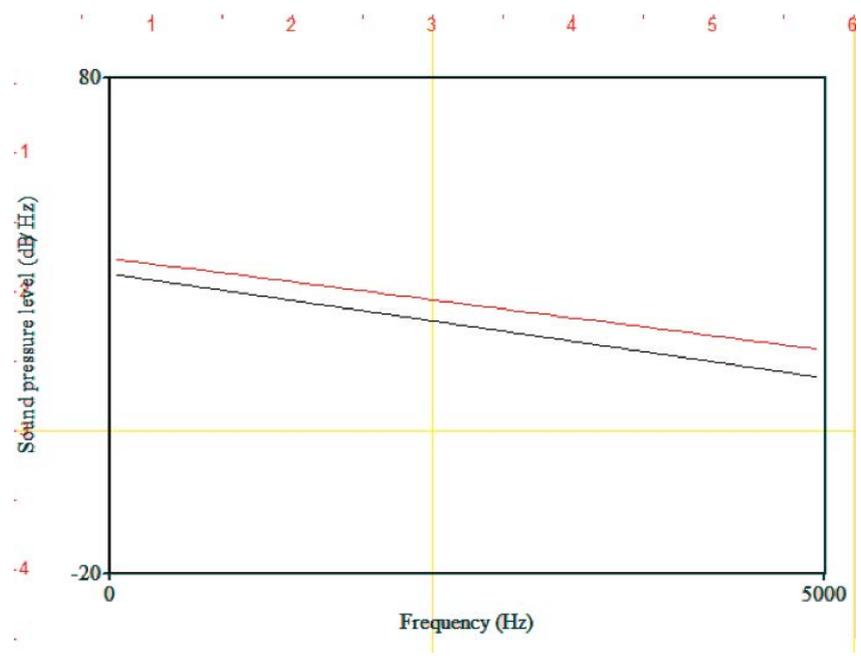
Nas figuras a seguir (Figuras 6, 7, 8, 9, 10), apresentamos as linhas de declinação (*trend lines*) referentes aos espectros de longo termo (LTAS) das emissões vocais de um dos enunciados pelos cinco sujeitos da pesquisa. As linhas em preto referem-se às emissões sem instrução e as linhas vermelhas, às emissões eliciadas pela metáfora da catedral.



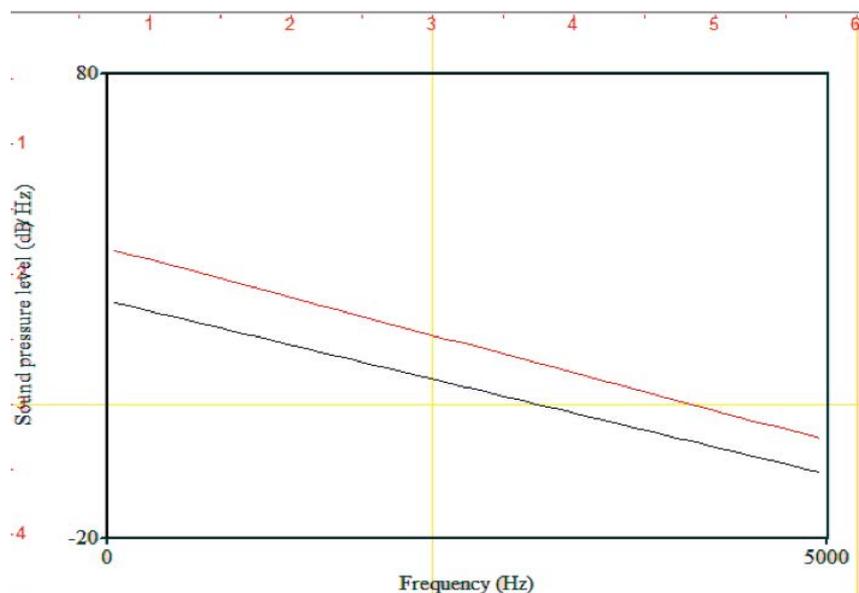
**Figura 6** – Linha de declinação (*trend lines*) referente aos espectros de longo termo (LTAS) das emissões vocais de um dos enunciados pelo sujeito da pesquisa S1. A linha na cor preta refere-se à emissão sem instrução e a linha na cor vermelha, à emissão eliciada pela metáfora da catedral.



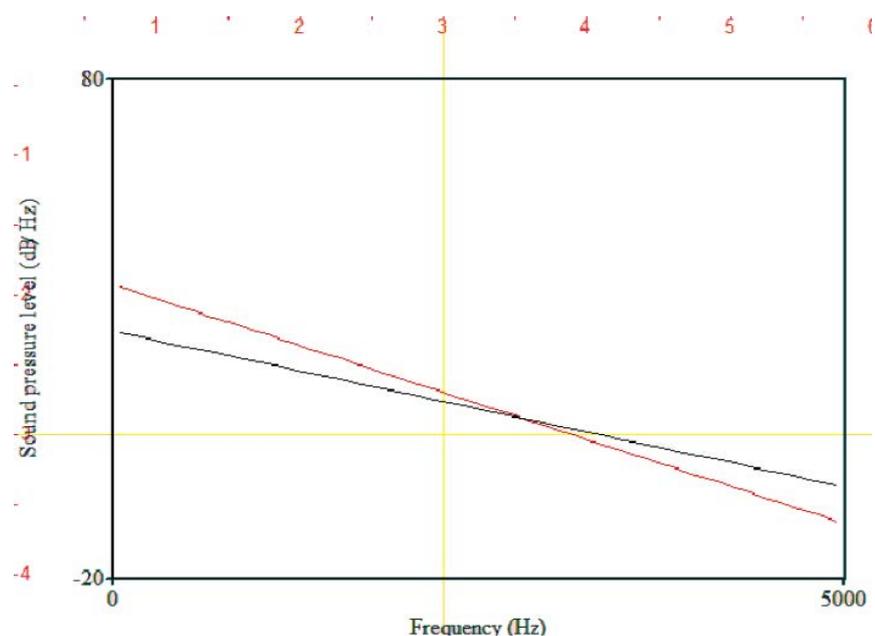
**Figura 7** – Linha de declinação (*trend lines*) referente aos espectros de longo termo (LTAS) das emissões vocais de um dos enunciados pelo sujeito da pesquisa S2. A linha na cor preta refere-se à emissão sem instrução e a linha na cor vermelha, à emissão elicitada pela metáfora da catedral.



**Figura 8** – Linha de declinação (*trend lines*) referente aos espectros de longo termo (LTAS) das emissões vocais de um dos enunciados pelo sujeito da pesquisa S3. A linha na cor preta refere-se à emissão sem instrução e a linha na cor vermelha, à emissão elicitada pela metáfora da catedral.



**Figura 9** – Linha de declinação (*trend lines*) referente aos espectros de longo termo (LTAS) das emissões vocais de um dos enunciados pelo sujeito da pesquisa S4. A linha na cor preta refere-se à emissão sem instrução e a linha na cor vermelha, à emissão elicitada pela metáfora da catedral.



**Figura 10** – Linha de declinação (*trend lines*) referente aos espectros de longo termo (LTAS) das emissões vocais de um dos enunciados pelo sujeito da pesquisa S5. A linha na cor preta refere-se à emissão sem instrução e a linha na cor vermelha, à emissão elicitada pela metáfora da catedral.

A inspeção das linhas de declínio relacionadas ao espectro de longo termo demonstra que ocorrem, como verificado pelas medidas e pelos testes estatísticos, alterações no grau de inclinação ou na intensidade ao longo das frequências.

## 6. Conclusão

Os resultados indicam que as duas metáforas, a do disco voador e a da catedral, levaram os sujeitos a modificar a configuração do trato vocal e, desse modo, a alterarem a qualidade vocal. No caso da metáfora da catedral, o ajuste de mandíbula aberta ou a combinação do ajuste de mandíbula aberta (MA) de laringe abaixada (LA) foram utilizados na totalidade das emissões dos sujeitos. Por vezes o ajuste de corpo de língua recuado também foi utilizado.

Todos esses ajustes apontam para a adoção de uma estratégia de ampliação das cavidades de ressonância, o que provoca alteração na qualidade vocal.

Sendo a metáfora da catedral amplamente utilizada no treinamento vocal em canto lírico, evidencia-se a sua real contribuição para a modificação de ajustes de qualidade de voz nesse tipo de treinamento.

Esta pesquisa traz contribuições para reflexões sobre os efeitos das metáforas no ensino do canto, especialmente do canto lírico. Ela fornece dados experimentais sobre os efeitos das metáforas nas implementações de ajustes que a técnica de ensino de canto lírico por meio de metáforas almeja. Em relação à emissão vocal sem tensão nas pregas vocais e no restante do trato vocal, efeito que é buscado pela apresentação da metáfora do disco voador, e à projeção adequada da emissão vocal no canto lírico dentro de auditórios sem a necessidade de utilização de microfones, efeito que é buscado pela metáfora da catedral, foram observados nos dados a implementação de ajustes de qualidade vocal, o que resultou em mudanças nas configurações do trato vocal.

Verificou-se maior impacto da metáfora da catedral na emissão vocal do que das demais metáforas. A sugestão da metáfora da catedral levou os sujeitos a utilizarem estratégias de ampliação de cavidades de ressonância, principalmente pelo ajuste de mandíbula aberta, mas também por meio de abaixamento da laringe e do corpo de língua recuada. Essas correlações entre medidas acústicas e resultados dos testes perceptivos ficaram evidenciadas nos testes estatísticos e nas alterações nos ajustes de qualidade de voz bem evidenciadas na verificação das linhas de declínio no espectro de longo termo.

As metáforas instrutivas provocaram alterações nas configurações articulatórias que se moldaram para contemplar a associação elicitada pela metáfora, percorrendo o trajeto inverso ao da metáfora sonora, conforme argumenta Madureira (2011).

Conhecer as configurações dos ajustes de qualidade de voz e seus efeitos acústicos pode auxiliar na obtenção de efeitos que almejem a projeção vocal, a diminuição da tensão vocal nas pregas e a emissão vocal lírica e ser um fator que agregue informação sobre as configurações vocais à aplicação de metáforas no ensino do canto.

## Referências Bibliográficas

- ALBANO, E. Fazendo sentido do som. **Ilha do desterro**. Florianópolis: UFSC, v. 18, p. 11-26, 1988.
- ALVES, C. C. **Efeitos das metáforas no ensino do canto: dados acústicos e de imagem do trato vocal**. Dissertação de Mestrado. São Paulo: PUCSP, 2013.
- BAKHTIN, M. **Estética da criação verbal**. Tradução Paulo Bezerra. São Paulo: Martins Fontes, 2 ed., 1992.
- BARBOSA, P. A. Detecting changes in speech expressiveness in participants of a radio program. In: **Interspeech**, 2009. Brighton. **Proceedings...** Brighton: United Kingdom, p. 2155-2158, 2009.
- BELLER, G. Transformation of expressivity in speech. In: LANG, P. (Ed.). **The role of prosody in the expression of emotions in English and in French**, 2008.
- BERBER SARDINHA, T. **Metáfora**. São Paulo: Parábola, 2007.
- CAMARGO, Z.; MADUREIRA, S. The acoustic analysis of speech samples designed for the Voice

Profile Analysis Scheme for Brazilian Portuguese (BP-VPAS): long-term f0 and intensity measures. In: BOTINIS, A. (Ed.). **ISCA**. Athens: v. 1. p. 33-36, 2010.

\_\_\_\_\_. Voice quality analysis from a phonetic perspective: Voice Profile Analysis Scheme Profile for Brazilian Portuguese (BPVPAS). In: **4th Conf on Speech Prosody**. 2008, Campinas. **Proceedings ...** Campinas: v. 1, p. 57-60, 2008.

\_\_\_\_\_; RUSILO, L. C., MADUREIRA, S. Evaluating speech samples designed for the Voice Profile Analysis Scheme for Brazilian Portuguese. In: **4th ISCA Tutorial and Res Workshop on Experim Ling**. 2011. Paris, **Proceedings...** Paris: p. 55-58, 2011.

CAMERON, L. **Metaphor in educational discourse**. London: Continuum, 2003.

CIENKI, A.; Muller, C. Metaphor, gesture and thought. In: GIBBS JR., R. W. (Ed.). **The Cambridge Hdbk of Metaphor and Thought**. Cambridge. Cambridge U P, 2008.

FANT, G. **Acoustic theory of speech production**. Paris: Mouton, 2. ed., 1970.

FÓNAGY, I. Lang.s within **Lang.s: an evolutive approach**. Philadelphia: John Benjamins Pub Co, 2000.

\_\_\_\_\_. **La vive voix**. Paris: Payot, 1983.

JAKOBSON, R. **Seis lições sobre o som e o sentido**. Lisboa: Moraes, 1977.

LAKOFF, George; JOHNSON, Mark L. **Metaphors we live by**. Chicago: The U of Chicago P, 1980.

LAVÉ, John. The phonetic evaluation of voice quality. In: KENT, R. D.; BALL, M. J. **Voice quality measurement**, San Diego: Singular Pub Group Inc., 2000.

\_\_\_\_\_. **Princ of phonetics**. Cambridge: Cambridge U P, 1994.

\_\_\_\_\_. **The phonetic description of voice quality**. Cambridge: Cambridge U P, 1980.

\_\_\_\_\_. **Voice quality: a classified Res bibliography**. John Benjamins Pub, Berlim, 1979.

LAVÉ, J.; MACKENZIE-BECK, J. **Vocal profile Analysis Scheme – VPAS**. Edinburgh: Queen Margareth U College, 2007.

MACKENZIE-BECK, J. Perceptual analysis of voice quality: the place of vocal profile analysis. In: HARDCASTLE W. J; MACKENZIE-BECK J. **A figure of speech: a festschrift for John Lavé**. Hillsdale: Lawrence Erlbrum Ass, p. 285-322, 2005.

MADUREIRA, S. The investigation of speech expressivity. In: Mello, H.; Panunzi, A.; Raso, T. (Eds.). **Illocution, modality, attitude, information patterning and speech annotation**. Firenze: Firenze U P, v. 1, p. 101-118, 2011.

\_\_\_\_\_; CAMARGO, Z. A. Explorando o simbolismo sonoro na investigação da fala expressiva. In: **encontro de linguística experimental – Internacional Speech Com Assoc**, v. 03., Atenas, 2010.

\_\_\_\_\_. Reciting a sonnet: production strategies and perceptual. In: **Proceedings of the 4th Conf on Speech Prosody**, Campinas: v.1, p. 697-700, 2008.

\_\_\_\_\_. **O sentido do som**. Tese de Doutorado. São Paulo: PUCSP, 1992.

MARCHESI, M. **Bel Canto: a theoretical & practical vocal method**. New York: Dover Pub Inc., 1970.

MARIZ, J. **Entre a expressão e a técnica: a terminologia do professor de canto**. Tese de Doutorado. São Paulo: Unesp, 2013.

ORLANDI, E. P. **O que é linguística**. São Paulo: Brasiliense, 1986.

PESSOA, A. N. **Correlatos perceptivos e acústicos de qualidade e dinâmica vocal na fala de crianças usuárias de implante coclear**. Tese de Doutorado. São Paulo: PUCSP, 2012.

SALOMÃO, G. L. **Registros vocais no canto: aspectos perceptivos, acústicos, aerodinâmicos e fisiológicos da voz modal e da voz de falsete**. Tese de Doutorado. São Paulo: PUCSP, 2008.

SAUSSURE, F. **Cours de linguistique générale**, Paris: Payot & Rivage, 1916.

SINCLAIR, J. M. **Corpus, concordance, collocation**. Oxford: Oxford U P, 1991.

SOUSA, J. M.; ANDRADA E SILVA, M. A.; FERREIRA, Leslie P. O uso de metáforas como recurso

didático no ensino do canto: diferentes abordagens. **Rev Soc Bras Fonoaudiologia**, São Paulo, 15, v. 03, p. 317-28, 2010.

SUNDBERG, J. **The Sci of the singing voice**. Illinois. Northern Illinois U P, 1987.

\_\_\_\_\_. My Res on the singing voice from a rear-view-mirror perspective. Stockholm: Voice Res Centre/KTH, 1985.

\_\_\_\_\_. Articulatory interpretation of the “singing formant”. **J Acoust Soc Amer**, n. 55, p. 838-844, 1974.

\_\_\_\_\_; GAUFFIN, J. Waveform and spectrum of the glottal voice source. In: LINDBLOM, B.; ÖHMAN, S. (Eds.). **Frontiers of speech Com Res**. London: Academic P, p. 301-320, 1979.

VIOLA, I. C. **O gesto vocal: a arquitetura de um ato teatral**. Tese de Doutorado. São Paulo: PUCSP, p. 209, 2006.

VYGOTSKY, L. **A construção do pensamento e da linguagem**. Tradução Paulo Bezerra. São Paulo: Martins Fontes, (1934) 2001.

ZBIKOWSKI, L. M. Metaphor and music. In: GIBBS JR., R. W. (Ed.). **The Cambridge Hdbk of metaphor and thought**. Cambridge: Cambridge U P, 200.

# ***Gestos Faciais e Corporais e os Tons Alto e Baixo: Qual a Relação?***

VERA PACHECO  
MARIAN OLIVEIRA

# 1. Introdução

Durante uma conversa, nós, falantes de qualquer língua, estamos a todo tempo realizando movimentos com as mãos e assumimos diferentes configurações faciais. Fazemos esses gestos sem nos dar conta, nós os fazemos de forma automática, inconsciente.

A presença de gestos durante a fala é tão forte que, quando estamos falando ao telefone, inevitavelmente mexemos as mãos, a cabeça e os olhos, mesmo cientes de que o nosso interlocutor não está nos vendo. Esse fato permite-nos pensar que a gesticulação é constitutiva da linguagem.

Outra evidência de que os gestos podem compor a linguagem é o incômodo que temos ao ver a Lu, a vendedora virtual da loja Magazine Luiza, ou a Eva Byte, repórter virtual do programa da rede Globo, Fantástico, que entrou no ar em abril de 2004 e que logo, logo deixou de apresentar reportagens. Nesses dois casos específicos, demoramos um pouco a prestar atenção no que elas dizem porque não há sincronia entre seus movimentos e a sua fala. Temos igualmente grande estranhamento quando som e imagem de um vídeo não estão sincronizados, bem como sentimos estranhamento nos primeiros minutos de um filme mal dublado.

O incômodo que nos traz a falta de sincronia entre fala e gestos é um forte indício de que os movimentos feitos por nós não são aleatórios ou simplesmente instintivos. Os gestos são portadores de significado, como filosoficamente demonstrou Merleau-Ponty (1945/1999) e, importantes coadjuvantes no processo de percepção da fala, como demonstrado experimentalmente por Sumby e Pollack (1954) McGurck e MacDonald (1976).

De fato os gestos realizados durante a fala carregam consigo intenções, interesses, sentimentos de quem fala. Para Kendon (2004, p1), “[...] it is through the orientation of the body, and especially, through the orientation of the eyes, that information is provided about a direction and nature of a person’s attention”, o que endossa a hipótese da inseparabilidade entre gestos e fala (Kendon, 1972, McNeill, 1992.). Para Ekman (1972), as expressões faciais realizadas em um ato comunicativo expressam diversas emoções. Assim somos capazes de expressar alegria, tristeza, surpresa só com as caras e bocas que fazemos, sem ser necessário emitir uma palavra sequer. Somos, dessa forma, capazes de deduzir os estados de espírito de quem nos fala.

Dentre as diversas possibilidades de significados que os gestos podem carregar consigo, as informações prosódicas do enunciado são informações linguísticas frequentemente reforçadas pela coocorrência de movimentos corporais e faciais. Nesse caso em particular, trata-se, segundo alguns autores, de prosódia visual (Munhal et al., 2004). Isso porque, ao ver e ouvir alguém falando, as gesticulações e expressões realizadas por quem fala constituem para o ouvinte pistas importantes para a apreensão de aspectos prosódicos.

Partindo destas constatações, nossa hipótese é que muitos dos gestos faciais e corporais realizados pelo falante podem ocorrer quando o mesmo intenciona destacar partes da situação comunicativa que ele julga pertinentes, aquelas partes para as quais ele quer chamar atenção do seu ouvinte, o que auditivamente é identificado pelo aumento do tom de voz, em que temos, assim, um tom alto, ascendente. Por outro lado, naqueles passagens que julgamos não serem tão importantes, quase secundárias, abaixamos o tom. Nesse caso estamos diante de um tom baixo, descendente.

Diante dessa hipótese, nossa proposta, neste trabalho, é avaliar que tipos de gestos (movimentos corporais e expressões faciais) são realizados durante a ocorrência de tom alto e tom baixo identificados auditivamente.

Avaliamos gravações de entrevistas e palestras observando trechos com tom alto e com tom baixos e a configuração gestual no exato momento da realização desses tons e qual o efeito de sentido obtido pela combinação de tons e gestos.

Para sermos capazes de avaliar com detalhes os movimentos corporais e faciais realizados, usamos o modo lento, a uma taxa de 90% de redução da velocidade, de apresentação de vídeo.

Apresentaremos aqui passagens de uma palestra proferida por William Bonner, âncora do Jornal Nacional da Globo, para ilustrar um pouco das análises realizadas por nós.

Para darmos conta desta empreitada, na seção seguinte avaliaremos o conceito de gestos, seu papel de coexpressividade na ato comunicativo e a proposta de um sistema cognitivo unificado que agrega gestos e fala.

## **2. O Que é um Gesto?**

Ao desenvolver uma pesquisa sobre gestos, precisamos primeiro delimitar o que de fato será considerado gesto, tema já bastante debatido por Kendon (1972, 1975, entre outros trabalhos), McNeill (1992).

O tipo de gesto que de fato nos interessa, seguindo Kendon (1972, 1975, entre outros) e McNeill (1972) é aquele que não é pensado e não é produto de uma convenção social, como, por exemplo, o uso do polegar estendido para indicar que está tudo bem, encontrado em algumas culturas, inclusive a brasileira. Nem tampouco nos interessam os movimentos feitos instintivamente, como é caso de fechar os olhos quando estamos diante de uma luz forte.

Interessa-nos, sim, os gestos feitos inconscientemente e automaticamente durante uma conversa, aqueles que, de acordo com McNeill (2002), não parecem resultar de um esforço comunicativo consciente. Nesse sentido, o gesto é parte integrante da expressão global do falante, é um facilitador verbal. (Kendon, 2004)

De acordo com Kendon (2004), esses gestos inconscientes, a gesticulação, podem codificar aspectos da estrutura discursiva, contribuindo na organização dos padrões discursivos que são necessários no ato comunicativo. É, dessa forma, para o falante um facilitador da organização das estruturas conceituais.

O papel facilitador dos gestos no ato comunicativo pode ser verificado no fato de haver grande associação entre ocorrência de gestos e planejamento de pausa, bem como na associação de gestos e escolha lexical, sendo o gesto de grande importância na compreensão de enunciado (Kendon, 2004).

Para McNeill (1992), esses gestos espontâneos, idiossincráticos e imagéticos são representações das imagens mentais do falante, o que para Kita (2000) pode ser entendido como emergências de ações manipulatórias que uma pessoa traz de um quadro de uma palavra virtual.

Para a classificação de um gesto, os humanos dependem dos parâmetros auditivos, além dos visuais. A concomitância de estímulos auditivo e visual é tão necessária que sem os parâmetros auditivos, a performance de identificação de gestos cai significativamente (Eisenstein; Davis, 2004).

Krahmer e Swerts (2007) também mostram que a percepção de proeminências prosódicas é aumentada quando elas ocorrem acompanhadas de gestos, denominados por eles visual beats (batidas, ritmos visuais) <sup>1</sup>.

Para Butcher e Goldin-Meadow (2004), apesar de estarmos diante de duas modalidades comunicativas diferentes, gesto e fala formam um sistema de comunicação simples e unificado no adulto. Há, segundo elas, duas evidências que justificam a integração entre dois meios comunicativos: 1) coerência semântica e 2) sincronia temporal.

---

<sup>1</sup> É importante lembrar que na tipologia gestual proposta por McNeill (1992), há os gestos beats, justamente aqueles que co-ocorrem na fala ligados ao ritmo da fala. Pode ser, por exemplo, a batida da mão, do braço ou de um dedo, sempre com a expectativa de repetição.

A primeira evidência, a coerência semântica dos gestos, refere-se ao fato de os mesmos serem portadores de sentidos que estão relacionados com a fala que lhe acompanha. Gestos e fala são, portanto, semântica e pragmaticamente coexpressivos (Butcher; Goldin-Meadow, 2004).

A sincronia temporal constitui a segunda evidência da integração e unificação entre gesto e fala. É notório o fato de que os gestos são produzidos em sincronia com a fala. Os gestos não ocorrem antes ou depois do sinal sonoro. Ambos são produzidos ao mesmo tempo (Butcher; Goldin-Meadow, 2004).

Assim, gestos e fala fornecem uma mensagem coerente e convergente para o ouvinte, tratando-se, pois, de um sistema de comunicação unificado (Butcher; Goldin-Meadow, 2004); o que só é possível porque os gestos e a fala compartilham uma representação cognitiva comum. Constituem ambos o desenrolar comunicativo, fazendo ambas partes de uma mesma ideia (McNeill, 1992).

Somente um sistema simples e unificado que integra gestos e fala consegue explicar o montante vultoso encontrado por McNeill (1992) de 90% dos gestos produzidos por adulto ocorrerem ao mesmo tempo em que o gesticulador estava falando. Diante do resultado expressivo encontrado pelo pesquisador, ele afirma que atos de fala e gesticulação estão ligados um ao outro no tempo e em um nível mais geral, que para ele se trata do mesmo sistema cognitivo.

Diante de um cenário em que gestos e fala pertencem a um único e simplificado sistema cognitivo, o que dizer da relação entre gestos e os tons altos usados pelo falante para chamar atenção de seu ouvinte para certos trechos da fala considerados pelo falante como importantes ou o uso de tons baixos que indicam informações menos importantes? Há algum padrão gestual que pode ser majoritariamente atribuído a um e a outro tipo de tom?

A seção seguinte traz dados de fala espontânea, no sentido em que não foi gravada com o intuito de se investigar a relação gestos e tons com função de destaque. Investigamos se é possível estabelecer algum padrão entre esses dois meios comunicativos.

### **3. Gestos e Tom Alto e Baixo: Qual a Relação?**

#### **3.1. Tom Alto**

Para Crystal (1975) e Laver (1975), muitos fatos semânticos podem se manifestar por meio de certas configurações tonais. Dessa forma, é possível afirmar que um enunciado pode assumir sentidos diferentes a depender da forma como ele é dito.

O tom mais alto ao longo da fala tende a indicar ênfase, destaque. A ênfase a determinadas passagens é usualmente marcada pela subida do tom melódico se comprado com o restante no enunciado (Cagliari, 1999). Essa configuração tonal é seguida, como mostram as análises realizadas, pelo levantar das sobrancelhas e das mãos, conforme exemplificado na figura 1:



Fonte – <http://www.youtube.com/watch?v=tpSw5emgbHM>

Figura 1 – Como isso se aplica ao jornalismo da Globo.

A figura 1 e as demais aqui apresentadas foram extraídas de uma palestra proferida pelo jornalista William Bonner, editor e âncora do Jornal Nacional, telejornal brasileiro, transmitido pela Rede Globo de Televisão, no CONARH/ABRH, no ano de 2008, sobre “Talentos de Valor”. O jornalista começa sua fala dizendo que irá mostrar como o tema da palestra se aplica na emissora onde ele trabalha.

Ao dizer o trecho “Como isso se aplica ao jornalismo da Globo”, Bonner aumenta o seu tom melódico exatamente na palavra se aplica. Nesse enunciado, essa palavra é a que tem uma significação diferenciada, ela é de fato a palavra mais importante.

Ao aumentar o tom de se aplica, o jornalista chama a atenção de seus expectadores, indicando, que a partir daquele momento, ele está dando pistas importantes sobre o funcionamento da rede de televisão onde ele trabalha. Assim, ele está dizendo aos seus ouvintes: estejam atentos!

A ênfase na palavra se aplica não se restringe ao aumento do tom melódico. Ao realizar essa palavra de forma enfática, o jornalista *levanta a sobrancelha*, como indicado em 1, na figura 1, e *levanta as mãos*, como em 3. Além disso, há uma contração dos músculos faciais, como em 2, evidenciando certa força ao pronunciar a palavra. Nesse momento, o efeito prosódico é reforçado e a ênfase desejada pelo falante é amplificada. Aumenta-se, assim, a chance de a mensagem que se deseja transmitir alcançar os seus ouvintes.

Além de ser usada para dar destaque, importância a certos trechos da fala, a ênfase é também comumente usada pelo falante para fazer correções. Ao se corrigir uma palavra que foi dita erroneamente, o falante usa a ênfase para dizer a palavra que realmente deveria ter sido usada. Nessa situação também, a palavra correta é dita enfaticamente, com um tom melódico mais alto e vem igualmente acompanhada de movimentos, como é ilustrado nas figuras 2a e 2b:



Fonte – <http://www.youtube.com/watch?v=tpSw5emgbHM>

Figura 2a – Engenharia do jornalismo. Não.



Fonte – <http://www.youtube.com/watch?v=tpSw5emgbHM>

Figura 2b – Engenharia para o jornalismo.

Nas sequências 2a e 2b, o jornalista corrige o uso inadequado do jornalismo. Ao se dar conta do equívoco do uso da preposição do, ele refaz a frase acentuando engenharia e acentuando ainda mais a preposição para. Ao enfatizar as palavras engenharia e para, os movimentos faciais e corporais são realizados no mesmo instante em que ele aumenta o seu tom de voz para fazer as correções. A ênfase utilizada na correção acontece com *inclinar da cabeça para a direita* (1); *o levantar das sobrancelhas* (2); *abrir os olhos* (3); *levantar a mão direita* (4); *contrair dos músculos da face* (5); *levantar a mão direita e estender o polegar* (6).

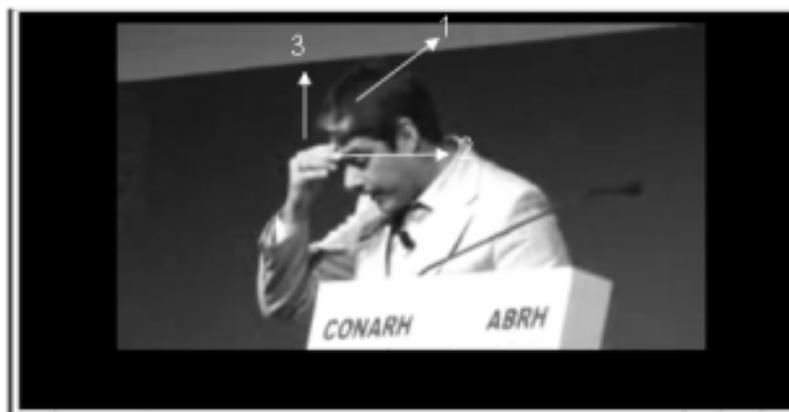
Como lembram Kramer e Sweters (2006), para indicar a palavra mais importante dentre aquelas que estão sendo faladas, os falantes usam diferentes meios para destacá-la e dentre essa diversidade de formas, incluem-se os parâmetros de acento de *pitch* e movimentos. Assim, a marcação de ênfase se faz a partir da articulação de vários parâmetros.

Nossas análises, aqui representadas por duas situações comunicativas, expressas nas figuras 1, 2a e 2b, evidenciam que a ênfase, normalmente marcada na fala pelo aumento do tom; pode ser também marcada por gestos ascendentes da cabeça e do corpo. Com os gestos ocorrendo simultaneamente, temos a sensação de que a ênfase é mais marcada, não abrindo margem para a possibilidade de não ser percebida por quem ouve.

### 3.2. Tom Baixo

Se por um lado, o falante tende a suspender o tom melódico para destacar determinados trechos de sua fala, por outro, ele tende a usar um tom melódico bem abaixo do normal quando tem por objetivo indicar trechos cujo conteúdo semântico “deve ser levado em consideração como algo ‘entre parênteses, ‘de menor importância’, ‘intercalado no meio de um outro trecho’, isto é, do assunto principal do discurso naquele momento” (Cagliari, 1999, p. 9).

E, à semelhança do aumento do tom melódico, o abaixamento também é acompanhado por gestos faciais e corporais característicos, como se pode verificar na figura 3:



Fonte – <http://www.youtube.com/watch?v=tpSw5emgbHM>

Figura 3 – A gente deixou Congonhas e.

Ao iniciar a história que havia anunciado, o jornalista *abaixa a cabeça* (1), *fecha os olhos* (2) e *ergue a mão até a testa* (3) ao dizer Congonhas (Figura 3), que foi dito de forma mais baixa se comparado com o restante do enunciado. Nesse contexto, em que ele narrava um grave acidente de avião, o nome do aeroporto era secundário. Assim, esse trecho foi realizado de forma mais atenuada, para que os trechos seguintes que continham de fato as informações mais importantes ficassem mais em evidência, mesmo sem qualquer aumento do tom melódico e realização de movimento corporal expressivo, como pode ser inferido pela figura 4.



Fonte – <http://www.youtube.com/watch?v=tpSw5emgbHM>

**Figura 4** – passou na Avenida Gutenberg, ao lado daquele lugar houve aquele acidente horrórico do ano passado.

Naquele contexto, a informação de onde ele tinha saído não era relevante, era de menor importância frente ao que ele queria de fato dizer: localizar uma avenida próxima de onde acontecera o acidente aéreo da Tam, aquele acidente horrórico, nas palavras do jornalista.

Abaixar o tom, abaixar a cabeça e fechar os olhos no trecho Congonhas foi uma pista para que o seu ouvinte soubesse que tão logo ele recuperasse a postura e o tom melódico, uma informação importante seria pronunciada que recebeu destaque por estar próxima a um enunciado dito com um tom mais abaixo e com movimentos que lhe são típicos.

O tom baixo sob uma passagem pode também indicar uma hesitação decorrente de uma incerteza, de uma dúvida acerca da informação que está sendo passada. Nesses casos, o tom baixo também coocorre com movimentos corporais, como pode ser observado na figura 5.



Fonte – <http://www.youtube.com/watch?v=tpSw5emgbHM>

**Figura 5** – Eram três ou três e meia da tarde.

A incerteza acerca do horário do acontecimento do evento narrado leva o palestrante a hesitar entre duas informações: três ou três e meia da tarde. Nesse trecho de sua fala, ele diminui seu tom e o mesmo é acompanhado pelo *fechar dos olhos* (1) e o *levantar da mão direita* (2), além de ser visível o relaxamento da musculatura facial (Figura 6).

Nossos dados mostram que, enquanto falamos, as caras e bocas que fazemos tendem a indicar os tons altos e baixos que estão ocorrendo na fala. Há sim sincronia entre as informações visuais e auditivas. Os movimentos não são aleatórios e caóticos. Esse é, pois, o motivo de a Lu e a Eva Byte não convencerem sobre aquilo que elas falam.

## 4. Conclusões

Frente à proposta de pesquisa aqui apresentada, a saber, avaliar a relação entre tom alto e tom baixo, gestos e seus respectivos sentidos, podemos concluir que, ao menos para o Português Brasileiro, os tons alto e baixo podem vir acompanhados por gestos faciais e corporais que tornam, respectivamente, a ênfase e atenuação mais marcadas.

Frente às análises realizadas do vídeo apresentado e de outros não apresentados aqui, podemos afirmar que o aumento do tom alto tende a ser acompanhado por gestos de elevação da cabeça, maior abertura dos olhos, maior contração dos músculos faciais. Já o tom baixo tende a ser acompanhado pelo abaixamento da cabeça, olhos fechados e semicerrados e maior relaxamento dos músculos faciais.

Enfim, podemos generalizar dizendo que tom alto, normalmente usado para indicar ênfase, é mais audível, exige mais esforço de produção, requer gestos que indiquem mais energia. Por sua vez o tom baixo, normalmente usado para indicar atenuação, é menos audível, exige menos esforço de produção, é acompanhado por gestos que indicam menos energia.

Podemos afirmar, portanto, que tom e gestos têm os mesmos significados, suas mensagens são convergentes, portanto. Assim, podemos afirmar que os tons alto e baixo e os gestos que os acompanham compõem a mesma ideia, o que o reforça a hipótese de McNeill (1992) de que a gesticulação e a fala compõem um sistema único.

## Agradecimentos

Agradecemos ao CNPq o suporte financeiro para a execução desta pesquisa.

## Referências Bibliográficas

- BUTCHER, C.; GOLDIN-MEADOW, S. Gesture and transition from one-to-two-words speech: when hand and mouth come together. MCNEILL, D. (Ed.) **Lang. and gesture**. Chicago: U Chicago, p. 235-257, 2000.
- CAGLIARI, L.C. **Acento em Português**. Edição 1. Edição do Autor, Campinas, 1999.
- CRYSTAL, D. **The English Tone of Voice**. London: Edward Arnold, 1975.
- EISENSTEIN, J.; DAVIS, R. Visual and Ling. information in gesture classification. In: Proceedings of the INT. CONF ON MULTIMODAL INTERFACES (ICMI'04). USA, p. 113-120, 2004.
- EKMAN, P. Universal and cultural differences in facial expression of emotion. COLE, J. (Ed.) **Nebrascka Symp of Motivation**, p. 207-283, 1972.
- KENDON, A. **Gesture: Visible Action as Utterance**, Cambridge, UK: Cambridge UP, 2004.
- \_\_\_\_\_. Gesticulation, speech and the gesture theory of Lang. origins. **Sign Lang Studies**, p 349-373, 1975.
- \_\_\_\_\_. Some relationships between body motion and speech. In: SEIGMAN, A.; POPE, B. (Eds.) **Studies in Dyadic Com**. Elmsford, NY: Perga P, p.177-216, 1972.
- KITA, S. How representational gestures help speaking. McNeill, D. (Ed.), **Gesture and Lang**, Cambridge: Cambridge UP, p. 162-185, 2000.
- KRAHMER, E.; SWERTS, M. Hearing and seeing beats. The influence of visual beats on the production and perception of prominence. In: Proceedings of the CONF SPEECH PROSODY. Dresden: 2006.
- LAVIER, J. 1975. **Individual Features in Voice Quality**. Edinburgh: University Edinburgh, 1975.
- MCGURK, H.; MACDONALD, J. Hearing lips and seeing voices. **Nature**, n. 264, p. 746-748, 1976.
- MCNEILL, D. Gesture and Lang dialectic. **Acta Ling.a Hafniensia: Int J of Ling**, v. 34, n. 1, p. 7-37,

2002.

MCNEILL, D. **Hand and mind: what gestures reveals about thought**. Chicago, London: U of Chicago P, 1992.

MERLEAU-PONTY, M. **Fenomenologia da percepção**. Tradução Carlos Alberto Ribeiro de Moura. São Paulo: Martins Fontes, 1999.

MUNHALL, K.G. et al. Visual Prosody and Speech Intelligibility: Head Movement Improves Auditory Speech Perception. **Psych Sci**, v. 15, n. 2, p. 133-137, 2004.

SUMBY W.H.; POLLACK, I. Visual contribution to speech intelligibility in noise. **J Acous Soc Am**, v. 26, n. 2, p 212-215, 1954.

# *Os Papéis das Prosódias Vocal e Visual na Expressão de Emoções na Fala*

MARIO AUGUSTO DE SOUZA FONTES

## 1. Introdução

Na fala, a comunicação envolve a transmissão de informações, mas também tem o papel de impressionar o ouvinte (Bolinger, 1986). O ser humano tem várias formas de se comunicar, sendo as prosódias vocal e facial dois modos que se complementam na expressão de sentidos, entre eles, os relacionados à expressão de emoções (De Gelder e Vroomen, 2000) e que se integram no processamento perceptivo das características prosódicas como argumentam Borràs-Comes e Prieto (2011) e Martínez, Rojas e Suárez (2012).

A prosódia vocal compreende a entoação, a acentuação, a taxa de elocução, as pausas, a qualidade de voz e o ritmo. Esses elementos interagem com os segmentos vocálicos e consonantais e com eles compõem a fala. A prosódia visual refere-se aos gestos corporais, tais como movimentos de cabeça, de olhos, de sobrancelhas, de lábios, de braços e de mãos, que acompanham a fala.

Entre as várias funções informativas, discursivas e expressivas das prosódias vocal e visual, destacam-se a veiculação de emoções. As emoções são veiculadas na fala como produto dos gestos vocais e visuais que impressionam o ouvinte, suscitando efeitos de sentido (Madureira, 2011). Desse modo são construídas as metáforas sonoras que Fonagy (2000) define como “sensações proprioceptivas, táteis e motoras que acompanham as produções dos sons”. As expressões de emoções na fala podem, segundo essa perspectiva da metaforicidade, por associação, representar as gestualidades que as derivaram. As metáforas são dinâmicas e podem ser expressas por gestos (Cienki e Müller, 2008) e as associações que se estabelecem são procedimentos cognitivos de compreensão de alguma coisa em termos de outra e remetem a possíveis leituras que podem ser interpretadas como uma extensa cadeia inferencial em que “efeitos emotivos podem ser inferidos a partir de efeitos físicos, o que é coerente com a natureza humana” (Zanotto, 2010:640).

Os papéis das prosódias vocal e visual na expressão de emoções na fala são objeto deste capítulo que aborda trabalhos relevantes da literatura concernida com aspectos fonéticos e visuais relacionados à expressão de emoções.

## 2. Prosódia vocal e expressão de emoções

Os elementos prosódicos, ou seja, a entoação, o ritmo, a acentuação, a qualidade de voz, a taxa de elocução e as pausas, são apontadas na literatura como relevantes para a expressão de emoções (Johnstone e Scherer, 2000; Scherer, Johnstone e Klasmeyer 2003; Patel et al, 2006; Viola, 2006; Martínez e Rojas, 2011).

Do ponto de vista fonético-acústico, a prosódia vocal pode ser analisada de acordo com os parâmetros acústicos de duração, frequência fundamental e intensidade. A entoação, ou melodia da fala, tem como parâmetro acústico principal a  $f_0$  (frequência fundamental) e pode ser medida em Hz, correspondendo, em nível fisiológico, ao número de vibrações das pregas vocais por segundo e, em nível auditivo, à sensação de altura ou *pitch*. A literatura sobre emoções e características acústicas aponta forte correlação entre frequência fundamental alta e alegria, e frequência fundamental baixa e tristeza (Scherer, 1995). Diferenças de nível de intensidade também foram apontadas por Audier, Aubergé e Rilliard (2008) como parcialmente correlacionadas com o nível de percepção da emoção e com a discriminação entre enunciados de fala espontânea e atuada.

As variações de taxa de elocução de rápida à lenta, que afetam tanto a duração quanto a qualidade dos segmentos vocálicos e consonantais, também são correlacionadas à expressão de diversas emoções como raiva, felicidade e tristeza (Rodrigues, 2007; Yildirim et al, 2004).

O papel expressivo da pausa silenciosa (Santos, 2010), elemento prosódico correlacionado à duração, é referido comumente como “pausa dramática” (Bolinger, 1978) por seu efeito perturba-

dor à continuidade do fluxo de fala.

Os elementos prosódicos podem ser utilizados de modo a expressar um contínuo de funções e efeitos que vão do extralinguístico ao linguístico, passando pelo paralinguístico, que é o nível que compreende a expressão vocal de emoções (Laver, 1994).

## **2.1. Qualidade de voz enquanto elemento da prosódia vocal e seu papel na Construção da expressão das emoções**

### **2.1.1. Conceito de qualidade de voz**

Embora menos explorada nos estudos fonéticos devido à ausência de modelos descritos de natureza fonética até o advento do trabalho de Laver (1980), a qualidade de voz desempenha papel muito relevante na expressão de emoções, e seus ajustes são comumente explorados por atores para a construção de personagens e seus estados emotivos.

O termo “qualidade de voz” é conceituado de acordo com Laver (1979, 1980, 2000). Para analisar as qualidades de voz, Laver utilizou de uma unidade analítica chamada *setting* (ajuste) que corresponde aos ajustes fonatórios e articulatórios que acompanham dois ou mais segmentos de fala, enunciados ou, mais extensivamente, a fala de um indivíduo.

Para a descrição dos ajustes de qualidade de voz, Laver (1980) propõe um modelo fonético descritivo que tem como unidade básica o ajuste neutro. Esse ajuste neutro tem como base a ideia de que o trato vocal não é modificado pela ação da musculatura, ou seja, parte de um ponto de repouso.

Todos os ajustes propostos por Laver são delineados a partir desse ajuste neutro que se caracteriza pelos seguintes aspectos: não envolve nem tensão nem relaxamento do trato vocal; os lábios não se projetam a não ser para caracterizar sons que intrinsecamente envolvam a labialização; não há nasalidade, a não ser na pronúncia de segmentos nasais; a laringe não se encontra nem levantada nem abaixada; o diâmetro do trato vocal supralaríngeo mantém-se praticamente uniforme ao longo do seu comprimento; as articulações orais anteriores são realizadas pela lâmina da língua, pois a raiz da língua não está nem avançada nem recuada; os músculos da faringe formam constrições no trato vocal; a mandíbula não está acentuadamente aberta ou fechada, a vibração das pregas vocais é regularmente periódica, sem fricção audível.

São considerados três tipos de ajustes no modelo fonético: supralaríngeos, laríngeos e de tensão. Os ajustes supralaríngeos englobam mobilizações dos órgãos do trato vocal, dos lábios, da língua, da mandíbula, do véu palatino e da laringe, modificando as configurações das cavidades oral, nasofaríngea, orofaríngea e laringofaríngea. Os ajustes de tensão concernem ao trato vocal e ao laríngeo. Os ajustes fonatórios compreendem ajustes de extensão, aproximação e contato realizados pelas pregas vocais e com variados modos de fonação.

No módulo fonético descritivo de qualidade de voz proposto por Laver (1980), dois princípios são postulados: o da susceptibilidade e o da compatibilidade. Eles são relevantes ao serem considerados na identificação dos ajustes.

Pelo princípio da susceptibilidade, postula-se que alguns segmentos fônicos são mais susceptíveis do que outros a determinados ajustes. Dessa maneira, os segmentos fônicos orais são mais susceptíveis ao ajuste de nasalidade do que os nasais, ou seja, a característica do ajuste da nasalidade será mais facilmente percebida se os segmentos fônicos forem orais.

A relação entre ajustes e segmentos deve ser observada para fins de identificação dos ajustes, pois, quando os ajustes têm atributos não comuns ao segmento, esses passam a ser mais susceptíveis à sua influência.

A análise acústica das qualidades de voz pode ser verificada por meio de variadas medidas de curto termo, tais como frequência de formantes e a relação entre harmônicos e o LTAS, medida

de longo termo pela qual se verifica a variação da intensidade ao longo do eixo da frequência.

Baseado no modelo fonético descritivo de qualidades de voz, Laver (1979, 1980, 2000) elaborou, para fins de avaliação das qualidades de voz, um roteiro denominado Voice Profile Analysis Scheme (VPAS), cuja versão mais recente data de 2007 (Laver e Mackenzie-Beck, 2007). Essa versão do roteiro foi traduzida e adaptada para o português brasileiro por Camargo e Madureira (2008).

Essa versão adaptada do roteiro foi aplicada em treinamento de juízes, e os resultados da aplicação foram relatados em Camargo e Madureira (2008). Para o treinamento, foi construído um corpus com frases que continham repetições de segmentos fônicos-chave (sons consonantais e vocálicos da língua portuguesa), de modo a facilitar a identificação dos ajustes de qualidade de voz. Dessa forma, frases constituídas somente por sons orais, por exemplo, tinha o objetivo de treinar a percepção de ajustes de qualidade de voz nasal; frases com vogais anteriores frequentes, para treinar a percepção de ajustes de corpo de língua recuada; e frases com sons alveolares, para treinar a percepção de vários ajustes de língua.

A versão do roteiro de Camargo e Madureira (2008) também foi aplicada em vários trabalhos desenvolvidos no Laboratório Integrado de Análise Acústica e Cognição (LIAAC), da Pontifícia Universidade Católica de São Paulo, entre eles: Pessoa (2012); Camargo, Rusilo e Madureira (2011); Madureira e Camargo (2010) e Madureira (2011) constituindo uma linha de pesquisa original sobre a investigação da qualidade de voz por considerar os ajustes articulatórios além dos fonatórios e de tensão comumente apreciados na literatura sobre qualidade de voz e relacioná-los a parâmetros acústicos robustos (Barbosa, 2009).

### **2.1.2. Qualidade de voz e expressão de emoções**

A qualidade de voz é um elemento prosódico de relevância paralinguística (Laver, 1980), aspecto que engloba a expressão de emoções. Goble e Chasaide (2003) investigam, a partir de estímulos originados por aplicação de procedimentos de síntese de fala, sete tipos de qualidades de voz (modal, harsh voice, tense voice, breathy voice, whispery voice, creaky voice e lax-creaky voice) e os efeitos impressionáveis desses estímulos. Esses tipos de qualidade de voz corresponderiam a ajustes fonatórios e de tensão laríngea descritos no VPAS-PB. Para avaliar os efeitos impressionáveis da qualidade de voz, conduzem um experimento perceptivo em que os juízes têm de avaliar as impressões causadas pelos diferentes tipos de qualidade de voz. Os descritores utilizados foram: relaxado/estressado, contente/zangado, amigável/hostil, triste/feliz, desinteressado/interessado, formal/informal, tímido/confiante e medroso/destemido. Como resultados, apontam que as qualidades de voz sinalizaram melhor emoções mais sutis do que emoções fortes, com exceção da raiva, e que não há biunivocidade entre qualidade de voz e tipo de afeto.

Johnston e Scherer (1999) investigaram a expressão de emoções em frase emitida por um ator com variadas emoções, realizando um experimento com técnicas de análise acústica e fisiológica (eletroglotografia, respirometria, eletrocardiograma e eletromiografia de superfície). Os resultados apontam valores mais baixos de frequência fundamental ( $f_0$ ) na expressão de emoções como tristeza, depressão, tédio e desapontamento, e mais altos para a alegria e raiva explosiva (hot anger). A extensão de  $f_0$  revelou-se maior na expressão de alegria e menor na expressão de estados afetivos negativos, como depressão, tensão, tédio, ansiedade e irritação. O que também é compatível com os de outros estudos: Kaiser (1962), Van Bezooijen (1984), Scherer (1986), Banse e Scherer (1996) e Laukka (2007), e complementar aos achados de Kienast e Sendlmeier (1999) em relação à expressão de tristeza e tédio sobre a existência de redução de segmentos, taxa de articulação lenta e diminuição nos valores de frequência de formantes.

Os valores de jitter (índice de instabilidade na forma da onda sonora glotal, medido na va-

riação ciclo a ciclo da frequência fundamental) correlacionaram-se com valores de  $f_0$ , mostrando-se mais altos com a elevação de  $f_0$ . Essa tendência não foi, todavia, verificada na expressão de ansiedade e tensão. Em relação ao parâmetro de jitter em análise de fala emotiva Nunes (2009) aponta que valores mais altos de jitter relacionaram-se com a expressão de desespero, medo raiva e tristeza e os mais baixos à alegria.

A análise do tempo de adução e abdução das pregas vocais mostrou que as emoções tiveram impacto no tempo de fechamento (adução), mas não no de abertura (abdução). Para as emoções com alto grau de ativação do organismo, caracterizadas por  $f_0$  alto e energia espectral alta, a glote fecha-se mais rapidamente. Isso revela esforço vocal maior ou tensão muscular laríngea.

Na expressão de irritação e felicidade, foi encontrada pouca energia espectral nas frequências baixas. Na expressão de tensão e ansiedade, entretanto, não foi verificada redução tão grande em baixas frequências, o que demanda, segundo os autores, a investigação de outros fatores além do esforço vocal.

A expressão de alegria foi caracterizada por maior variação de  $f_0$  e maior concentração de energia em frequências altas, mas menor quociente de fechamento (adução das pregas vocais). O parâmetro de  $f_0$  tem se mostrado muito relevante na análise de expressão de emoções, sendo um dos parâmetros mais investigados.

Em termos de número (12) e relevância (medidas de  $f_0$ , intensidade e duração) de parâmetros acústicos analisados, o trabalho de Barbosa (2009) traz uma importante contribuição para a análise da expressividade oral em geral e a investigação de emoções na fala.

Nesse trabalho, a expressividade de fala de entrevistados em um programa de rádio é analisada por um método que detecta variação na expressividade da fala por meio da conjugação de três tipos de análise: uma Análise Dimensional, uma Análise de Componentes Principais e uma Análise de Regressão Múltipla.

O corpus constituído de 206 enunciados foi avaliado em termos de 4 dimensões expressivas por meio de teste perceptivo aplicado a 12 juízes. Três das dimensões concerniam os primitivos emocionais de ativação (relaxado-agitado), valência (agradável-não agradável) e controle (controlado-sem controle) e uma quarta dimensão relacionava-se ao envolvimento (envolvido-não envolvido).

Entre os resultados, foi verificado que o aumento do esforço vocal (valores mais baixos de inclinação espectral) diminui na fala empática, enquanto valores altos de  $f_0$  (representados pelo quantil 0,995) aumentam a empatia.

A análise de PCA explicou 97% dos dados e a Análise de Regressão Múltipla revelou-se capaz de prever os valores de duas dimensões derivadas do PCA (a prontidão e a empatia) a partir dos parâmetros acústicos obtidos automaticamente do conjunto de enunciados.

Os parâmetros propostos por Barbosa (2009) podem ser utilizados para a análise de fala emotiva. Podem ser integrados nessa análise os ajustes de qualidade vocal e aspectos da dinâmica vocal descritos pelo VPAS, a expressão de valência e emoções e aspectos visuais. Com essa integração pode-se contribuir para a reflexão sobre a função paralinguística da fala e sua relevância comunicativa.

## **2.2. Prosódia Visual e seu Papel na Construção da Expressão das Emoções**

Os gestos visuais, como as movimentações dos olhos, da cabeça, da sobrancelha e dos lábios, estão correlacionados à expressão de emoções. Trabalhos como os de Ekman (1972) trazem evidências sobre as correlações entre os gestos visuais e as emoções.

Ekman (1979, 1992, 1993, 1997, 1998, 1999) defende que as emoções são expressões involuntárias que foram selecionadas e refinadas na evolução das espécies para fins comunicati-

vos e, devido sua carga informativa, desempenham papel muito relevante na interação social. Na expressão de emoções, Ekman considera serem a face e a voz os dois aspectos mais importantes da expressão de emoções. Para a análise da expressão de emoções, Ekman e Friesen (1978) desenvolveram o Facial Action Coding Systems (FACS), um roteiro para analisar os movimentos da face. Esse roteiro foi revisado por esses mesmos autores décadas mais tarde (Ekman e Friesen, 2002).

### 2.2.3. O Protocolo do FACS - Facial Action Code System

A unidade analítica do protocolo do FACS são as AUs, unidades de ação que descrevem os movimentos dos músculos da face. Cada uma das 64 AUs está associada à movimentação de um dos músculos que estão presentes na face.

A avaliação da movimentação das AUs sempre é tomada com relação a uma imagem da face neutra da pessoa que está sendo analisado a movimentação facial. Na Figura 1, podemos observar 15 das 64 AUs utilizadas por Ekman associadas a movimentação do músculo da face associada ao ponto da face que está sendo observado o deslocamento.

Lower Face Action Units					
AU9	AU10	AU11	AU12	AU13	AU14
					
Nose Wrinkler	Upper Lip Raiser	Nasolabial Deepener	Lip Corner Puller	Cheek Puffer	Dimpler
AU15	AU16	AU17	AU18	AU20	AU22
					
Lip Corner Depressor	Lower Lip Depressor	Chin Raiser	Lip Puckerer	Lip Stretcher	Lip Funneler
AU23	AU24	*AU25	*AU26	*AU27	AU28
					
Lip Tightener	Lip Pressor	Lips Parts	Jaw Drop	Mouth Stretch	Lip Suck

Figura 1 – Extraído de Khadoudja Ghanem, Alice Caplier, M.k. Kholadi (2012)

O deslocamento do ponto analisado é quantificado em uma escala de intensidade que varia de A até E, sendo a intensidade A correspondente a marcação de traço na movimentação da AU e a intensidade E representa o extremo máximo na variação da AU. A escala de variação não tem suas intensidades de forma linear, demandando do avaliador um treinamento exaustivo para a familiarização e aprendizado na quantificação da classificação das AUs.

Ekman salienta em seu manual de treinamento do FACS a necessidade do aprendiz do protocolo investir no mínimo 100 horas de atividades de estudo. Recomenda que essas atividades, as quais são propostas para habilitar o aprendiz a prestar os exames de certificação do referido protocolo manual, não tenham duração maior do que 3 a 4 horas diárias.

Atualmente temos além do processo manual do protocolo do FACS a possibilidade de utilizar softwares, utilizando algoritmos de reconhecimento de padrões fazer a identificação das AUs de forma automática. Esses ambientes de análise permitem a análise em tempo real de um vídeo de 30 frames por segundo, possibilitando o estudo dinâmico das emoções da face por meio do protocolo do FACS.

O grande diferencial da análise automática é a possibilidade de realização de estudos da evolução dinâmica das emoções. Enquanto uma análise manual do FACS por um especialista cer-

tificado demanda para a análise de 1 minuto de vídeo duas horas de análise a análise automática permite fazer estudos em tempo real, viabilizando a compreensão das alterações que se verificam na maneira de se expressar ao longo do tempo. Acoplado esse tipo de análise à análise dinâmica acústica, é possível fazer correlações entre a linguagem facial e a linguagem oral.

### 2.2.2. Análise das emoções

A co-ocorrência de AUs permitem identificar as emoções. O quadro 1, a seguir, representa a relação entre as emoções básicas e as AUs que são identificadas pelo FACS na face de uma pessoa que está sob essa emoção.

**Quadro 1** – Extraído de Freitas Magalhães (2011)

Emoções	AUs
Alegria	6+12+25
Tristeza	1+4+15
Côlera	4+5+7+10+26
Medo	1+2+4+5+20+25
Surpresa	1+2+5+25 ou 26
Aversão	7+10+25

Kohler et al (2004), utilizando as FACS em estudo sobre quatro emoções (alegria, tristeza, raiva e medo), confrontaram os aspectos visuais relacionados à expressão de alegria com os relacionados à expressão de tristeza, e apontaram como relevantes os gestos de sobrancelha, lábio e bochechas. A alegria compreendeu a elevação das sobrancelhas, das bochechas, do lábio superior e das comissuras labiais; a tristeza, sobrancelhas franzidas, lábios tensionados, comissuras labiais inclinadas para baixo e extensão de lábios diminuída; a raiva, sobrancelhas abaixadas, olhos bem abertos, pálpebra superior levantada, pálpebra inferior comprimida, lábios entreabertos, dentes à mostra e comissuras labiais inclinadas para baixo; e o medo, olhos esbugalhados, sobrancelhas vincadas e elevadas e lábios esticados, elevação dos lábios e dilatação das narinas.

As investigações sobre os movimentos da face demonstraram a multifuncionalidade desses movimentos, pois, além de revelarem emoções, são também utilizados como ilustradores (movimentos que acompanham a fala, exercendo variadas funções). Entre os ilustradores, encontramos os batons, uma das subcategorias de ilustradores que compreendem os movimentos utilizados para enfatizar ou acentuar palavras e outras unidades linguísticas. Os ilustradores e as outras categorias (emblemas, manipuladores e reguladores) propostas por Ekman e Friesen (1969) basearam-se na pesquisa de Efran (1968) e foram largamente aplicadas na análise do discurso oral e na análise da conversação. Cotes (2000) baseou-se nas categorias propostas por Ekman e Friesen (1969) para analisar a locução em telejornais.

Ekman (1999a) ressalta que as expressões de emoção analisadas em frases isoladas dão informações tanto sobre o contexto que as antecede quanto sobre o que as sucede, e que a expressão de emoções a partir do julgamento das expressões faciais é mais robusta em certas emoções do que em outras: raiva, medo, nojo, desprezo, surpresa, desgosto e alegria são mais facilmente reconhecidas por meio de julgamentos dos movimentos faciais do que outras emoções.

Resultados que corroboram a robustez dos julgamentos de expressões faciais de alegria encontram-se em um experimento desenvolvido por Swertz et al (2012), no qual juízes avaliaram expressões de emoção deficientes visuais e por pessoas com visão normal em três modos de apresentação: áudio, imagem (imagem de vídeo sem som) e vídeo (áudio mais imagem). A expressão

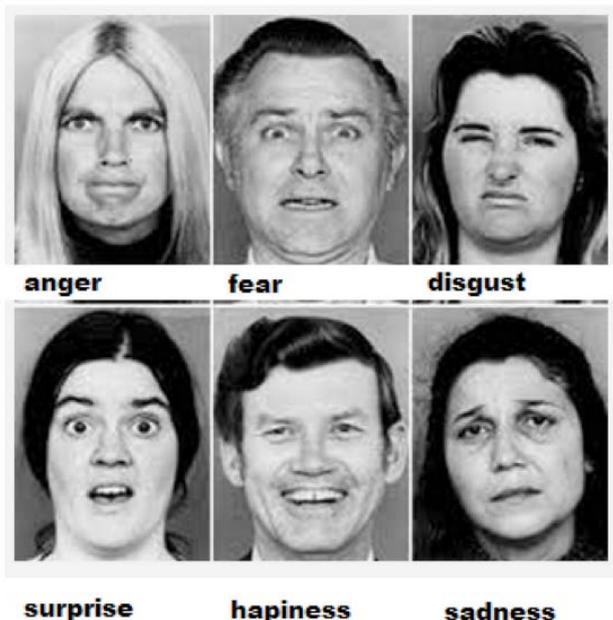
de alegria foi bem avaliada pelos dois grupos, enquanto que as expressões de tristeza e de medo foram muito confundidas. Esse estudo também revelou que as expressões de emoções pelos deficientes visuais foi melhor percebida pela expressão vocal do que pela visual.

A confusão na identificação de tristeza e medo, ambas emoções negativas, podem ser ponderadas a partir do apontamento de Ekman (1992) sobre a existência de expressões faciais comuns em um grupo de expressões de emoções positivas, tais como a alegria, a felicidade e o alívio, e em grupo de emoções negativas, entre outras, a tristeza, a culpa, a vergonha e o desgosto.

Trabalhos sobre as relações entre movimentos faciais e expressão de emoção como os de Darwin (1872/1965), Izard (1978), Ekman (1972,1973), Boucher e Ekman (1975), Ekman e Friesen (1978), Ekman, Roper e Hagen (1980), Ekman, Friesen e Ancoli (1980), Reeve (1993), Ekman e Friesen (2003), Silvia (2008), Scherer e Ellgring (2007), Mortillaro, Mehu e Scherer (2011) e Jack et al (2014) têm revelado tendências de pareamentos entre aspectos visuais e categoria de emoção.

Entre os pareamentos mencionados, destacamos: felicidade/alegria (happiness/joy): pálpebra inferior levantada, cantos dos lábios inclinados para cima e abertura de boca; raiva (anger): olhos saltados, testa franzida, pálpebras superiores e inferiores levantadas, abertura de boca, sobrancelhas abaixadas e lábios pressionados firmemente; surpresa (surprise): sobrancelhas levantadas em arco, olhos bem abertos, pálpebra superior levantada, abertura de boca e queixo abaixado; nojo (disgust): testa franzida e pálpebra inferior levantada; medo (fear): pálpebra superior e sobrancelha levantadas, testa franzida, olhos abertos e boca ligeiramente aberta; tristeza (sadness): testa franzida, sobrancelha levantada e cantos dos lábios inclinados para baixo.

A seguir, na Figura 2, fotos de expressões de seis emoções básicas descritas por Ekman (1978) e Ekman e Friesen (1978, 2003).



**Figura 2** – Expressões de seis emoções básicas, extraídas de scienblog.com

A expressão real das emoções, segundo Ekman (1999b), diferencia-se das expressões atuadas, por exemplo a encenação de atores, as quais são definidas em relação a movimentos faciais que se referem a emoções não sentidas pelas pessoas que as expressam. A expressão atuada de uma emoção assemelha-se à expressão real dessa emoção e a diferença entre elas pode não ser percebida facilmente.

As emoções reais diferenciam-se das emoções atuadas por serem ativadas pelo sistema límbico produzindo uma ação de movimento muscular facial que independe da vontade da pessoa, de maneira que a face passa a apresentar todas as variações de emoções que sentimos ou vivemos.

ciamos.

O estudo da dinâmica das emoções reais medidas pelos gestos vocais e faciais permitem a compreensão da funcionalidade da anatomia funcional neurológica na área das amídalas, que tem o mesmo nervo condutor para o acionamento da musculatura da face e do aparelho fonador. Também trazem subsídios sobre como se dá a integração ou a complementaridade entre os gestos.

Concluimos, a partir da literatura visitada, que investigar a expressão de emoções, correlacionando os aspectos visuais e vocais abre uma perspectiva de alta relevância comunicativa e reputamos serem tais correlações potencialmente facultadas e viabilizadas pela aplicação de análise estatística multidimensional.

## **Agradecimentos**

Nossos agradecimentos à CAPES, pelo apoio referente ao processo BEX 0770/15-9 da CAPES.

## **Referências Bibliográficas**

AUDIBERT, N.; AUBERGÉ, V.; RILLIARD, A. Acted vs. spontaneous expressive speech: perception with inter-individual variability. **Proceedings of the 2nd International Workshop on Corpora for Research on Emotion and Affect**, Marrakech: Morocco, 23-27, 2008.

BANSE, R.; SCHERER, K. R. Acoustic profiles in vocal emotion expression. **Journal of Personality and Social Psychology**, 70:614-636, 1996.

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

BOLINGER, D. L. Intonation across languages. In: GREENBERG, J. H.; FERGUSON, C. A.; MORAVCSIK, E. A. (eds.) **Universals of Human Language Phonology**. Stanford, CA: Stanford University Press, 2:471-524, 1978.

BOUCHER, J. D.; EKMAN, P. Facial areas and emotional information. **Journal of Communication**, 25(2):21-29, 1975.

CAMARGO, Z. A.; MADUREIRA, S. Voice quality analysis from a phonetic perspective: Voice Profile Analysis Scheme Profile for Brazilian Portuguese (BP-VPAS). In: **Proceedings of the 4th Conference on Speech Prosody**. Campinas, Brazil, 1:57-60, 2008.

\_\_\_\_\_. The acoustic analysis of speech samples designed for the Voice Profile Analysis Scheme for Brazilian Portuguese (BP-VPAS): long-term f0 and intensity measures. In: BOTINIS, A. (ed.) **Proceedings of the third ISCA Tutorial and Research Workshop on Experimental Linguistics**. Athens, 1:33-36, 2010.

CAMARGO, Z. A.; RUSILO, L. C.; MADUREIRA, S. Evaluating speech samples designed for the Voice Profile Analysis Scheme for Brazilian Portuguese. **Proceedings of the Fourth ISCA Tutorial and Research Workshop on Experimental Linguistics**. Paris, p. 55-58, 2011.

CIENKI, A.; MULLER, C. Metaphor, gesture and thought. In: GIBBS JR., R. W. (ed.) **The Cambridge Handbook of Metaphor and Thought**. Cambridge: Cambridge University Press, p. 483-502, 2008.

COTES, C. S. G. **Apresentadores de telejornal: análise descritiva dos recursos não-verbais e vocais durante o relato do jornal na televisão**. Tese de Doutorado. São Paulo, PUC-SP, 2000.

DARWIN, C. **The expression of the emotions in man and animals**. Chicago University of Chicago Press, 1872/1965.

EFRAN, J. S. Looking for approval: effects on visual behavior of approbation from persons differing in importance. **Journal of Personality and Social Psychology**, 10 (1), 21-25, 1968.

EKMAN, P. Universals and cultural differences in facial expressions of emotion. In: COLE, J. K. (ed.)

**Nebraska symposium on motivation.** Lincoln: University of Nebraska Press. p. 207-282, 1972.

\_\_\_\_\_. Cross-cultural studies of facial expression. In: EKMAN, P. (ed.), **Darwin and facial expression: A century of research in review.** New York: Academic. p. 169-222, 1973.

\_\_\_\_\_. About Brows-Emotional And Conversational Signals in Human Ethology. In: VON CRANACH, M.; FOPPA, K.; LEPENIES, W.; PLOOG, D. (eds.) **Human ethology.** Cambridge: Cambridge University Press. p. 169- 248, 1979.

\_\_\_\_\_. An argument for basic emotion. **Cognition and Emotion.** 6:169-200, 1992.

\_\_\_\_\_. Facial expressions of emotion: new findings, new questions. **Psychological Science.** 3:34-38, 1992.

\_\_\_\_\_. Facial expression of emotion. **American Psychologist.** 48:384-92, 1993.

\_\_\_\_\_. All Emotions are basic. In: EKMAN, P; DAVIDSON, R. (ed.) **The nature of emotion: fundamental questions.** New York: Oxford University, 1994.

\_\_\_\_\_. What we have learned by measuring facial behavior. In: EKMAN, P. & ROSENBERG, E. L. (eds.) **What the face reveals: basic and applied studies of spontaneous expression using the facial action coding system (FACS).** New York: Oxford University Press. p. 469-485, 1997.

\_\_\_\_\_. Universality of emotional expression? A personal History of The Dispute. In: EKMAN, P. (ed.) **Third edition of Charles Darwin the expression of the emotions in man and animals, with introduction, afterwords, and commentaries.** London: Harper Collins. p. 363-393, 1998.

\_\_\_\_\_. Basic emotions. In: DALGLEISH, T.; POWER, T. (eds.) **The Handbook of Cognition and Emotion.** Sussex, UK: John Wiley & Sons. p. 45-60, 1999a.

\_\_\_\_\_. Facial expressions. In: DALGLEISH, T.; POWER, T. (eds.) **The Handbook of Cognition and Emotion.** Sussex, UK: John Wiley & Sons. p. 301-320, 1999b.

EKMAN, P.; FRIESEN, W. V. The repertoire of nonverbal behavior: categories, origins, usage, and coding. **Semiotica,** 1:49-98, 1969.

\_\_\_\_\_. **Facial Action Coding System: A Technique for the Measurement of Facial Movement.** Consulting Psychologists Press: Palo Alto, 1978.

\_\_\_\_\_. **Unmasking the Face: A Guide to recognizing Emotions from Facial Expressions.** Los Altos: ISHK, 2003.

EKMAN, P.; FRIESEN, W. V.; ANCOLI, S. Facial sign of emotional experience. **Journal of Personality and Social Psychology.** 39(6):1125-1134, 1980.

EKMAN, P.; ROPER, G.; HAGER, J. C. Deliberate Facial Movement. **Child Development.** 51:886-891, 1980.

FREITAS MAGALHÃES, A. **O Código de Ekman, o Cérebro a Face e a Emoção.** Ed. Fernando Pessoa - Porto, Portugal, 118, 2011.

FONAGY, I. **Languages within languages: an evolutive approach,** Amsterdam-Philadelphia: John Benjamins Publishing Company, 2000.

GOBL, C.; NÍ CHASAIDE, A. The role of voice quality in communicating emotion, mood and attitude. **Speech Communication.** 40:189-212, 2003.

IZARD, C. On the ontogenesis of emotions and emotion-cognition relationship in infancy. In: LEWIS, M. & ROSENBLUM, L. A. (eds.) **The development of affect.** New York, NY: Plenum Press, 389-413, 1978.

JACK, R. E.; GARROD, O. G. B.; SCHYNS, P, G. Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. **Current Biology.** 24(2):187-192, 2014.

JAMES, W. What is an emotion? **Mind.** 19:188-205, 1889.

JOHNSON, K.; SCHERER, K. The effects of emotion on voice quality. **Proceedings of 14th International Congress of Phonetics,** San Francisco, pp 2029-2032, 1999.

KAISER, L. Communication of affects by single vowels. **Syntheses.** 14:300-319, 1962.

KANADE, T.; COHN, J.; TIAN, Y.L. Comprehensive database for facial expression analysis. **Proce-**

edings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00) pp46-53, 2000.

KIENAST, W. F.; SENDLMEIER, M. F0 contours in emotional speech. **Proceedings of the ICPHS in San Francisco**, S.929, 1999.

KOHLER, C. G.; TURNER, T.; STOLAR, N. M.; BILKER, W. B.; BRESINGER, C. M.; GUR, R. E. Differences in facial expression of four universal emotion. **Psychiat Res.** In: press, 2004.

LAUKKA, P. Uses of music and psychological well-being among the elderly. **Journal of Happiness Studies.** 8:215-241, 2007.

LAVIER, J.; MACKENZIE-BECK, J. **Vocal Profile Analysis Scheme – VPAS.** Edinburgh: Queen Margareth University College, 2007.

LAVIER, J. **Voice quality: a classified research bibliography.** Berlin: John Benjamins Publishing, 1979.

\_\_\_\_\_. **The phonetic description of voice quality.** Cambridge: Cambridge University Press, 1980.

\_\_\_\_\_. The phonetic evaluation of voice quality. In: KENT, R. D.; BALL, M. J. **Voice quality measurement.** San Diego: Singular Publishing Group Inc. p. 37-48, 2000.

\_\_\_\_\_. **Principles of Phonetics.** Cambridge: Cambridge University Press, 1994.

MADUREIRA, S. The investigation of speech expressivity. In: MELLO, H.; PANUNZI, A.; RASO, T. (eds.) **Illocution, modality, attitude, information patterning and speech annotation.** Firenze: Firenze University Press, 1:101-118, 2011.

MORTILLARO, M., MEHU, M.; SCHERER, K. R. Subtly different positive emotions can be distinguished by their facial expressions. **Social Psychological & Personality Science.** 2:262-271. doi:10.1177/1948550610389080, 2011.

NUNES, A. M. B. **Voz e Emoção em Português Europeu.** Dissertação de Mestrado. Aveiro, Universidade de Aveiro, 2009.

PATEL, A.D.; IVERSEN, J.R.; Rosenberg, J.C. Comparing the rhythm and melody of speech and music: The case of British English and French. **Journal of the Acoustical Society of America,** 119:3034-3047, 2006.

PESSOA, A. N. **Correlatos perceptivos e acústicos de qualidade e dinâmica vocal na fala de crianças usuárias de implante coclear.** Tese de Doutorado. São Paulo, Linguística Aplicada e Estudos da Linguagem, PUC-SP, 2012.

REEVE, H. K.; SHERMAN, P.W. Adaptation and the goals of evolutionary research. **Quarterly Review of Biology.** 68:1-32. (Excerpted in: *Evolution: An Oxford Reader*, p. 120-121. Mark Ridley, Ed., Oxford University Press, 1997), 1993.

RODRIGUES, A. J. M. **As emoções na Fala.** Dissertação de Mestrado. Aveiro, Universidade de Aveiro, 2007.

SANTOS, I. **Expressividade da fala: o desvelar da locução de um poema a partir da análise acústica e da filosofia de Spinoza.** Dissertação de Mestrado. São Paulo, Linguística Aplicada e Estudos da Linguagem, PUC-SP, 2010.

SCHERER, K. R.; ELLGRING, H. Multimodal Expression of Emotion: Affect Programs or Componential Appraisal Patterns? **American Psychological Association.** 7(1):158-171, 2007.

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

\_\_\_\_\_. Expression of Emotion in Voice and Music. **Journal of Voice.** 9(3):235-248, 1995.

SCHERER, K. R.; JOHNSTONE, T.; KLASMEYER, G. Vocal expression of emotion. In: DAVIDSON, R. J.; SCHERER, K. R. & GOLDSMITH, H. (eds). **The Handbook of the Affective Sciences.** New York, NY: Oxford University Press. p. 433-456, 2003.

SILVIA, P. J. Interest – The curious emotion. **Current Directions in Psychological Science.** 17:57-

60, 2008.

SWERTS, M. G. J.; MARSJ, E. C. Prosodic evaluation of accent distributions in spoken news bulletins of Flemish newsreaders. **Journal of the Acoustical Society of America**. 132(4):2616-2624, 2012.

VAN BEZZOIJEN, R. **The characteristics and recognizability of vocal expression of emotions**. Dordrecht: Foris, 1984.

VIOLA, I. C.; MADUREIRA, S. Voice quality and Speech Expressiveness. In: **Proceedings of the Paralinguistic Speech07- between models and data**. Saarbrücken: DFKI Publication, p. 65-70, 2007.

YILDIRIM, S.; BULUT, M. C. M.; KAZEMZADEH, A.; BUSO, C.; DENG, Z.; Lee, S.; NARAYANAN, S. An acoustic study of emotions expressed in speech. **Proceedings of Interspeech**. Jeju Island: Korea, p. 889-992, 2004.

ZANOTTO, M. S. **The Multiple Readings of Metaphor in the Classroom: co-construction of inferential chains**. D.E.L.T.A. 26 (especial):615-644, 2010.

*Aspectos Teóricos para Elaboração  
de Programa de Treinamento da  
Performance Comunicativa*

LETÍCIA CORREA CELESTE  
AVELINY MANTOVAN LIMA-GREGIO

## 1. Introdução

Ao se pensar em “comunicação” é possível voltar a discussão para diferentes aspectos, como comunicação humana e animal, comunicação entre homem e máquina, dança, música etc. Tentaremos, neste capítulo, focalizar a comunicação humana no que se refere ao desempenho da transmissão de informações entre indivíduos de maneira mais formal e em torno do desenvolvimento pessoal da oratória. Para delimitar o espaço de discussão deste capítulo consideraremos apenas a zona do discurso do tipo conferência (Ortega, 1994), já que se trata da comunicação realizada com distância acima de 3 metros. Feitas tais considerações iniciais, partiremos do pressuposto que a comunicação abrange, necessariamente, estados mentais, intenção e elementos comunicativos.

Os chamados estados mentais estão relacionados, de forma geral, às emoções e atitudes do falante (Sperber e Wilson, 1995). Ambos podem ser manifestados de forma voluntária ou involuntária (Ekman e Friesen, 1969), porém, as emoções, como alegria, raiva e tristeza, são mais involuntárias, enquanto as atitudes, como a dúvida, a certeza e a ironia, podem ser mais facilmente controladas (Sperber e Wilson, 1995). Esses estados mentais são inerentes ao falante mediante determinada circunstância comunicativa.

Nesse sentido, podemos incluir as atitudes comunicativas como “intenção do falante”. Ainda dentro da intenção comunicativa, incluímos aqui também o vestuário. Tomemos como exemplo um professor universitário. Não entraremos aqui na questão do “estilo próprio”, mas esse professor pode se vestir de forma diferente a partir de sua intenção comunicativa: seu vestuário em uma palestra para gestores da universidade e seu vestuário para uma conversa informal com seus alunos pode refletir sua intenção comunicativa. As escolhas, neste ponto, mostram a intencionalidade no vestir-se: um traje mais social para a primeira ocasião e um mais esportivo para segunda.

Por fim, quanto aos elementos comunicativos, incluem-se os aspectos de conteúdo e forma. A princípio, nos parece claro que a morfossintaxe, a prosódia e até mesmo os gestos poderiam ser incluídos na forma. Entretanto, tratar desses aspectos sem antes detalhar os elementos da comunicação seria, no mínimo, imprudente.

Dessa forma, iremos discutir neste capítulo os elementos da comunicação de forma mais abrangente e com especial ênfase àqueles relacionados ao desempenho comunicativo delimitados no início da introdução. Levantaremos, ainda, a questão da avaliação da performance comunicativa e seus possíveis indicadores.

## 2. Elementos Linguísticos, Paralinguísticos e Extralinguísticos da Comunicação

A comunicação humana engloba os recursos verbais e não verbais da linguagem. Os recursos verbais incluem os elementos linguísticos e os não verbais os elementos paralinguísticos e extralinguísticos. Na zona do discurso do tipo conferência, em que temos comunicação frente-a-frente, ambos os recursos, verbais e não verbais da linguagem, interagem para a transmissão da mensagem. Neste trabalho, dentre os recursos verbais destacam-se os elementos linguísticos relacionados à fala e à voz.

Para Laver (1999), os recursos não-verbais que incluem elementos paralinguísticos são as características de gesto, expressão facial, contato visual e a direção do olhar, orientação corporal, postura, proximidade, contato físico e, ainda, as características do tom de voz e sinais vocais não linguísticos, mas com significados comunicativos específicos. O autor destaca como elementos extralinguísticos outras características não-verbais, tais como corte de cabelo e o estilo de vestir.

Esses elementos linguísticos, paralinguísticos e extralinguísticos interagem para auxiliar o ouvinte no processo de comunicação. Contudo, poucos estudos investigam essa relação. Laver

(1999) destaca as funções dos elementos paralinguísticos no apoio e na complementação – por vezes contraditórias, segundo ele – da mensagem falada. De acordo com o autor, algumas dessas funções seriam enfatizar partes das expressões linguísticas, sinalizar os aspectos atitudinais das mensagens, expressar emoção, gerenciar o tempo compartilhado entre falante e ouvinte durante as interações e gerenciar as transições entre a saudação, o assunto principal e o término das fases da interação.

Nesse contexto, a escolha dos elementos desses três aspectos da comunicação deve ser considerada para a elaboração de um treinamento que visa ao aprimoramento comunicativo. Destacaremos, portanto, os seguintes elementos:

- I. Extralinguísticos: vestuário.
- II. Paralinguísticos:
  - i) Expressão facial: contato de olhos, relação entre a direção do olhar e a ocorrência de disfluências de fala, expressividade.
  - ii) Expressão corporal: postura, caminhar, diversidade gestual, relação entre gesto e discurso, amplitude gestual.
- III. Linguísticos:
  - i) Fala e voz: articulação, velocidade de fala, pausas, melodia, ênfase, intensidade.

É importante ressaltar que escolhemos esses elementos pensando em um treinamento para grupos de até 15 pessoas, exclusivamente prático, de curta duração (em torno de 25 horas) e, por isso, de fácil compreensão e aproveitamento para os participantes. Exploraremos cada um desses elementos nos tópicos a seguir.

## **2.1. Vestuário**

Dissemos inicialmente neste capítulo que o vestuário pode ser analisado como intenção comunicativa. Em um estudo com voluntários a fim de verificar a possível modificação comportamental de pacientes com relação a médicos de acordo com o vestuário, foi verificado, com uma proporção muito alta, que os pacientes tendem a confiar mais em médicos que usam jalecos brancos com roupas formais por baixo (Rehman et al, 2005). Em ambientes corporativos, o vestuário carrega uma série de esquema simbólicos, além de questões comportamentais laborais (Rafaeli et al, 1997). Nesse sentido, o vestuário pode ser interpretado como comportamento comunicativo que, apesar de não transmitir diretamente uma mensagem, comunica.

Não propomos aqui que um treinamento das habilidades comunicativas trace roupas “certas” e “erradas” para determinadas situações. Defendemos que a discussão sobre o assunto pode trazer efeito positivo a partir do debate em torno da questão.

## **2.2. Expressão facial**

O que dizem nossas expressões faciais? Estudos recentes (Keltner e Elkman, 2000, 2003; Kendler et al, 2008; Ekman, 2016) apontam para uma resposta comum: emoções. E o reconhecimento dessas emoções é rápido e universal, ou seja, qualquer ser humano reconhece as emoções nas expressões faciais de qualquer ser humano, independentemente de raça, nacionalidade, sexo (Keltner e Elkman, 2000), salvo algumas patologias específicas. Voltando a análise para o foco deste capítulo, é natural nos perguntarmos se existe a possibilidade de controlarmos nossas ex-

pressões faciais para não demonstrá-las.

Ao se levantar para apresentar um trabalho, um estudante universitário pode sentir diferentes emoções: medo (de esquecer o ensaiado, de tirar uma nota baixa, de se expor perante a turma etc.), desespero (uma vontade extrema de voltar a se sentar na cadeira que acabou de levantar), tranquilidade (simplesmente pode ser uma pessoa sem qualquer problema com falar em público, além de ter domínio do tema) etc. Dessas e de tantas outras, a expressão facial de “tranquilidade” pode ser almejada, independentemente da emoção presente naquele estudante. Podemos “enganar” nosso interlocutor a ponto de transmitir outra emoção em uma situação de vida real? Ekman e Friesen (1969) afirmaram que a expressão facial é a forma de transmissão mais involuntária e, ao mesmo tempo, a que mais tentamos alterar, podendo resultar em uma fonte “confusa” de informações. Porém, mais do que expressar emoções gerais, como destacadas acima, as sutilezas da mímica facial trazem informações sobre questões comportamentais (Ladkin e Taylor, 2010; Stewart e Fodr, 2013).

Dessa forma, a reflexão e o treino da expressão facial podem auxiliar no aprimoramento das habilidades comunicativas e, por esse motivo, devem ser incluídos em programas específicos dentro do tema.

### **2.3. Expressão corporal**

Antes de mais nada, tratamos o termo “expressão corporal” como postura e gestos. Distância, orientação corporal, posição dos braços e o estado de relaxamento ou tensão corporal são variáveis relevantes da comunicação (Mehrabian, 1969). Em um estudo recente com apoio de ressonância magnética funcional (fMRI), pesquisadores verificaram que a sensação de que o outro está direcionando a fala a você provém da díade gestos-postura (Nagels et al, 2015).

A dicotomia corpo e objeto do discurso pode ser analisada sob a perspectiva da inter-relação corpo e língua nos processos de significação da comunicação (Vezali, 2012). Essa complexa relação é discutida por meio de diferentes teorias. A proposta de McNeill (1992), na qual o autor destaca a relação entre linguagem, gesto e cognição, traz quatro tipos principais de gestos: gestos icônicos (gestos exemplificam a fala), gestos metafóricos (também representacionais, porém de elementos abstratos), gestos dêiticos (direcionais ou demonstrativos) e gestos ritmados (não são associados ao conteúdo, mas à musicalidade da fala). Naturalmente, as informações transmitidas pelos gestos que acompanham o discurso ficam gravadas e auxiliam na compreensão do discurso. Mais do que isso, porém, ouvintes retêm informações que são passadas apenas por gestos e, ainda, quando há discrepância entre gestos e fala, os gestos ficam mais marcados que a fala – paradigma da incompatibilidade (Cassel e McCullough, 1999).

Com isso, defendemos que o trabalho com a comunicação oral implica, necessariamente, pensar no equilíbrio entre postura, expressão facial, expressão corporal e aspectos linguísticos.

### **2.4. Fala e Voz**

A linguagem verbal humana é usualmente dividida em dois níveis estruturais: o mais alto, chamado de gramatical, e o mais baixo, fonológico. O nível gramatical é o que decide as unidades capazes de fazer referências externas no campo da semântica, tais como as palavras, frases, orações e sentenças (Laver, 1994). Não estudaremos os elementos do nível gramatical aqui. Por sua vez, no nível fonológico estão as unidades abstratas que representam os sons da língua em questão (Laver, 1994). Esse nível, que é segmental, inclui a pronúncia dos sons da fala. Ainda, inclui os aspectos suprasegmentais, que abarcam os elementos prosódicos relacionados, principalmente, à duração e à frequência fundamental. A prosódia é, portanto, um aspecto intrínseco determinante

da forma da linguagem falada (Cutler, Oahan, Donselaar, 1997).

Mesmo os estudos sobre produção de fala que possuem foco nos aspectos paralinguísticos, assim como o reconhecimento da personalidade do falante – i. e., se é competente ou extrovertido, por exemplo (Mairesse et al., 2007) –, relatam a importância de pistas linguísticas, tais como os parâmetros acústicos de intensidade de voz, variação de frequência fundamental, ocorrência e duração de pausas e velocidade de fala. Embora alguns atributos pessoais contribuam para a caracterização da imagem de um “bom falante”, em um treinamento de aprimoramento da comunicação o cerne do trabalho dos elementos linguísticos é a melhora da inteligibilidade de fala. No conceito de inteligibilidade de fala abarcamos todo o entendimento da mensagem de acordo com o que o falante realmente deseja transmitir, desde a intenção até a articulação final dos sons e gestos.

No dia-a-dia o falante modifica conscientemente a sua fala de modo que ela fique mais inteligível sempre quando percebe alguma dificuldade de percepção do ouvinte. Conforme temos tratado neste trabalho, na fala para conferências a modificação consciente acontece para adequá-la à modalidade da apresentação. Tais mudanças incluem a diminuição da velocidade, o aumento da intensidade e a melhor articulação dos sons, com o intuito de prover maior saliência das pistas acústicas e permitir ao ouvinte compreender mais claramente a mensagem (Smiljanić e Bradlow, 2009).

Um dos efeitos da diminuição da velocidade de fala é o aumento da duração dos segmentos, e estudos demonstram que há melhora na inteligibilidade de fala. Fowler e Housum (1987) observaram que a primeira ocorrência de palavras em um monólogo de rádio teve o parâmetro acústico “duração” maior. Em um experimento adicional, os mesmos autores viram que uma simples tarefa de colocar aquelas palavras fora do contexto foi suficiente para que elas fossem percebidas como mais inteligíveis do que na segunda ocorrência delas no monólogo.

Ademais, a diminuição da velocidade de fala implica em movimentação mais lenta dos articuladores, especialmente o lábio inferior e a língua (Dromey e Ramig, 1998), ambos solidários ao movimento da mandíbula. A redução da velocidade de fala pode influenciar, portanto, na articulação da fala. Em um treino de performance comunicativa essa informação pode ser relevante, tendo em vista que a instrução para falar mais lentamente associada com a instrução para prestar atenção na maneira de articular os sons pode ter um resultado positivo para o aprimoramento da comunicação.

Além da velocidade de fala, a pausa é um parâmetro relacionado aos aspectos temporais da estrutura da fala. Viola e Madureira (2008) apresentaram várias perspectivas para a descrição das pausas, a saber: (a) estrutural, que podem ser preenchidas por silêncio ou por pistas fonético-acústicas, tais como alongamento, alterações na qualidade da voz e variação de  $f_0$ ; (b) funcional, que podem ser classificadas como respiratória, discursiva ou expressiva; (c) distribucional, que podem ocorrer intra ou entre as frases e outros constituintes morfossintáticos; (d) temporal, que podem ser curto ou longo. Um estudo de Celeste e Reis (2013), que investigou os aspectos temporais envolvidos na expressão de certeza e dúvida, indicou que as pausas e disfluências da fala estiveram presente como estratégia comunicativa na expressão de dúvida. Portanto, em conferências, o falante deve planejar adequadamente a utilização das pausas mais apropriadas a essa modalidade.

Outro elemento importante para o treinamento da performance comunicativa é a ênfase. Kohler (2006) caracterizou a função comunicativa dos acentos e fez uma proposta preliminar para a categorização da ênfase em dois tipos principais: a ênfase para o foco e a ênfase para a intensidade. Ambos os tipos envolvem vários parâmetros acústicos que não apenas a frequência fundamental ( $f_0$ ), conforme pode-se observar em algumas propostas sobre ênfase da fonologia tradicional. No caso do foco há a possibilidade de ser falante ou ouvinte-orientado e serve para seleção de informação, contraste fatural, contraste de afeto, surpresa fatural e surpresa de afeto. Por sua vez, a intensidade pode se dividir em positiva e negativa, indicando a avaliação expressiva de aprovação

e desaprovação do falante, respectivamente.

Um parâmetro acústico diretamente relacionado à frequência fundamental é a melodia da fala. De acordo com Haer, Collier e Cohen (2006), a melodia da fala é um atributo que varia continuamente e abrange unidades linguísticas enquanto enunciados completos. Esses autores afirmam que não há evidências se, e de que forma, ela pode ser dividida em unidades menores. Em um estudo brasileiro sobre a variação melódica, Celeste e Reis (2011) investigaram a produção e percepção de atitudes de certeza e dúvida em indivíduos gagos e não gagos. Os autores observaram que os grupos de falantes gagos produziram menor variação melódica do que o grupo de falantes não gagos. No teste de percepção, os participantes que representaram os interlocutores em um diálogo não conseguiram interpretar os enunciados de certeza e dúvida, como uma expressão explícita dessas atitudes, quando os enunciados foram produzidos pelos grupos de falantes gagos. Esse achado, embora tenha sido relacionado a indivíduos com desordem na fala, revela a pertinência de se abordar esse parâmetro em um treino de aprimoramento da comunicação.

## **2.5. Avaliação da performance comunicativa**

Não existe um olhar único e “perfeito” para todos os públicos em treinamento. Exemplificando essa afirmação, um excelente avaliador para um treino com docentes seria o olhar do discente, da mesma forma que para avaliar o discente, o olhar do docente forneceria um bom respaldo prático; e para um jornalista um telespectador, e assim por diante. Entretanto, tal procedimento é inviável na prática daqueles que irão fazer do treinamento o seu trabalho cotidiano.

Sob essa perspectiva, a autoavaliação, preferencialmente conduzida para os aspectos trabalhados, ganha força na verificação da eficácia do treinamento para o aperfeiçoamento dos padrões comunicativos. De acordo com Santos (2002), a autoavaliação permite que o aluno reflita de forma consciente sobre o seu próprio processo de aprendizagem, desenvolvendo a capacidade de autoquestionamento.

Por outro lado, sob o panorama da pesquisa, seria interessante contrapor dados da autoavaliação com avaliações de profissionais especializados na temática. Nesse sentido, os resultados dos dois tipos de avaliação seriam complementares.

## **3. Considerações finais**

A partir do embasamento teórico apresentado neste capítulo, acreditamos que a elaboração de um treinamento das habilidades comunicativas com foco na comunicação do tipo conferência deve considerar a abrangência dos seguintes tópicos:

- (1) Conhecer o grupo
- (2) Levar o grupo a se conhecer
- (3) Treinamento
- (4) Feedback

Conhecer o grupo implica em avaliar a performance comunicativa do participante antes do treinamento. É o momento para identificar as potencialidades e também os elementos que merecem mais atenção. Sugerimos que seja feita uma gravação com duração média de 3 minutos. Essa gravação pode ser utilizada para comparação da performance após o treinamento.

Em um segundo momento é importante que o participante faça a sua autoavaliação a partir da mesma gravação. É interessante que o grupo avalie os seus colegas e aponte livre e gentilmente os aspectos que consideram que precisam ser melhorados. Ou, ainda, podem-se atribuir notas (exemplo: notas de 0 a 5 para cada parâmetro que será incluído no treinamento). Para pesquisas

sobre o tema, avaliações por profissionais especializados podem contribuir para a validação das autoavaliações.

O treinamento pode ser realizado via encontros semanais ou intensivo diário. No intensivo não seriam necessários os exercícios para casa, já no semanal eles seriam úteis para que haja fixação do tema abordado. Ressaltamos que seria de extrema valia estudos que comparassem a efetividade de programas de cunho intensivo e extensivo.

Para a elaboração do treinamento propriamente dito é necessário que se conheça o grupo e se trace estratégias que possam incluir tanto os aspectos verbais quanto os não verbais da comunicação. Acreditamos que seja interessante a realização de exercícios em grupos menores e treino de apresentações com todos os participantes presentes para que cada um tenha a oportunidade de experimentar as técnicas com mais liberdade (dentro de grupos menores), mas também se coloque em situação de maior estresse (apresentações para todos os participantes). A prática de apresentações orais deve ter o objetivo de direcionar o treino para o elemento que está sendo trabalhado naquele momento. Essas apresentações devem ser gravadas para que o participante possa refletir sobre as suas peculiaridades.

Ao final do treinamento sugerimos que seja realizada uma gravação nos mesmos moldes da primeira gravação: com duração em torno de 3 minutos e, preferencialmente, de mesmo tema da primeira. Essa espécie de reavaliação deve ser realizada pelo próprio participante (segundo momento de autoavaliação) e por seus pares (os membros do grupo). É fundamental olhar para os dois momentos de gravação e ter clareza dos pontos que foram aprimorados, dos que não foram, e traçar novos “desafios”.

## Referências Bibliográficas

CELESTE, L.C.; REIS, C. Variação melódica na expressão de atitudes de indivíduos com gagueira. **Rev de Est da Ling**. Belo Horizonte: v.19, n.1, p. 191-217, 2011.

\_\_\_\_\_. Expressão de certeza e dúvida na gagueira: estudo dos aspectos temporais da fala. **Rev CEFAC**. São Paulo: v.15, n.6, p. 1609-20, 2013.

CUTLER, A.; OAHAN, D.; DONSELAAR, W. Prosody in the comprehension of spoken Language. **Lang and Speech**, v.40, n.2, p. 141-201, 1997.

DROMEY, C.; RAMIG, L.O. Intentional changes in sound pressure level and rate: their impact on measures of respiration, phonation, and articulation. **J Speech Lang Hear R**, v.41, n.5, p. 1003-18, 1998.

EKMAN, P. What Scientists Who Study Emotion Agree About. **Ass Psych Sci**, v.11, n.1, p. 31-34, 2016.

EKMAN, P; FRIESEN, W. V. Nonverbal leakage and clues to deception. **Psychiatry**, v.32, p. 88-105, 1969.

FOWLER, C. A.; HOUSUM, J. Talkers' signaling of “new” and “old” words in speech and listeners' perception and use of the distinction. **J of Mem Lang.**, v.26, p. 489-504, 1987.

HART, J.; COLLIER, R.; COHEN, A. **A perceptual study of intonation: an experimental-phonetic approach to speech melody**. Cambridge: Cambridge U P, 2006.

KELTNER, D.; EKMAN, P. Introduction: Expression of Emotion. In: DAVIDSON, R. J.; SCHERER, K. R.; GOLDSMITH, H. H. **Hdbk of Affective Sci**. New York: Oxford University Press, 2003.

\_\_\_\_\_. Facial Expression of Emotion. In: LEWIS, M.; HAVILAND-JONES, J. **Hdbk of emotions**. New York: Guilford Publications, p. 236-249, 2000.

KENDLER, K. S.; HALBERSTADT, L.J.; BUTERA, F.; MYERS, J.; BOUCHARD, T.; EKMAN, P. The similarity of facial expressions in response to emotion-inducing films in reared-apart twins. **Psych**

**Med**, v. 38, n. 10, p. 1475-1483, 2008.

KOHLER, K. What is Emphasis and How is it Coded? In: **SPEECH PROSODY**, Dresden. **Proc Speech Prosody Conf.** Dresden, 2006.

LADKIN, D.; TAYLOR, S. Enacting the 'true self': Towards a theory of embodied authentic leadership. **The Leadership Quarterly**, v. 21, n.1, p. 64-74, 2010.

LAVIER, J. **Princ of phonetics**. Cambridge: Cambridge University Press. 1994.

\_\_\_\_\_. Towards an integrated theory of nonverbal Com. In: ICPHS, 1999. **Proceedings of the XIVth Int Cong of Phonetics Sci.** San Francisco: p. 2433-2436, 1999.

MAIRESSE, F.; WALKER, M. A.; MEHL, M. R.; MOORE, R.K. Using Ling. Cues for the Automatic Recognition of Personality in Conversation and Text. **J of Artif Intellig Res**, v. 30, p. 457-500, 2007.

MEHRABIAN, A. Significance of posture and position in the Com of attitude and status relationships. **Psych Bulletin**, v. 71, n.5, p. 359, 1969.

NAGELS, A; KIRCHER, T.; STEINES, M.; STRAUBE, B. Feeling addressed! The role of body orientation and co-speech gesture in social Com. **Human Brain Mapping**, v. 36, n. 5, p. 1925-1936, 2015.

ORTEGA, A. El discurso político: retorica, parlamento. **Dialectica**. Madri: Fundation Canovas del Castillo. Coll XXI. 1994.

Rafaeli, A.; Dutton, J.; Harquial, C.; Mackie-Lewis, S., Navigating by attire: the use of dress by administrative employees, **Academy of Management Journal**, Vol. 40 No. 1, pp. 19-45, 1997.

REHMAN, S. U.; NIETER, P. J.; COPE, D. W.; KILPATRICK, A. O. What to wear today? Effect of doctor's attire on the trust and confidence of patients. **The Am J of Med**, v.118, n.11, p. 1279-1286, 2005.

SANTOS, L. Auto-avaliação regulada: porquê, o quê e como?. In: ABRANTES, P.; ARAÚJO, F. **Avaliação das Aprendizagens. das concepções às práticas**. Lisboa: DEB ME. p. 75-84, 2002.

SMILJANIĆ, R.; BRADLOW, A. R. Speaking and Hearing Clearly: Talker and Listener Factors in Speaking Style Changes. **Lang Ling Compass**, v. 3, n. 1, p. 236-264, 2009.

SPERBER, D.; DEIRDRE W.: **Relevance: Communication and Cognition**, second edition, Oxford, Cambridge: Blackwell Publishers, 1995.

STEWART, P. A.; FORD, P. K. Interpreting President Barack Obama's Facial Displays of Emotion: Revisiting the Dartmouth Group. **Political Psych**, v. 34, n. 3, p. 369-385, 2013.

VEZALI, P. O corpo: considerações acerca da relação entre fala e gesto. **ILINX-Rev LUME**, v. 2, n.1, 2012.

VIOLA, I. C.; MADUREIRA, S. The roles of pause in speech expression. In: **SPEECH PROSODY**, 2008, Campinas. **Proceedings of the Speech Prosody Conf.** Campinas:p. 721-724, 2008.

***Dez Tópicos Sobre Expressividade Oral  
para o Fonoaudiólogo: uma Proposta de  
Debate***

IZABEL VIOLA  
LÉSLIE PICCOLOTTO FERREIRA

# 1. Introdução

O fonoaudiólogo, ao avaliar o sujeito que o procura com uma determinada queixa, tem que buscar estratégias para reverter as questões que precisam de atenção, e nem sempre esta prática “dialoga” na mesma sintonia com a teoria.

Partimos, portanto, de algumas evidências: a primeira, de que, nos dias de hoje, o fonoaudiólogo não tem mais dúvida da presença e importância da vertente psicossocial da voz e que o estudo e a pesquisa sobre a expressividade oral, assim como sobre os gestos corporais, são possibilidades de escancarar tal vertente, pois, afinal, voz e linguagem não podem ser vistas isoladamente; a segunda, de que ao trabalhar com os chamados profissionais da voz, percebe-se que o fonoaudiólogo, mais que registrar, como faz em sua intervenção, caminha na direção da descrição, da caracterização do perfil dos diferentes profissionais, com maior ênfase entre os teleapresentadores e atores; a terceira, que a pesquisa e a atuação com os profissionais da voz foram subsidiadas, em sua maioria, pela Linguística, mais especificamente pelos estudos que usam o referencial da análise fonética-acústica e aqueles que preconizam uma interface com a Sociologia; quarta, que a avaliação perceptivo-auditiva, de forma isolada ou em conjunto com a acústica, tem contribuído na realização de importantes trabalhos e o julgamento de leigos ou especialistas tem sido também um procedimento metodológico recorrente, na tentativa de entender as relações entre recursos fônicos e sentidos, pois um mesmo recurso pode gerar diferentes efeitos de sentido, ou ainda um determinado sentido pode ser materializado por diferentes recursos (Ferreira, 2004).

Dessa forma, separamos dez perguntas que frequentemente nos fazem e que sempre permitem discutir aspectos inerentes à pesquisa e à prática do fonoaudiólogo.

- 1 - O que é expressividade?
- 2 - Qual termo seria mais indicado para definir o trabalho do fonoaudiólogo na área: expressividade oral, vocal ou verbal?
- 3 - Pode-se dizer que um sujeito é mais ou menos expressivo?
- 4 - O que vem a ser ênfase?
- 5 - A qualidade de voz pode ser considerada um parâmetro para marcar a expressividade?
- 6 - Qual a diferença entre ritmo e velocidade?
- 7 - Como compreender as pausas a serviço da expressividade?
- 8 - A articulação pode ser vista como marcador de expressividade?
- 9 - O que observar e como proceder na avaliação da expressividade oral?
- 10 - Qual é a diferença entre oratória e expressividade?

Para alguns leitores, as respostas podem ser simples, e, em alguns momentos, parecem estar repetindo aspectos mencionados em outras produções bibliográficas. Contudo, nosso intuito foi organizar um texto que, de forma mais didática, facilite o entendimento daqueles que estão se formando nessa área.

## 1.1. O que é expressividade?

O significado da palavra “expressão”, no sentido etimológico, é apertar com força, espremer e tirar comprimindo. Como interpreta Fónagy (1983) é “ex-pressão” ou eliminação do que cria tensão, entendida como uma força que impulsiona. Assume também o significado de reproduzir, representar, retratar, exprimir, dizer, expor e enunciar claramente. Desta forma, do ponto de vista do falante, o objetivo da expressão é eliminar as tensões internas, reproduzindo-as em comporta-

mentos semióticos, como metáforas sonoras (Fónagy, 1983) ou como gestos vocais (Viola, 2006).

A expressão é controlada em diferentes níveis de processamento cognitivo. Atitudes e emoções são expressas pelo locutor, e podem ser distintas pela natureza do controle na sua expressividade, do voluntário ao involuntário. Entende-se por atitudes (designado, atualmente, também como prosódia audiovisual), os comportamentos determinados, conscientes, controlados, e que têm um componente moral e intelectual, como ironia, reprovação e justificação. Por outro lado, as emoções são entendidas como descargas espontâneas da tensão psíquica de fonte pulsional, como raiva, alegria, tristeza, surpresa e angústia.

Entretanto, a expressão da emoção e da atitude está vinculada às exigências formais da língua, denominadas modalidades. Essas são atitudes fundamentais com relação ao objetivo da comunicação, ou seja, a apresentação do enunciado como narração de um fato, hipótese, desejo, declaração, exclamação, ordem e pergunta. As atitudes básicas são marcadas nas diferentes línguas por meio de morfemas determinados, pela ordem das palavras ou por formas de entoação constantes, delimitadas de maneira nítida.

A diferenciação entre emoções, atitudes e modalidades é feita de acordo com os graus de regularidade convencional. As emoções primárias são expressas em todos os níveis do aparelho vocal, por mudanças fisiológicas. A expressão de atitude repousa sobre um nível mais elaborado e regularizado, em termos de convenções comunicativas e a expressão de modalidade é altamente estruturada num sistema gramatical (Chung, 2000). Uma frase de mesma modalidade pode expressar diferentes emoções ou atitudes. Por exemplo: um padrão entoacional ascendente-descendente numa pergunta do tipo “Você vai lá?” (modalidade), ao se associar a diferentes qualidades de voz e/ou duração vocálica de algumas vogais, pode expressar tanto uma atitude assertiva como uma emoção de surpresa.

Como alguns autores se referem à fala expressiva como uma fala caracterizada por variabilidade de padrões melódicos e rítmicos, gera um entendimento que a definição seria ligada a uma fala de prosódica rica e que haveria graus de expressividade, as mais e as menos expressivas. Discutimos este ponto na Questão 3.

Cabe lembrar que, na Fonoaudiologia, outros termos são utilizados para fazer referência à expressividade oral. Porém, partimos dos pressupostos de Fónagy (1983), teórico pioneiro em estudos da área.

## **1.2. Qual termo seria mais indicado para definir o trabalho do fonoaudiólogo, na área: expressividade oral, vocal ou verbal?**

No ato da interlocução, a expressividade está presente em diferentes níveis, por meio de elementos verbais (palavras que fazem parte da língua falada), não verbais (elementos visuais, tipo corporal e facial), vocais (elementos sonoros audíveis) e, informações de natureza semântica, regulativas e evidenciais. As informações semânticas, de natureza verbal, são advindas do significado direto da expressão falada, ligada ao léxico escolhido pelo falante. As regulativas têm função de controlar a interação e são manifestas por informações vocais e não verbais, por exemplo, a entoação, o sincronismo entre os participantes e as manifestas pelo canal visual (gestos, posturas, movimentos de cabeça e contato de olho). As informações evidenciais de natureza vocal, manifestas, por exemplo, na qualidade de voz e na prosódia, revelam características pessoais do falante, como as marcas físicas (sexo, idade, condições físicas e estado de saúde), as sociais (filiação regional e nível sócio-educacional, ocupação e papel social) e psicológicas (personalidade, estado afetivo e humor) (Pittam, Scherer, 1993; Laver, 1994).

A complementação do termo expressividade, com palavras como corporal, gestual, da fala, vocal, oral ou verbal, reporta-se a diferentes áreas de domínio, e a um entendimento dissociado en-

tre fala e voz, como se pudéssemos dissociar as características fonéticas dos sons do gesto glotal e dos articuladores.

Propomos, dessa forma, a adoção do termo expressividade oral, para não carregar uma ou outra acepção (fala-voz), mas por se referir “ao som que sai da boca” (etimologicamente do radical latino *os, oris* “boca”) (Viola, 2006).

### **1.3. Pode-se dizer que um sujeito é mais ou menos expressivo?**

O uso do termo expressividade remete a uma fala caracterizada por variabilidade de padrões melódicos e rítmicos, em contraposição a uma fala com menos variação, referida, algumas vezes, como neutra. Contudo, ser expressivo não significa manter um uso variado de recursos orais, da mesma forma que não existe adequação ou inadequação no uso desses recursos, enquanto forma expressiva.

Como afirma Madureira (2004; 2011), toda fala é expressiva, porque veicula alguma atitude, emoção, crença, estado físico ou condição social. Mesmo a comumente referida fala monótona pode ser interpretada pelo ouvinte como indicadora de falta de entusiasmo, apatia, desinteresse, entre outros significados. Um bom exemplo, são as vozes de comando, que, próprias ao seu contexto de uso, têm prosódia com pouca variabilidade, caracterizada por alta intensidade, articulação precisa, *pitch* agudo (Santos, 2005; Souza LMC, 2007), com tessitura extensa na tônica proeminente (Souza LMC, 2007).

Beller (2009) trata a expressividade como um nível de informação na mensagem vocal, que agrupa as manifestações externas, controladas ou não, que podem ser atribuídas a estados internos não controlados e inacessíveis, como emoções, sentimentos, atitudes, humores e estados psicológicos. O autor exemplifica dizendo que um indivíduo (identidade do orador) ao pronunciar (modalidade e proeminência) uma mensagem linguística (mensagem semântica), em um determinado ambiente físico e sob certas restrições sociais (estilo de fala) e, de acordo com seu estado interno, decide expressar ou suprimir (expressividade) algo da mensagem. A existência de um nível de expressividade zero, ou seja, neutra, seria a ausência de informação emocional ou psicológica interna.

Fónagy (1983) nomeia essa informação desprovida de emoções e/ou atitudes como modalidade, por suas características mais sintáticas. Em discursos de estilo pedagógico é desejável pautar-se na divisão sintática, pois, desta forma, se torna mais objetivo.

A pesquisa de Marquezin *et al.* (2015) reforça a ideia de que um discurso mais organizado sintaticamente é percebido como mais seguro, objetivo, empático e convincente. As autoras analisaram a expressividade da fala de quatro executivos (S1 a S4)<sup>1</sup>, a partir de dados perceptivos e acústicos da dinâmica vocal. A amostra de fala coletada foi caracterizada por um relato com sugestões de produtos de investimentos, finalizando com uma frase de convencimento sobre o que foi apresentado na consultoria. Esses relatos foram ouvidos por leigos que, ao final, concluíram que as características de fala e as atitudes comunicativas de dois sujeitos de forma contraposta se destacaram (S1 e S4), de modo que a taxa de elocução lenta e as quebras dos grupos prosódicos transmitiram insegurança, pouca objetividade e não convencimento, enquanto o aumento da intensidade, e, principalmente, o aumento da  $f_0$  foram percebidos como recursos importantes na expressão das emoções.

### **1.4. O que vem a ser ênfase?**

Dois dos significados que o dicionário apresenta para a palavra ênfase merecem ser comentados. O primeiro, vigor de expressão que salienta ou torna mais impressionante um assunto e pro-

eminência especial dada na leitura ou oração a uma ou mais palavras; e o segundo, sílabas para atrair a atenção ou fazer sobressair a importância especial lógica ou emocional dessas palavras. São sinônimos de ênfase: destaque, relevo, alarde, entusiasmo, ostentação, abordagem, foco, exaltação, louvor, efervescência, exibição. São antônimos: naturalidade e simplicidade (Michaelis, 2016).

A oposição ênfase x naturalidade/simplicidade, proposta pelo dicionário, não faz sentido, porque elas nem sempre acontecem de forma chamativa, a ponto de se perder a naturalidade ou se dar rebuscamento, a ponto de se perder a simplicidade. Toda fala apresenta ênfase (proeminências), embora algumas em menor quantidade se comparadas a outras, como as falas monótona e triste. As ênfases ocorrem por diferentes estratégias prosódicas que modificam o andamento melódico ou temporal, seja por mudança na qualidade de voz, variação de *pitch* e *loudness* ou na distribuição das pausas e taxa de elocução. A mudança da ênfase pode estar associada ao assunto em andamento (tópico); à introdução de um elemento novo na conversação (foco); ao tipo de modalidade frasal (negativa, interrogativa, exclamativa e afirmativa) e para exprimir atitudes e emoções assumidas pelo locutor (Viola, 2008). A ênfase afeta apenas o item lexical que está em destaque.

Outro sentido dado ao termo ênfase é apresentado por Reis (2005), que diferencia ênfase de foco. O autor afirma que foco seria o termo usado nas situações de diálogos espontâneos, enquanto a palavra ênfase seria utilizada em contexto de leitura em voz alta, como em telejornal, uma vez que na ausência de pistas contextuais o locutor tem liberdade de destacar a palavra que lhe interessar.

Alguns exemplos podem esclarecer o que foi dito. Numa resposta do tipo PAULO ESTEVE AQUI HOJE, pode-se destacar: a palavra PAULO, se a pergunta era “quem esteve aqui hoje?”. ESTEVE, se quem perguntou estivesse duvidando que Paulo esteve aqui hoje; AQUI, ao se perguntar onde Paulo esteve; e finalmente HOJE, ao responder quando Paulo esteve aqui. Essas palavras destacadas, ao serem emitidas, podem carregar diferentes recursos de ênfase. Por exemplo, ao responder à pergunta “quem esteve aqui?”, a palavra PAULO, que inicia a frase, pode ser expressa utilizando-se o recurso de aumento de *loudness*, destacando-se, assim, uma afirmativa, ou com *loudness* diminuído, no caso da intenção ser passar a ideia dessegredo ao interlocutor, ou ainda utilizando um prolongamento da primeira vogal, ou uma modulação ascendente e descendente, numa expressão que o outro pode interpretar como “você não vai acreditar”. Enfim, diferentes recursos (e, na maioria das vezes, combinados entre si) podem ser utilizados para destacar uma determinada palavra, lembrando que cada um deles irá gerar, no ouvinte, diferentes sentidos.

### **1.5. A qualidade de voz pode ser considerada um parâmetro para marcar a expressividade?**

Ao se considerar a interação das estruturas do trato vocal na produção de uma determinada qualidade de voz, duas formas podem ser apresentadas:

---

<sup>1</sup> S1, inseguro, não objetivo, não empático e não convincente, sendo que, na análise dos recursos, foi identificado o uso inadequado de pausas, formadas, principalmente, por hesitações; separação inadequada dos grupos prosódicos, com quebra de constituintes sintagmáticos; S2, seguro, mas pouco objetivo, empático e convincente, apresentando uso regular de pausas para recarga respiratória, organização de frases e ênfase; S3, apontado como seguro, objetivo, empático e convincente, com uso de pausas regulares para recarga respiratória e organização de frases e hesitações; e, finalmente, S4, o mais seguro, objetivo, empático e convincente, com uso adequado das pausas para recarga respiratória, planejamento e ênfase, com grupos prosódicos que coincidiram com o enunciado, sem separar os constituintes sintagmáticos.

**1** - Concepção mais tradicional, na Fonoaudiologia, em que as fontes glótica e supraglótica são analisadas separadamente. Desta forma, é necessário especificar o ajuste glótico (tipo de coaptação das pregas vocais, para saber se há fenda, sobreposição e/ou desnivelamento); os ajustes ressonanciais (laríngeo, faríngeo, oral, nasal e nasal compensatório); a forma de articulação, referente aos pontos articulatorios (precisa e imprecisa) e grau de abertura mandibular (travada, fechada, adequada, aberta e sobrearticulada) (Viola e Madureira, 2007). Ao final, tem-se uma impressão global, sendo registrado o aspecto de maior evidência, descrito pelo elemento que ela traduz quanto à qualidade da fonte glótica ou dos ressoadores (como, por exemplo, vozes sussurradas, estridentes ou nasais), ou por uma característica física ou emocional do falante (voz infantilizada, virilizada, afeminada ou pastosa). Nesta vertente, podem-se destacar os inúmeros descritores impressionísticos, apresentados por Boone (1996). Eles se remetem à cor (clara, escura, dourada, incolor, branca, brilhante), forma (aberta, redonda, pontuda, cortante), velocidade (lenta/rápida), peso (leve/pesada), sexo (efeminada/masculinizada), temperatura (quente/fria), força (forçada, comprimida, relaxada, mole, fraca, forte), textura (áspera, macia, suave, dura), umidade (seca/úmida), espessura (grossa/fina), tamanho (alta, grande, pequena), qualidade (feia, bonita, boa, ruim, crepitante, gostosa), movimento (flutuante, efervescente, oscilante, trêmula), idade (velha, infantil, jovial), atitude (ameaçadora, cruel, falsa, convincente, sedutora, antipática, descontrolada), emoção (alegre, triste, medrosa) e outras categorias, que implicam em oposição, como, por exemplo, pobre/rica; suja/limpa; ardida/melosa.

**2** - Concepção mais atual, que entende a qualidade de voz como produto final da integração da fonte (glote) e dos filtros (ressoadores e articuladores), conforme o Modelo Fonte-Filtro de Fant (1970), seguido por Laver (1980; 1994). A partir desta proposta, a produção de uma determinada qualidade de voz vai depender de como se comporta o trato vocal, em toda sua extensão. Para Laver (1980; 1994), estes ajustes se constituem em *settings*, e são delineados a partir de um ajuste denominado neutro, que se caracteriza pela ausência de tensão e relaxamento do trato vocal; protrusão dos lábios (a não ser para a emissão de sons bilabiais); nasalidade (a não ser para a emissão de sons nasais); movimentação da laringe (nem levantada, nem abaixada); diâmetro do trato vocal supralaríngeo, uniforme ao longo do seu comprimento; articulações orais anteriores, com raiz da língua nem avançada nem recuada; mandíbula, nem aberta nem fechada e vibração das pregas vocais periódica, sem fricção audível.

Uma frase que não apresenta nasalidade, como, por exemplo, “o carro bateu”, pode ser pronunciada com um ajuste de nasalização, por um determinado falante que tem esta característica de voz, ou pela presença de uma alteração de palato ou por características regionais. Diferentes ajustes de qualidade de voz podem ser registrados (labialização, laringe levantada, laringe abaixada, voz rangida, voz de falsete, entre outros), e Camargo e Madureira (2008a; 2008b), ao adaptar o instrumento para o Português Brasileiro, incluiu, na proposta, as ocorrências de diplofonia, quebras e instabilidades.

Do ponto de vista da expressão de emoção e atitudes, a qualidade vocal é um dos elementos acústicos que compõe a expressividade oral de um sujeito. Esse, mesmo estando num processo inconsciente, permite que o ouvinte reconstrua suas características inatas, como sexo, idade e estado físico, além de informações de ordem comunicativa e/ou emocional. Há ajustes usados habitualmente pelo falante, denominados de longo prazo, em oposição aos de curto prazo, usados temporariamente, para as manifestações de cunho expressivo ou de um estado físico, como afecções respiratórias.

Exemplos disso são apresentados por Viola e Madureira (2007), quando analisam a interpretação dada por um ator à leitura de um poema. Quatro personagens, que dialogam entre si, são analisados, sem contar com recursos cênicos nem com a presença de outros atores ou público

para interagir. Apesar de o ator, ao interpretar a personagem do narrador, usar um ajuste fonatório modal, com mandíbula aberta, quando assume os pensamentos de outras personagens, e em situações que interpreta semanticamente o significado de uma palavra, realiza ajustes, em que usa a voz sopro, para criar um clima de suspense e dramaticidade ou expressar o pensamento ansioso do pai. A voz do ator muda totalmente, quando, ao saber da atitude covarde do filho de chorar frente à morte, enfurece, e, nesse momento, produz uma voz com hiperfunção, aspereza e crepitação. As autoras concluem que, embora os ajustes de voz usados pelas diferentes personagens podem ser correlacionados com determinadas emoções e atitudes, não é possível estabelecer uma relação biunívoca, pois uma mesma emoção pode ser manifestada por diferentes ajustes, combinados com outros.

Importante destacar que a qualidade de voz pode reforçar o sentido de uma determinada palavra quando há compatibilidade entre a forma sonora e o sentido, por exemplo, ao se falar a palavra “sensual” com uma voz sopro.

### **1.6. Qual a diferença entre ritmo e taxa de elocução?**

O ritmo é um termo complexo e, para simplificar, vamos pensar no ritmo cardíaco, que acontece entre a ocorrência do som do batimento do coração e o silêncio. Se esse som é registrado em sequência próxima a do silêncio, permite sabermos que o coração está impulsionando o fluxo de sangue de forma constante. Contudo, um ritmo não harmônico denuncia uma arritmia e, portanto, assinala que algo não está indo bem com aquele organismo.

Ritmo é um termo comum entre os músicos, definido no dicionário como uma série de fenômenos que ocorrem com intervalos regulares, com periodicidade. Toda peça musical, além da melodia e da harmonia tem um ritmo, que vai determinar a duração de cada som na música e também a duração dos silêncios. Uma mesma sequência de três notas iguais pode dar origem a três composições musicais diferentes, apenas pela variação do ritmo.

Se na música essa definição não é tão simples, quando se trata do ritmo da fala tudo fica ainda mais complexo. Para o fonoaudiólogo, a fala de um sujeito afetada por algum transtorno neurológico ou psiquiátrico pode ser um exemplo mais evidente do que vem a ser uma alteração de ritmo, mas ao avaliar um profissional da voz, muitas vezes, confunde o ritmo com a velocidade.

Um exemplo prático, que usamos para diferenciar esses dois aspectos (ritmo e velocidade), é bater palmas numa sequência. Essa sequência caracteriza um determinado ritmo, que pode ser executado de maneira mais rápida ou mais lenta e, aí sim, a velocidade estará diferenciando um mesmo ritmo.

O ritmo na fala é um dos elementos prosódicos em que as noções de regularidade e estruturação se interligam. As regularidades são percebidas nas unidades proeminentes da fala (entre sílabas acentuadas e não acentuadas ou entre sílabas breves e longas), e essas conferem certa estruturação à fala. As línguas são divididas em dois grupos: as de ritmo acentual (intervalo entre os acentos), como o inglês; e as de ritmo silábico (a duração da sílaba que é percebida), como o espanhol e o francês. No caso do Português, temos uma língua híbrida. A partir de uma avaliação auditiva, se considerarmos graus em uma escala contínua entre o silabado e o acentual, o ritmo dos falantes gaúchos seria mais silabado do que o do paulistano. Analogamente, o mesmo acontece entre o português do Brasil (menos acentual) e o de Portugal (mais acentual). (Barbosa, 1999).

O termo velocidade tem sido substituído por taxa de elocução, por ser uma melhor tradução para o termo “*speech rate*” (Barbosa, 2000). O ritmo é influenciado pela taxa de elocução e pelo estilo de fala usada, uma vez que ela interfere na duração dos segmentos, sílabas, palavras e enunciados. Ele é mais acentual nos estilos coloquiais e rápidos, enquanto tende a silábico, nos estilos formais e mais lentos. Laver (1994) diferencia taxa de elocução de taxa de articulação,

sendo que a taxa de elocução refere-se ao tempo total da fala, incluindo as pausas preenchidas, os prolongamentos de sílabas e os silêncios, enquanto a taxa de articulação refere-se somente ao tempo audível (movimentos de articuladores), o que exclui as pausas silenciosas.

Não há uma taxa de elocução única para os falantes de uma mesma língua ou para um mesmo dialeto, de forma que uns falam mais rápido que outros. Segundo Meireles (2007), falantes do dialeto mineiro apresentaram 7,5 sílabas por segundo, taxa essa que, em média, foi 17% mais rápida que as do dialeto paulista, que registraram 6,4 sílabas por segundo. Quando o autor analisou os dados considerando a variável sexo, os homens (5,8 sílabas/s) falaram, em média, 11,5% mais rápido do que as mulheres (5,2 sílabas/s).

### 1.7. Como compreender as pausas a serviço da expressividade?

Antes de responder sobre os tipos de [pausa](#), é importante defini-la. Pode-se dizer que a pausa estrutura o pensamento e acaba por melhorar a compreensão da mensagem. Ela pode ser registrada tanto na língua escrita quanto na falada. Na escrita, ela é registrada por meio da pontuação, enquanto que, na fala, se constitui na interrupção do fluxo.

Cotes (2007), ao analisar as narrações de apresentadores de telejornais, em diferentes décadas, considerando a relação entre gestos orais e corporais presentes na construção da expressividade, encontra diversos usos de recursos de expressividade. Em especial, quanto ao uso da pausa, nas narrações, chama a atenção do leitor para o fato de que programas interativos utilizam mais pausas, porque suscitam um diálogo, enquanto nos telejornais atuais o número de pausas é menor, e, no caso dos apresentadores mais antigos, havia maior respeito à pontuação presente no texto, fato que dava à fala maior formalidade.

Como outros recursos apresentados neste capítulo, a pausa também é um elemento prosódico, que veicula informações, e que, muitas vezes, torna o que é dito mais fácil de ser percebido e, conseqüentemente, compreendido (Madureira, 2004). A combinação do local de incidência da pausa, na oração, o tipo de pausa e a variação de sua duração estão a serviço do sentido que se quer atribuir ao dito.

As pausas, por exemplo, dividem o sinal de fala contínuo em grupos de palavras coerentes. Esta segmentação do fluxo de fala reduz a ambigüidade, aumenta a inteligibilidade, além de possibilitar, ao ouvinte, um tempo para processar a informação (Madureira, 2004; Viola e Madureira, 2008).

Segundo a mesma autora, a percepção de uma pausa, pode ser sinalizada por variações de *pitch*, alongamento da sílaba acentuada imediatamente anterior à pausa ou por presença de silêncio, ou, pelo conjunto destas características.

As pausas estão relacionadas à coordenação pneumofonoarticulatória, mas a demarcação de pausas, num texto, deve ir além da necessidade de pausas respiratórias. Em textos do gênero informativo, as demarcações seguem um raciocínio mais fisiológico e sintático, mas para se transmitir emoções e atitudes, a uma determinada marcação, as pausas devem corresponder à ênfase e dramaticidade que se quer empregar.

Os tipos de pausa são:

1. Silenciosa, quando há um silêncio ou o falante respira; os silêncios ou as interrupções demonstram respeito, atenção, importância, ansiedade, ou atitude de poder, do falante para com o ouvinte.
2. Preenchidas ou de planejamento, na presença do uso de expressões como *né*, *tá*, ou um alongamento da vogal para lembrar uma determinada ideia (tipo [a:], [o:], [m:]), ou preen-

chida por respiração ruidosa; cabe destacar que entre os profissionais da voz o uso de pausas preenchidas em excesso diminui a fluência da fala, dificultando melhor entendimento por parte do ouvinte.

3. Perceptiva, na presença da interrupção da fala, gerada por diferentes pistas acústicas, a saber: elevação ou diminuição de  $f_0$ , maior duração de sílabas ou mesmo alterações de qualidade de voz.

O falante calcula e delibera o uso da pausa, em seu turno de fala, e, gramaticalmente, ela marca a fronteira do grupo entoacional e coincide com a fronteira sintática, isto é, a fronteira de frase ou de constituinte, quando no interior da frase. Quando, porém, a pausa cai internamente e quebra a coerência da estrutura entoacional, a fala é interrompida, tornando o discurso não fluente ou hesitante (Viola, 2006).

Tanto as pausas de mudança de turno como as de fronteira, dos enunciados, podem apresentar função respiratória e/ou expressiva e serem silenciosas, preenchidas ou perceptivas, ou seja, pode haver acúmulo de funções.

Viola (2006), ao analisar a construção da expressividade de um ator, interpretando quatro diferentes personagens, constatou que as pausas silenciosas apresentaram, no discurso, as funções discursiva, enfática e dramática. As pausas discursivas segmentaram o texto em unidades semânticas, demarcando ou não as mudanças de tópico, ou seja, mudança em relação à pessoa ou à coisa sobre a qual se fala. As pausas enfáticas destacaram e salientaram a informação dada, enquanto as pausas dramáticas criaram uma expectativa, certo suspense ao que foi dito ou ao que ainda seria falado. As pausas mais longas foram do tipo discursivas, posicionadas em fronteira de frases, onde houve mudança de tópico, e, as mais curtas, ocorreram nas frases que indicavam a manutenção do assunto em pauta.

Para Steinberg (1988), embora as pausas silenciosas se constituam em ausência de som, a ocorrência delas pode gerar diversas interpretações, considerando o contexto em que são produzidas. Assim, o silêncio pode se relacionar a convenções de uma sociedade, ou, em situações de interação, pode representar recusa em responder, forte espanto, preceder uma tomada de decisão, exercer autoridade, regular a distância pessoal, entre outras possibilidades.

No teatro, os atores costumam realizar algumas marcações no texto e as pausas são representadas por traços simples (/) ou duplos (//), usados para representar a primeira, a pausa lógica, que divide a frase em períodos, reunindo palavras em grupos, conforme o que se quer fazer entender do texto; e, a segunda, denominada pausa psicológica, mais longa, que dá vida aos pensamentos e ajuda a transmitir o conteúdo subtextual das palavras (Gayotto, 1996).

Fraga e Jurado Filho (2014), a partir do material coletado de quatro atores de teatro que interpretaram, individualmente, um mesmo texto teatral, analisaram em que medida semelhanças e diferenças ocorrem entre os locais julgados como de ocorrência de pausas na interpretação. Os achados demonstraram que os quatro sujeitos apresentaram diferentes quantidades de pausa no trecho em comum que interpretaram (variando de seis a 14 ocorrências), e cinco locais de pausa em comum.

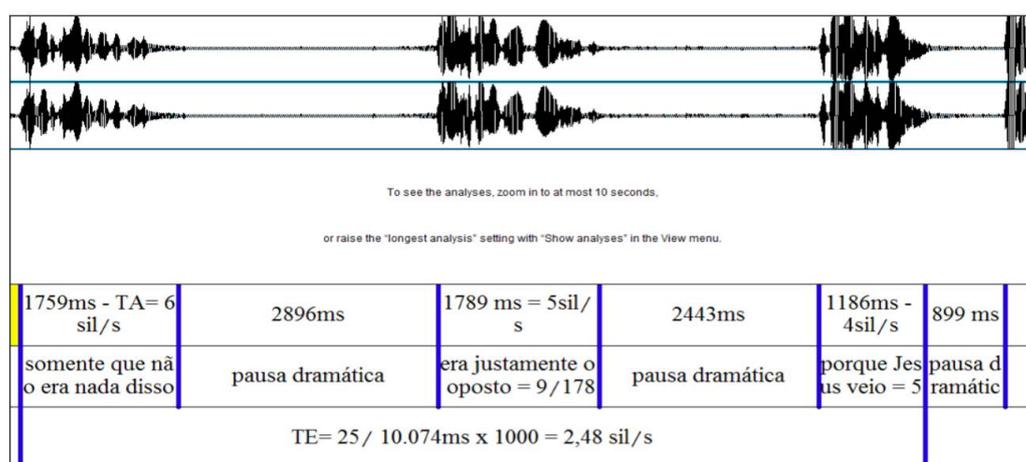
Em locutores, a distribuição das pausas é mais homogênea em comparação aos não locutores (Medrado *et al.*, 2005); e nos idosos, a presença de hesitações e pausas, em lugares anormais das frases, é característico de sua linguagem.

Viola e Madureira (2008) propõem um roteiro para analisar, de forma mais detalhada, as pausas presentes num enunciado (Quadro 1).

**Quadro 1 – Proposta de classificação das pausas, segundo Viola e Madureira (2008).**

Divisão	1. Fluxo da fala		
	2. Estrutura do falante	2.1. Monólogo	
		2.2 Interação entre falantes	2.2.1 Mudança de turno
Fisiológica	Respiratória	Controlada pelo falante	1. Troca de turno 2. No final de grupo entoacional
Fato gramatical	Dentro da fala de um falante	1. Marca fronteira de grupo entoacional	
		2. Coincide com fronteiras sintáticas	2.1. Dentro e entre sentenças 2.2. Queda interna na sentença e quebra Fluência ou hesitação
Estrutural	1. Silenciosa	Qualquer silêncio com o mínimo de duração de 200 ms	
	2. Preenchida	Presença de conteúdo não linguístico	1. Alongamento de partes da palavra
			2. Sons como [m:] 3. Inspiração com ruído audível
	3. Sinalizada por propriedades fonética- acústicas	3.1. Alongamento 3.2. Mudança de qualidade de voz 3.3. Variação de f0	Duração silábica longa, principalmente em sons plosivos.
			Alteração de qualidade de voz no final dos enunciados
Queda abrupta			
Distribucional	Dentro ou entre sentenças ou outros constituintes morfo-sintáticos		
Funcional	1. Respiratória		
	2. Discursiva	Planejamento do discurso ou estruturação de partes	
	3. Expressiva	Dramática, Enfática e Reflexiva	
Temporal	1. Comprimento		Curta ou longa
	2. Intervalos		Regular ou irregular

O exemplo, a seguir, ilustra três pausas dramáticas, sequenciais, de tamanhos diferentes, de um mesmo discurso.



**Figura 1 –** traçado da forma da onda e camadas de segmentação “somente que não era nada disso”. “Era justamente o oposto”. “Porque Jesus veio”, mostrando a duração dos enunciados, das pausas e as taxas de elocução e articulação.

## 1.8. A articulação pode ser vista como marcador de expressividade?

A fala é composta pelos segmentos fonéticos (vogais e consoantes) e pela voz (qualidade e dinâmica), que constituem unidades de sentido, chamados, por Bakhtin (1997), enunciados. Um enunciado pode consistir de apenas um fonema, por exemplo, a palavra monossilábica “é” (do verbo ser) ou de uma frase com vários elementos.

A articulação refere-se à posição anatomofisiológica que os diferentes órgãos (língua, lábios, palato etc.) assumem durante a fonação das vogais e consoantes, sendo que algumas variações no grau de (im)precisão dos pontos articulatórios (ou alvo articulatório), no grau de tensão muscular (mais ou menos força) e na duração dos sons, podem ser entendidos como distorções, sotaques, estilo de um sujeito ou de um grupo ou, ainda, um gesto expressivo.

Tanto as características fonéticas quanto as prosódicas carregam elementos simbólicos, que podem ser exploradas expressivamente pelo orador, da mesma forma que a escolha do vocabulário, que não é somente uma opção social.

As chamadas propriedades sinestésicas relacionam-se às associações acústicas dos segmentos sonoros com outras dimensões da sensação e da percepção (auditiva, visual e tátil). Daí a denominação de vogais e consoantes como: oclusivas, sibilantes, chiantes, vibrantes e aspiradas. O valor expressivo também pode estar no uso da forma incomum, reduplicada ou marcada de um ou mais segmentos<sup>2</sup>.

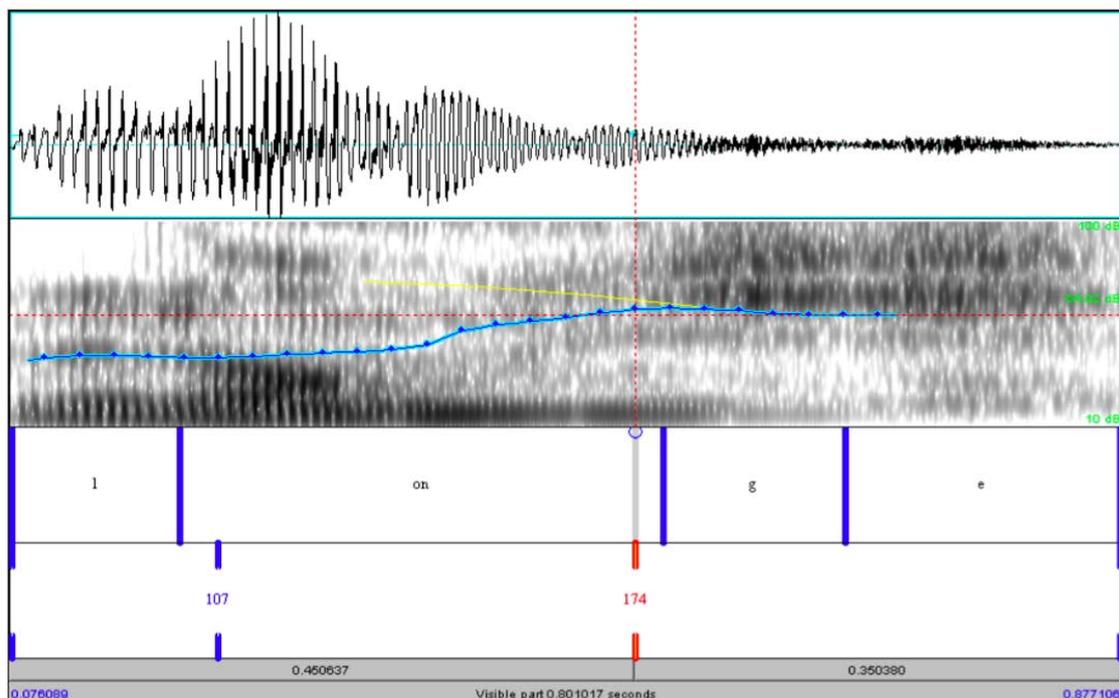
Alguns exemplos:

1. Uma fala com ceceio, se não for um traço da língua ou parte dos hábitos sociais da fala da comunidade na qual o sujeito está inserido, pode inconscientemente simbolizar traços de personalidade.

2. A variação  $f_0$  e a intensidade podem ocorrer conforme a pressão exercida nos articuladores, como nas oclusivas, que se alongam, e emprestar ao discurso um caráter “*staccato*” e duro (Fónagy, 1983).

3. Um alongamento vocálico ou consonantal (a plosão ou fricção) se apresentado com entonação desnivelada é mais perceptível do que quando nivelado (Fónagy, 1983), por exemplo, vixe nivelado e viiii iiii iiii xe.

4. O alongamento de sons pode ser usado como recurso para intensificar as palavras, e, num modo imitativo, reforça a própria palavra que está sendo enunciada. Quando aumentado, sugere lentidão, quando diminuído, rapidez. Se combinado com o *pitch* monotônico e baixo, torna-se um ingrediente de histórias de fantasma (Bolinger, 1985). Um exemplo dado por Viola (2006) ilustra o reforço. A palavra alongada é “longe”, o alongamento vocálico do segmento “on” de 348 ms e contorno ascendente de  $f_0$  (107 –174 Hz) reforça a distância, dando a impressão de ser demasiadamente longe (Figura 2).



**Figura 2** – Traçado da forma da onda, espectrograma de banda larga e camadas de segmentação do enunciado “longe”, mostrando a duração dos segmentos e os contornos de f0 e da intensidade.

5. Por fim, Madureira (2004) nos propõe a imaginar possíveis maneiras de se pronunciar SIM em resposta a uma pergunta do tipo VOCÊ GOSTOU DO FILME? Imaginemos as possibilidades de respostas:

- a. Você não gostou muito e quer dar uma resposta evasiva: uma das maneiras de demonstrar isso seria alongar um pouco a vogal e modificar, gradativamente, o tom em direção ascendente;
- b. Você gostou e quer afirmar isso categoricamente: a direção do tom seria, provavelmente, descendente e mais abrupta;
- c. Você gostou e se entusiasmou com o filme: o segmento seria alongado, com uso de tom ascendente-descendente;
- d. Você não está a fim de responder: seria registrada uma qualidade de voz utilizando um ajuste de mandíbula fechada, com mudança gradativa do tom, em direção descendente e alongamento da duração do segmento vocálico.

### 1.9. O que observar e como proceder na avaliação de expressividade?

A voz, sendo um índice gestual (gesto vocal), responde às dimensões corporais, simbólicas, linguísticas e estilísticas e, por ser dinâmica, flexível e adaptável, não pode ser analisada por elementos ideais e normativos (Viola, 2006).

Quando se trata de avaliar a expressividade, a primeira questão a ser considerada é alinhar a avaliação a pressupostos teóricos e metodológicos, ou seja, partir da concepção de linguagem, que definirá como serão abordadas as relações entre forma e conteúdo, na fala.

<sup>2</sup> Viola (2006) faz uma revisão da literatura e apresenta a tipologia e as formas do simbolismo sonoro, propostas por Hilton et al. (1994).

Numa concepção de linguagem em que forma e conteúdo da língua são tomados de maneira independente, a manifestação expressiva, também chamada estilo, é tratada como uma característica particular do indivíduo, que se desviou de uma norma ou de um contexto. Quando, entretanto, concebida a língua como dialógica, histórica e variável, forma e conteúdo se entrelaçam, na expressividade oral, e esta exibirá a subjetividade e a singularidade do locutor ou do grupo social ao qual ele se insere. A expressividade apresentará características orais, corporais e discursivas recorrentes e salientes, no discurso, e estas irão gerar efeitos próprios de sentido.

Embora, neste capítulo, não nos atenhamos às questões da expressividade corporal, é importante destacar que, nos processos comunicativos, não há cisão entre corpo e fala, pois, na fala, as informações indexicais interagem e são co-produtoras de um sistema total de comunicação. Enquanto parte do sistema total de signos expressivos, os gestos vocais podem ou não se apresentar em harmonia aos gestos corporais e faciais e, portanto, a compatibilidade e discrepância dos vários domínios da gestualidade irão denotar conflito ou simulação do indivíduo.

Duas direções simbólicas são orientadoras da compatibilidade ou discrepância entre som e sentido e os vários recursos acústicos, segundo Viola (2006):

1. Pequeno, alto, rápido, claro, longo, fino, tenso, forte, pontudo, início, feminilidade, alegria, prazer, atitude não ameaçadora, polida e submissa.

2. Grande, baixo, lento, escuro, curto, grosso, relaxado, fraco, arredondado, fim, masculinidade, tristeza, desprazer, atitude ameaçadora, agressiva e autoritária.

Por exemplo: o uso de *pitch* grave é compatível, simbolicamente, com atitude ameaçadora ou autoritária (grupo 2), enquanto um *pitch* agudo, com alegria ou atitude submissa (grupo 1).

Ao se avaliar a interação do corpo, oralidade e discurso e o alinhamento entre eles, essas duas direções revelarão os sentidos, como a ironia, que se manifesta na incompatibilidade da informação semântica com a qualidade e dinâmica de voz ou a dúvida na veracidade da mensagem a ser dada pela discrepância entre as informações falada e visual. Num exemplo concreto, uma vendedora, ao falar dos benefícios de um determinado produto, pode fazê-lo com voz aguda, forte e rápida, evidenciando uma qualidade positiva dele, porém, se, ao mesmo tempo, ela expressar um sorriso “de lado” ou sobrancelhas franzidas e gestos tensos, estes denotarão que o produto não é tão positivo como apresentado.

Souza CCL (2007), ao caracterizar a expressividade de quatro vendedoras de planos de saúde, no momento da negociação, concluiu, com base no julgamento pessoal de 31 possíveis clientes, que a comunicação não-verbal foi a que mais chamou a atenção, valendo a primeira impressão expressa por ela.

Outro aspecto importante a ser mencionado é que a existência de expressões vocais específicas (por exemplo, alegria e tristeza) indicam que há um jogo diferenciado de perfis vocais, que mudam com o resultado da avaliação do indivíduo e da ativação do organismo. Isto significa que não há cor, uma a uma, entre expressão vocal e parâmetros avaliados. Em outras palavras, buscar perfis vocais para as emoções é ineficiente, além de demandar rigorosos cuidados com a nomenclatura dada às emoções e atitudes (Viola, 2006).

A avaliação, num ambiente de atendimento clínico ou de assessoria, de profissionais da voz, pode apresentar diferenças se comparada à avaliação feita num contexto de pesquisa. Essas diferenças estão presentes no momento da coleta, seleção e análise dos dados.

Após apresentar essas questões conceituais, um aspecto a ser destacado é a coleta de amostra da fala. Esta deve representar um contexto habitual, em que o sujeito esteja se comunicando, que pode ser com o próprio fonoaudiólogo ou em seu ambiente profissional. Coletar diversas amostras, em diferentes contextos, pode auxiliar o fonoaudiólogo a analisar melhor os dados, con-

siderando o que é repetitivo. Duas possibilidades antagônicas se apresentam: no sentido estilístico, seria uma marca de estilo, que pode ou não ser mantida, ou, no sentido negativo, repetitiva por hábito inconsciente, exemplo de locutores que terminam todas as frases com entonação em direção ascendente.

É importante lembrar que para compreender a expressividade, de forma mais objetiva, e, ao mesmo tempo, atender ao rigor científico, além da análise perceptivo-auditiva, o uso da fonético-acústica, como recurso instrumental, complementa e amplia o entendimento do material. A análise fonético-acústica fornece evidências ao dito e auxilia na interpretação dos dados, uma vez que possibilita um “ouvir” melhor e oferece um registro possível de ser “visto e revisto”, quantas vezes forem necessárias. Assim sendo, cuidar para que o registro seja feito da melhor forma possível, pode auxiliar na análise feita a seguir. Gravações *in loco* nem sempre cumprem com exigências para a análise do sinal acústico, que deve ser coletado com gravador de boa qualidade e em ambiente silencioso.

O primeiro passo, para a análise de um material de fala, deve ser repetidas escutas, para se familiarizar com a interpretação do locutor e a percepção dos elementos salientes e de suas relações com o sentido veiculado. O segundo passo, será mapear o material relacionando os tipos de variantes segmentais utilizadas, o emprego de elementos prosódicos e as relações entre o som e os sentidos, onde emergem as atitudes e emoções. Deste mapeamento geral, surgirão partes mais salientes, que serão destacadas no terceiro passo e inspecionadas, mais profundamente. Num trabalho de assessoria, este material servirá para exemplificar ao cliente quais elementos precisam ser ajustados, revistos ou reforçados. Num trabalho de pesquisa, o material ainda passará por inúmeras inspeções, de acordo com o objetivo do trabalho científico.

A interpretação dos dados deve ser feita com cuidado. Primeiro, por ser habitual o fonoaudiólogo, considerando sua natureza mais clínica, voltar-se para o levantamento dos *déficits* apresentados pelo sujeito. Para tal avaliação, é comum fazer recortes de partes das informações coletadas, utilizando instrumentos que favorecem o registro restrito de presença/ausência de um determinado parâmetro, fato que conduz a avaliar a fala fora de seu contexto. O segundo, e o mais comum, é a contaminação dos desejos interpretativos do interpretante e a própria limitação do ouvido humano, em captar alguns eventos da fala.

Um instrumento para avaliar o gesto vocal não deve se restringir a itens fechados, pois, desta forma, a consequência será uma avaliação mais pontual do que global. Na literatura, Laver (2000) apresenta uma proposta de avaliação de Camargo e Madureira (2008b), cujo protocolo é usado internacionalmente. No Brasil, vários autores propõem instrumentos ou sugerem tópicos.

Apresentamos um roteiro para orientar a avaliação dos aspectos presentes na expressividade oral.

## **2. Roteiro para Avaliação da Expressividade**

### **2.1. Discurso**

1. Contexto e gênero do discurso;
2. Uso inadequado do idioma: erros gramaticais (“de modo que”, “a nível de”, aglutinações do tipo “eu vou ca menina”), concordância verbal, concordância nominal, conjugação verbal;
3. Vocabulário: uso de gírias, pouca variabilidade, rebuscamento;
4. Escolhas lexicais regionais;
5. Figuras de linguagem;
6. Estereotípias (uso repetitivo): né? tá? ok? querida, fofa, amor, gracinha, tudo bem, que lindo, ótimo;

7. Expressões que menosprezam a capacidade do ouvinte: Entendeu? Percebeu? Vou simplificar para você;
8. Alterações fonológicas.

## 2.2 Expressividade Oral

I. Características de longo prazo ou quase predominantes, em situação coloquial e na atividade profissional mais importante:

- Qualidade de voz
- *Loudness* habitual
- *Pitch* habitual, variabilidade! Taxa de elocução
- Alongamentos de segmentos! Segmentos marcados
- Sotaque (ou regionalismo)
- Pausas
- *Pitch* (curvas de entoação)

II. Características de curto prazo para marcar expressão de emoção ou atitude (variabilidade, combinações e contrastes de recursos):

1. Qualidade de Voz
2. *Loudness*
  - a. Variabilidade x repetição
  - b. Extensão
  - c. Compatibilidade ou discrepância com a semântica veiculada
  - d. Compatibilidade ou discrepância com a emoção
  - e. Compatibilidade ou discrepância como grau de formalidade do contexto
  - f. Compatibilidade ou discrepância com a distância do interlocutor
3. *Pitch*
  - a. Variabilidade x repetição
  - b. Extensão
  - c. Predominância do tipo de contorno (ascendente, descendente; ascendente/descendente e vice versa, nivelado)
  - d. Compatibilidade ou discrepância do contorno com a semântica veiculada
  - e. Compatibilidade ou discrepância do contorno com a emoção
  - f. Compatibilidade ou discrepância do contorno com o grau de formalidade do contexto
4. Taxa de elocução
5. Segmentos e sons corporais
  - a. Respiração ruidosa
  - b. Bocejo
  - c. Pigarro
  - d. Estalos de língua ou lábios
  - e. Alongamento de vogais com fins expressivos
6. Segmentos marcados – trocas, distorções e possível recorrência estilística
7. Articulação – grau de abertura (de travada a sobrearticulada)
8. Pausas:
  - a. Inter e intra pessoal (pausas de mudanças de turno)
  - b. Função – respiratória, discursiva, enfática, dramática, reflexiva
  - c. Distribuição ao longo de uma fala contínua – regular e irregular
  - d. Tipo – respiratória e/ou expressiva;

e. Modo – silenciosa, preenchida ou perceptiva.

f. Duração – curta à longa

III. Há predominância de algum do tipo de recurso para se dar ênfase (ou proeminência)? Qual?

IV. Presença de traço estilístico oral

#### 10 - Qual a diferença entre oratória e expressividade?

Oratória não é sinônimo de expressividade oral. Oratória se refere à arte do bem falar a um determinado público e expressividade é a ação de expressar emoções e atitudes. O quarto componente do sistema retórico (hoje, equivalente à argumentação), proposto pelos gregos e romanos, é o que se aproxima da expressividade oral e corporal, na expressão das atitudes: *actio*<sup>3</sup> é a “ação que atualiza o discurso, a sua execução e constitui o próprio alvo da Retórica” (Mosca, 1997), que é convencer pelo raciocínio e persuadir com base na emoção, por meio dos elementos suprassegimentais e da gestualidade.

Assim, podemos pensar na oratória como um conjunto de procedimentos que prepara o locutor para a consecução de um trabalho, de uso voluntário e controlado da sua expressão oral, contudo, que pode, de forma, diríamos, traiçoeira, ser subjugado por emoções, de caráter incontrollável e involuntário, como os famosos brancos, a gagueira, o tremor na voz, a sudorese etc.

A Linguística, de hoje, atualizou e fez a revisão dos conceitos antigos, através de novas metodologias e teorias, como as análises estilísticas, as teorias discursivas e a teoria da argumentação. A diferença entre a retórica da Antiguidade e a de hoje está na importância posta no orador e no receptor da mensagem. Hoje, é mais importante conhecer os diferentes contextos e selecionar os materiais para a construção retórica, adequados a cada situação, aproximando-se da realidade do público, de modo que ele se identifique com o comunicador e sua proposta (Mosca, 1997).

É preciso evitar tratar da expressividade oral de modo fragmentado e sem referência à linguagem, em seu uso efetivo, com enaltecimento das normas e dos padrões preconizados coletivamente, como se todos os oradores se expressassem da mesma forma e pudessem se beneficiar das mesmas regras e normas.

Bons oradores não necessariamente usam dos mesmos recursos vocais e corporais e podem, inclusive, apresentar perfis muito diferentes, conforme mostrou a pesquisa de Panico (2001), com três senadores da República, escolhidos por jornalistas, como os melhores oradores da bancada, na época. Todos conhecidos pela capacidade de argumentação, inteligência e sagacidade, mas com perfis vocais diferentes: 1. prolixo, exagerado na quantidade de gestos articulação nem sempre definida; 2. altamente intelectualizado e erudito, poucos gestos corporais, qualidade de voz modal com *loudness* fraco com articulação e pausas bastante definidas; e, por fim, 3. intermediário, se assim podemos definir, pelo passado de radialista, com qualidade de voz modal de peito, com boa projeção, muito marcante.

Tendo em mente a variabilidade da expressão, é evidente que os primeiros passos, para promover mudanças mais profundas, será adquirir um conjunto de informações importantes para desenvolver a atividade profissional, como a conscientização de fatores corporais (in)adequados em uma apresentação pública, das variações da fala nos contextos de interlocução (como público, domínio do assunto etc.) e da compreensão dos efeitos da sua expressão sobre os ouvintes (o

---

<sup>3</sup>As partes retóricas são: *inventio* (sujeitos, argumentos, lugares, técnicas de persuasão e de amplificação); *dispositivo* (arrumação das grandes partes do discurso: exórdico, narração, discussão, peroração); *elocutio* (escolha da disposição das palavras na frase, organização em pormenor); *pronuntiatio* ou *actio* (enunciação do discurso) e *memoria* (memorização).

que as pessoas pensam da minha fala). Contudo, esta prática atua no plano cognitivo, que, similar às técnicas das terapias cognitivistas, poderá causar impacto no inconsciente (na expressão das emoções), de modo a minimizar as reações fisiológicas ativadas pelo medo de falar em público.

## Referências Bibliográficas

- BAKHTIN, M. M. **Estética da criação verbal**. São Paulo: Martins Fontes, 1997.
- BARBOSA, P. A. Revelar a estrutura rítmica de uma língua construindo máquinas falantes: pela integração de ciência e tecnologia de fala. In: SCARPA, E (Org). **Estudos de prosódia**. Campinas: Unicamp, 1999.
- \_\_\_\_\_. **Syllable-timing in Brazilian Portuguese: uma crítica a Roy Major**. São Paulo: D.E.L.T.A.; 6 (2) p. 369-402, 2000.
- BELLER, G. Transformation of expressivity in speech. **Ling. Insights**: 97: 259–284, 2009.
- BOLINGER, D. **Intonation and its parts: melody in spoken English**. Edward Arnold, 1985.
- BOONE, D. R. **Sua voz está traindo você?** Porto Alegre: Artes Médicas, 1996.
- CAMARGO, Z. A; MADUREIRA, S. **Avaliação vocal sob a perspectiva fonética: investigação preliminar**. *Distúrb Comun*, São Paulo, 20. V. 01: 77-96, ab, 2008.
- \_\_\_\_\_. Voice quality analysis from a phonetic perspective: Voice Profile Analysis. Scheme Profile for Brazilian Portuguese (BP-VPAS). In: **Proceedings of the Speech Prosody**, Campinas: p. 57-60, 2008.
- CHUNG, S. J. **L'expression et la perception de l'émotion extradite de la parole spontanée: evidences du coréen et de l'anglais**. Tese. Paris: U de la Sorbonne Nouvelle, 2000.
- COTES, C. O uso das pausas nos diferentes estilos de televisão. In: *Rev CEFAC*. São Paulo: 9, v. 02, 2007.
- FANT, G. **Acoustic theory of speech production**. Paris: Mouton, 1970.
- FERREIRA, L. P. **Expressividade: a trajetória da Fonoaudiologia brasileira**. In: KYRILLOS, L. R. *Expressividade: da teoria à prática*. Rio de Janeiro: Revinter. 2004.
- FÓNAGY, I. **La vive voix**. Paris: Payot, 1983.
- FRAGA, M; JURADO FILHO, L. C. **Distribuição de pausas na interpretação de atores**. *Distúrb Comun.*, 26, v. 03, 2014. Disponível em: <<http://revistas.pucsp.br/index.php/dic/article/view/16175>>
- GAYOTTO, L. H. C. **Voz do ator a partitura da ação**. Dissertação de Mestrado. São Paulo: PU-CSP, 1996.
- HILTON, L.; NICHOLS, J.; OHALA, J. Introduction: sound-symbolic process. In: HILTON, L.; NICHOLS, J.; OHALA, J. (Eds)s. **Sound symbolism**. Cambridge: Cambridge University, 1994.
- LAVIER, J. Phonetic evaluation of voice quality. In: KENT, R. D.; BALL, M. J. **Voice Quality Measurement**. San Diego: Singular Publishing Group, 2000.
- \_\_\_\_\_. **Princ of phonetics**. Cambridge: Cambridge University, 1994.
- \_\_\_\_\_. **The phonetic description of voice quality**. Cambridge: Cambridge University, 1980.
- MADUREIRA, S. Expressividade da fala. In: KYRILLOS, LR. **Expressividade: da teoria à prática**. Rio de Janeiro: Revinter, 2004.
- \_\_\_\_\_. The Investigation of Speech Expressivity. In: MELLO, H.; PANUNZI, A.; RASO, T. (Eds). (Org.) **Illocution, modality, attitude, information patterning and speech annotation**. Firenze: Firenze University Press, v. 01, p. 101-118, 2011.
- MARQUEZIN, D.M.S.S; et al. **Análise da dinâmica vocal de executivos**. *CoDAS*; 27, v. 02; p. 160-9. 2015.
- MEDRADO, R.; FERREIRA, L. P.; BEHLAU, M. Voice-over: Perceptual and Acoustic Analysis of Vocal Features. **J of Voice**. 19, v. 03; p. 40-349, 2005.
- MEIRELES, A. R. **Reestruturas rítmicas da fala no português brasileiro**. Tese de doutorado.

Campinas: IEL, UNICAMP, 2007.

MICHAELIS – **Dicionário Brasileiro da Língua Portuguesa**. Editora Melhoramentos Ltda., 2016. Disponível em: <<http://michaelis.uol.com.br/busca?r=0&f=0&t=0&palavra=enfase>>.

MOSCA, L. L. S. Velhas e novas retóricas: convergências e desdobramentos. In: **Retóricas de ontem e de hoje**. São Paulo: Humanitas, p. 29, 1997.

PANICO, A. C. M. C. **A voz no contexto político análise dos recursos vocais e gestuais no discurso de senadores**. Dissertação de Mestrado. São Paulo: PUCSP, 2001.

PITTAM, J.; SCHERER, K. R. Vocal expression and Com of emotion. In: LEWIS, M., HAVILAND, J. M., editors. **Hdbk of emotions**. New York: Guilford, 1993.

REIS, C. Prosódia e telejornalismo. In: GAMA, A. C. C.; KYRILLOS, L.; FEIJÓ, D. (Org.) **Fonoaudiologia e Telejornalismo. Relatos do IV encontro nacional de Fonoaudiologia da Central Globo de Jornalismo**. Rio de Janeiro: Revinter, 2005.

SANTOS, F. A. **Caracterização da voz de comando de oficiais de polícia militar da academia de polícia militar do Barro Branco**. Monografia. São Paulo: Centro de Estudos da Voz, 2005.

SOUZA, C. C. L. **Identificação e Caracterização da Expressividade de Vendedoras de Planos de Saúde no Momento da Negociação**. Dissertação de Mestrado. São Paulo: PUCSP, 2007.

SOUZA, L.M.C. **A Prosódia no Comando Militar**. Dissertação. Belo Horizonte: Fac Letras UFMG, 2007.

STEINBERG, M. **Os elementos não-verbais da conversação**. São Paulo: Ed Atual, 1988.

VIOLA, I.C. **Expressividade, estilo e gesto vocal**. Lorena: Inst Santa Teresa, 2008.

\_\_\_\_\_. **O gesto vocal: a arquitetura de um ato teatral**. Tese de Doutorado. São Paulo: PUCSP, 2006.

VIOLA, I.C.; MADUREIRA, S. The roles of pause in speech expression In: **Speech Prosody 2008**, 2008. Campinas. **Proceedings...** RG, p.721 – 724, 2008.

\_\_\_\_\_. Voice quality and Speech Expressiveness. In: **ParaLing. Speech07- between models and data**, Saarbrucken. **Proceedings...** Saarbrucken: DFKI Publication, p. 65-70, 2007.

# *A Fala do Intérprete*

LAYLA PENHA

## 1. Introdução

Pesquisas sobre qualidade em interpretação simultânea desenvolvidas a partir de 1986 mostram que embora uma fala expressiva seja característica desejável na produção do intérprete, as questões prosódicas assumem papel secundário na percepção de qualidade que intérpretes e ouvintes sobre do trabalho de interpretação.

A percepção de qualidade de uma interpretação é afetada por muitos fatores subjetivos, incluindo a expectativa, formação e papel de cada agente em uma situação de interpretação (palestrantes, público, organizadores e intérpretes) (Hartley et al, 2003). Ademais, precisamos levar em conta a situação onde acontece a interpretação. Variáveis como o tamanho do evento, nível de formalidade e/ou técnico e nível de integração entre os participantes, entre outros, também podem afetar a percepção de qualidade da interpretação (Kahane, 2000). Pouquíssimos estudos sobre a percepção de qualidade na interpretação simultânea abordam tais variáveis, motivo pelo qual seus resultados não podem ser considerados absolutos. Isso, no entanto, não diminui a importância de conhecermos as pesquisas desenvolvidas na área até hoje e sua evolução ao longo do tempo.

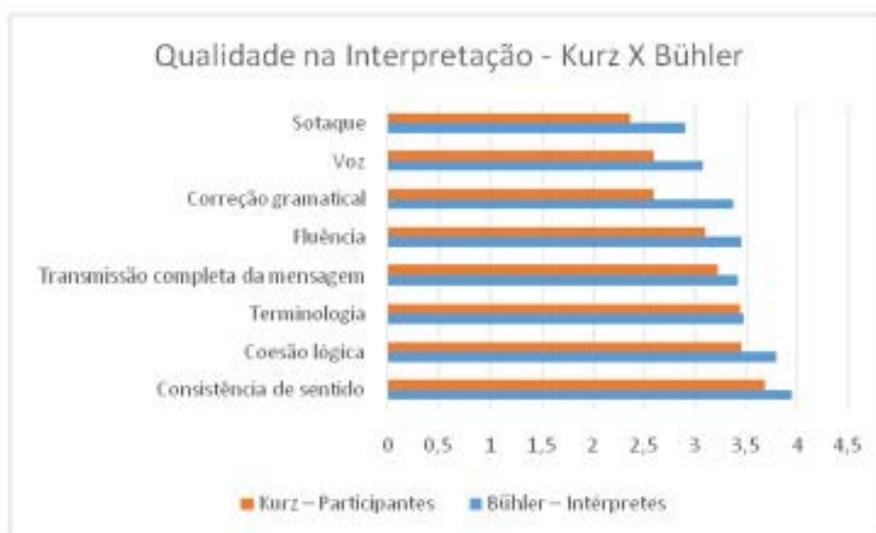
O primeiro estudo sobre qualidade na interpretação foi conduzido por Bühler (1986). Baseou-se em um questionário de avaliação a ser respondido pelos membros da Associação Internacional de Intérpretes de Conferência (AIIC) que objetivava levantar os critérios utilizados por estes para avaliar a qualidade de uma interpretação. Foram incluídos 17 parâmetros, que deveriam ser classificados em ordem de importância em uma escala de 1 a 4 (irrelevante/muito importante) – sotaque nativo; voz agradável; fluência de discurso; coesão lógica do enunciado; consistência de sentido com a mensagem original; transmissão completa da mensagem original da interpretação; correção gramatical; utilização de terminologia correta; utilização de estilo apropriada; preparação meticulosa do material da conferência; resiliência na interpretação; postura; aparência agradável; confiabilidade; capacidade de trabalhar em equipe; retorno positivo dos participantes da conferência; outros. É interessante observar que os nove primeiros critérios se referem à qualidade da interpretação em si, enquanto os seis restantes fazem mais referência ao trabalho de preparação do intérprete ou, como coloca Collados (2001,p.37) “ao processo de como alcançar qualidade”.

Pouco se conhece a respeito do perfil dos intérpretes que responderam ao questionário, distribuído em uma reunião de membros da AIIC, à exceção do número total de respondentes, quarenta e sete, incluindo aí sete membros da CACL – Comissão de Admissão e Classificação Linguística, em teoria intérpretes mais experientes no processo de avaliação de qualidade de interpretação. No entanto, o trabalho de Bühler é seminal. Além de ter sido o primeiro, muitos de seus parâmetros, nomeadamente aqueles mais voltados à percepção de qualidade da produção do intérprete em si, foram reproduzidos em estudos posteriores, o que nos permite comparar a importância atribuída a cada item ao longo do tempo e por públicos diferentes. Isso nos parece fundamental porque, como afirma Seleskovitch (1986, p.236) a interpretação deve sempre ser avaliada da perspectiva do ouvinte alvo e nunca como um fim em si mesmo, afinal, “a cadeia da comunicação não termina na cabine.”

A primeira oportunidade para tal acontece em 1989, com a pesquisa de Ingrid Kurz, que utiliza oito dos critérios de Bühler e os aplica em um questionário a ser respondido, desta vez, por 47 usuários dos serviços de interpretação, no caso, participantes de um congresso médico, com achados bastante consistentes com a pesquisa original. Para ainda verificar se diferentes públicos reagiriam de forma diferente aos mesmo critérios, ou se mudariam sua ordem de importância, Kurz volta a aplicar a mesma pesquisa em 1993, primeiro a um grupo de 29 participantes de um congresso sobre Controle de Qualidade e posteriormente a 48 participantes de uma reunião do Conselho da Europa.

O gráfico da Figura 1 mostra os resultados acumulados das três pesquisas de Kurz em com-

paração aos achados de Bühler, listados em ordem de importância. Dois pontos chamam imediatamente nossa atenção nos resultados. O primeiro é que os intérpretes sistematicamente atribuem maior pontuação (e consequentemente, maior importância) a cada critério questionado, levando a crer que os intérpretes têm um nível de exigência em relação ao processo de interpretação superior às expectativas de seu próprio público alvo. O segundo é que, independentemente dos tipos de respondente da pesquisa – sejam eles de diferentes públicos ou intérpretes – a ordem de importância dos critérios é exatamente a mesma, com consistência de sentido com a mensagem original sendo o parâmetro mais valorizado, seguido por coesão lógica (muito importantes). Aspectos mais ligados à fala em si, como sotaque nativo e voz agradável, aparecem em último lugar em ambas as listas.



**Figura 1** – Avaliação dos critérios de qualidade para a interpretação simultânea por intérpretes e participantes de conferência

Na mesma época em que Kurz desenvolvia sua pesquisa entre diferentes públicos, temos o questionário de qualidade de Daniel Gile, aplicado em 1990 em um congresso de Oftalmologia, com interpretação para o inglês e o francês. Também trabalhando com a percepção de usuários, desta feita em um ambiente real de conferência, Gile elenca seis critérios a serem pontuados pelos usuários do serviço de interpretação em uma escala de 1 a 5 (muito ruim a muito bom): qualidade global da interpretação, qualidade linguística, qualidade terminológica, fidelidade e voz, ritmo e entoação. Também adiciona dois pontos, principais defeitos da interpretação e outros comentários, a serem preenchidos de forma eletiva pelos participantes. Vinte e três participantes respondem ao questionário, dezoito ouvintes do inglês para o francês e cinco do francês para o inglês. Os resultados foram bastante homogêneos, com os ouvintes da interpretação para o inglês atribuindo pontuações maiores aos intérpretes em todos os quesitos. Entre os que ouviram a interpretação para o francês, as menores notas como um todo foram atribuídas a voz, ritmo e entoação, incluindo três comentários eletivos de interpretação ‘monotônica’. Apesar disso, a impressão geral da qualidade da interpretação não foi afetada adversamente por este parâmetro Gile (1990). A hipótese de Gile para este resultado é que talvez, para um público mais técnico-científico, as questões de voz assumirem menor importância, “é possível formular a hipótese de que o público mais científico (e técnico) é menos sensível à qualidade de voz, ritmo e entoação da interpretação que outros públicos, que lhe podem atribuir maior importância.” (Gile, 1990, p.68)

Entre 1993 e 1994, a pedido do Comitê de Pesquisa da AICC, Jennifer Mackintosh lidera um estudo para entender as expectativas de qualidade de diferentes usuários (Moser, 1995). Uma equipe de 94 intérpretes entrevistou mais de 200 participantes em 84 conferências diferentes realizadas no mundo todo. Praticamente metade dos entrevistados participavam apenas como ouvintes das

mencionadas conferências (50,5%), enquanto a outra metade participava como ouvinte e palestrante (49,5%). Apenas 9% utilizava os serviços de interpretação pela primeira vez, com 60% tendo participação frequente em conferências multilíngues com interpretação. Novamente, os resultados seguem a mesma tendência de estudos anteriores, com maior importância sendo atribuída aos critérios relacionados ao conteúdo do que à forma, independentemente do tipo de conferência. O critério fidelidade da interpretação ao sentido original foi o parâmetro mencionado mais vezes espontaneamente pelos participantes, com 45% o julgando o mais importante. No entanto, questões como fala vivaz e não monótona, pronúncia clara e ‘uma voz que soa natural’ foram mencionadas por 66 dos participantes como algo que um bom intérprete deve produzir. Em contrapartida, oito participantes criticaram a entoação exagerada e fala histriônica dos intérpretes, indicando que os intérpretes não deveriam se identificar em excesso com o conteúdo emocional do discurso original. De fato, o desejo de ouvir uma fala mais neutra em relação à mensagem sendo interpretada é verbalizado por 21 participantes. Outro fator relacionado às questões da fala criticado adversamente por 30 participantes é o mau uso do microfone aberto, com vazamento de barulhos da cabine, incluindo tosse, cochichos entre intérpretes, risadas, papéis sendo folheados, batuque de caneta ou lápis na mesa ou microfone entre outros. Finalmente, quando submetida aos participantes uma lista de possíveis fontes de irritação, as longas pausas silenciosas ou pausas preenchidas (ums e ahs) e atraso na fala do intérprete em relação ao discurso original (falta de sincronia) aparecem em primeiro lugar, com mais de 70% dos participantes classificando-os como irritantes ou muito irritantes.

Em 1995, Kurz e Pöchhacker repetem os mesmos 8 critérios baseados em Bühler para investigar um novo público ouvinte – desta vez um grupo de representantes de TV da Alemanha e Áustria – e verificar se os critérios em relação à forma e expressão do discurso do intérprete teriam o mesmo peso e grau de importância que os públicos anteriores (Collados, 1998). Embora mais uma vez a consistência de sentido com a mensagem original e coesão lógica fossem os parâmetros mais valorizados, os resultados mostraram que esse público era particularmente sensível a critérios relacionados à voz, sotaque e fluência.

Em 1998, Collados decide explorar mais a fundo as questões da fala do intérprete, conduzindo um estudo laboratorial controlado comparando interpretações monotônicas com versão mais ‘vivazes’ do mesmo discurso. Foram elaboradas três versões a serem avaliadas: versão monotônica sem falhas de conteúdo, versão vivaz com falhas de conteúdo, versão vivaz sem falhas de conteúdo (Collados, 1998). Sistemáticamente, as interpretações com uma fala mais vivaz, independentemente da presença de alguns erros, foram consideradas superiores em comparação àquela mais monotônica, mesmo que essa apresentasse uma transmissão completa da mensagem original da interpretação, sem falhas de conteúdo. As versões foram testadas em estudos e públicos diferentes, com resultados bastante consistentes, o que nos leva a crer que, apesar de atribuírem, em teoria, maior importância a aspectos mais relacionados ao conteúdo da interpretação, quando os usuários têm a chance de ouvir versões orais de uma interpretação com conteúdo prosódico mais ou menos monótono, estes tendem a atribuir maior qualidade àquelas versões menos monótonas, sugerindo que talvez alguns elementos prosódicos sejam importantes na percepção de qualidade da interpretação.

A partir do ano 2000, uma nova leva de estudos é produzida a fim de definir qualidade na interpretação. Um dos primeiros a ter destaque no novo milênio é o estudo de Chiaro e Nocella, de 2004. Similarmente a outros estudos aqui descritos, Chiaro e Nocella conduzem um estudo empírico da percepção de qualidade na interpretação com base nos critérios de Bühler, trabalhando aqui com o mesmo público alvo do referido estudo, ou seja, intérpretes profissionais de conferência. Com uma amostra de 286 respondentes de um universo de 1000 questionários enviados, o estudo conseguiu atrair participantes de todo o mundo (embora com grande concentração na Europa e Américas, 44% e 46%, respectivamente) e foi completamente conduzido online. Através de uma

escala multidimensional, os pesquisadores conseguiram produzir um mapa perceptivo com base nos resultados de como os intérpretes classificavam uma lista de critérios linguísticos e não linguísticos de acordo com sua percepção de importância. Mostra-se aqui a maior diferença entre este e o estudo de Bühler. A tarefa dos intérpretes desta feita era a de simplesmente classificar os critérios listados em uma ordem do mais ao menos importante, em vez de simplesmente classificá-los como importantes ou não como havia sido feito no estudo dirigido por Bühler.

Apesar de diferenças importantes em alguns quesitos, a simetria dos resultados é impressionante, de novo trazendo questões linguísticas a frente de qualquer questão extralinguística. A Tabela 1 compara os resultados de Chiaro e Nocella com aqueles de Bühler com bases nas porcentagens de “muito importante” e “importante” do estudo original.

**Tabela 1** - Comparação da classificação de critérios de qualidade (Pochhacker, 2013)

Chiaro e Nocella 2004	Bühler 1986
1. Consistência de sentido	Consistência de sentido
2. Transmissão completa da informação	Coesão lógica
3. Coesão lógica	Terminologia
4. Fluência	Fluência
5. Correção gramatical	Correção gramatical
6. Terminologia	Transmissão completa da mensagem
7. Estilo apropriado	Voz agradável
8. Voz agradável	Sotaque
9. Sotaque	Estilo apropriado

Ainda seguindo os critérios de escala de classificação de importância de Bühler, Zwischenberger e Pöchhacker (Zwischenberger e Pöchhacker, 2010) conduzem um novo estudo online com intérpretes membros da AAlC. Os critérios ditos linguísticos são mantidos como no estudo original, mas alguns critérios extralinguísticos são adicionados já levando em consideração os resultados obtidos por Moser (1996) e Collados (1998).

*“Observando as pesquisas sobre as expectativas de qualidade publicadas ao longo dos anos, aumentamos a lista de critérios originais para incluir “entoação vivaz”, como particularmente estudado por Ángela Collados Aís na Universidade de Granada, e “sincronia”, que aparece como uma característica esperada na interpretação simultânea pelos respondentes da pesquisa de expectativas dos usuários patrocinada pela AAlC e realizada por Moser<sup>1</sup>. (Pochhacker, 2012)*

Um total de 704 membros da AAlC em todo o mundo participaram da pesquisa e os resultados são muito próximos aos de Bühler, com consistência de sentido em primeiro lugar, seguido de coesão lógica. Neste ponto, já sem nenhuma surpresa, encontramos novamente os critérios voz agradável e sotaque, avaliados como menos importantes. Além disso, apesar de considerados importantes em estudos anteriores, os critérios adicionados ao estudo, entoação vivaz e sincroni-

dade ficam muito próximos a ambos, recebendo nível baixo de importância.

Finalmente, ainda em 2010, dois estudos da Universidade de Viena são publicados com foco na investigação da percepção de qualidade da interpretação e compreensão da fala do intérprete aliada a sua produção vocal. O primeiro (Holub, 2010), traz a comparação de uma fala interpretada e esta mesma fala manipulada no software Praat para se tornar monotônica. Apresentada em uma configuração que mimetizava uma conferência, o estudo contava com um vídeo sendo apresentado a 63 ouvintes alunos de marketing da instituição que eram randomizados em dois grupos, cada um ouvindo uma das versões e com a tarefa de completar um questionário de compreensão e avaliação do desempenho de intérpretes. Embora não estatisticamente diferente, os resultados mostram que a interpretação monotônica dificultou a compreensão. A análise mostrou correlação entre vivacidade da interpretação e a avaliação do desempenho do intérprete como um todo. Dados sugerem que a monotonia tem impacto negativo na avaliação de qualidade da interpretação e pode ter impacto negativo na compreensão do ouvinte.

No segundo, Rennert (2010) um intérprete profissional produz várias versões de interpretação de discurso do inglês para o alemão, com diferentes níveis de fluência. A versão mais fluente é escolhida como base para análise no software Praat e identificação de suas características. Novamente, consoante aos achados de Schlensinger, a produção da interpretação traz uma maior quantidade de pausas, alongamento de vogais e consoantes, respiração audível e outras características consideradas disfluências pela autora. A partir desta análise, duas novas versões são produzidas no Praat – uma eliminando completamente quaisquer disfluências e outra dando maior relevância a elas. As versões mais e menos fluente são testadas em questionários de compreensão e análise subjetiva, incluindo questões sobre idade, sexo, avaliação da familiaridade dos intérpretes com o tópico e outros. Os resultados da avaliação mostram que uma maior fluência está atrelada a uma maior percepção de correção na interpretação e consequentemente associada a melhor compreensão. A conclusão de Rennert neste trabalho é a de que fluência não é só uma questão de estilo e pode impactar a opinião do ouvinte na qualidade da interpretação em termos de desempenho do intérprete e qualidade da interpretação.

No segundo, Rennert (2010) um intérprete profissional produz várias versões de interpretação de discurso do inglês para o alemão, com diferentes níveis de fluência. A versão mais fluente é escolhida como base para análise no software Praat e identificação de suas características. Novamente, consoante aos achados de Schlensinger, a produção da interpretação traz uma maior quantidade de pausas, alongamento de vogais e consoantes, respiração audível e outras características consideradas disfluências pela autora. A partir desta análise, duas novas versões são produzidas no Praat – uma eliminando completamente quaisquer disfluências e outra dando maior relevância a elas. As versões mais e menos fluente são testadas em questionários de compreensão e análise subjetiva, incluindo questões sobre idade, sexo, avaliação da familiaridade dos intérpretes com o tópico e outros. Os resultados da avaliação mostram que uma maior fluência está atrelada a uma maior percepção de correção na interpretação e consequentemente associada a melhor compreensão. A conclusão de Rennert neste trabalho é a de que fluência não é só uma questão de estilo e pode impactar a opinião do ouvinte na qualidade da interpretação em termos de desempenho do intérprete e qualidade da interpretação.

---

*“Taking note of research on quality expectations published over the years, we extended the original list of criteria to include “lively intonation”, as studied in particular by Ángela Collados Aís (1998) at the University of Granada, and “synchronicity”, which had emerged as a feature expected of simultaneous interpreting by respondents in the AICC-sponsored user expectation survey carried out by Moser (1996).”*

A pesquisas realizadas até agora indicam que uma única lista de critérios talvez seja insuficiente para a definição dos critérios mais importantes na percepção de qualidade de um trabalho de interpretação. Todos os estudos que utilizaram algum tipo de amostra de fala de intérprete a ser submetida à análise de ouvintes trouxeram questões extralinguísticas à tona. Como mencionamos no caput deste capítulo, a base da interpretação é a fala do intérprete.

A fala não transmite apenas o conteúdo estritamente linguístico das sentenças, mas também veicula outras informações, tais como a expressão de atitudes e emoções do falante. A expressividade da fala de um intérprete influencia a maneira com que esse é avaliado por seu público em termos de agradabilidade, credibilidade, assertividade ou nível de conhecimento. Passar a mensagem com elementos prosódicos adequados, incluindo tom de voz, entoação e acento, é crucial para o sucesso da interpretação. É portanto fundamental investigarmos essa fala a fundo, incluindo parâmetros como taxa de elocução, pausas, padrões entoacionais e outros para podermos entender questões como percepção de qualidade e compreensão de seu trabalho. Talvez aí possamos discutir as bases para futuras pesquisas na área.

## **2. A fala no trabalho de Interpretação Simultânea**

No trabalho de interpretação simultânea, o intérprete produz uma fala em que utiliza palavras de sua escolha, mas que ao mesmo tempo está pautada pelo discurso do outro. O intérprete precisa esperar o discurso original para compreender seu significado, lógica e registro. Só a partir de então poderá construir sua fala – que compreenderá vocabulário, estrutura gramatical, entoação, posicionamento de adjetivos e advérbios, registro de formalidade, etc. – e que poderá ser parecida ou completamente diferente da estrutura que usaria fora da cabine de interpretação.

Tradicionalmente, embora se considere desejável que o intérprete produza uma fala expressiva, as questões prosódicas assumem um papel secundário na definição e avaliação da qualidade do trabalho do intérprete. Estudos realizados entre intérpretes e ouvintes usuários da interpretação mostram que os aspectos prosódicos são considerados menos relevantes na avaliação dos itens que contribuem para a qualidade da interpretação e aparecem em uma lista de importância atrás dos critérios linguísticos, mais atrelados à consistência de sentido, coesão lógica e outros (Bühler, 1986; Kurz, 1989, Moser 1996; Chiaro e Nocella, 2004; Pöchhacker e zwischenberger, 2010). No entanto, quando os mesmos avaliadores têm como tarefa analisar a produção oral dos intérpretes, estes tendem a atribuir maior qualidade e mais alto nível de compreensibilidade àquelas versões cujas características prosódicas – como voz agradável, vivacidade e fluência – são bem avaliadas. Isto sugere que os elementos prosódicos sejam talvez igualmente importantes na avaliação de qualidade e compreensibilidade da interpretação.

Nunca podemos nos esquecer que, diferente da tradução, a interpretação é uma atividade completamente baseada na fala, tanto como fonte quanto como resultado final. A abordagem tradicional nas pesquisas sobre a qualidade da produção do intérprete tenta dissociar os elementos prosódicos de outros elementos linguísticos em uma produção que é exclusivamente verbal/vocal. Tentar traçar uma linha entre sentido e produção vocal sem levar em consideração as características prosódicas da fala é ignorar que a produção vocal carrega também sentido em si, e que esse sentido contribui para a avaliação de qualidade atribuída unicamente aos quesitos linguísticos encontrados em tais pesquisas.

*“Para interpretar, o intérprete deve primeiro entender”* (Seleskowitch 1978, p. 11). Como iremos verificar, a interpretação não é um processo mecânico, mas sim *“uma operação cognitiva altamente complexa”* (Kurz 2008: 180) em que os intérpretes primeiro devem entender a mensagem que recebem para depois verbalizá-la no idioma alvo. Essa complexidade de processamento mental é muito bem descrita no modelo teórico de Daniel Gile (1995), que explica como o trabalho

do intérprete simultâneo deve equilibrar diferentes esforços – de Audição e Análise (*Listening and Analysis*), de Memória (*Memory*), de Produção (*Production*) e de Coordenação destes esforços (*Coordenation*) (GILE, 1995)– de forma a chegar a um resultado final satisfatório, uma tarefa que Gile compara àquela de um malabarista em sua “Hipótese da Corda Bamba” (*Tightrope Hypothesis*):

*“A Hipótese da Corda Bamba preconiza que na maior parte do tempo os intérpretes trabalham próximo ao ponto de saturação, seja em termos de requisitos de capacidade de processamento mental como um todo, seja em relação a Esforços individuais, devido à alta exigência de um esforço em particular e/ou uma alocação subotimizada dos recursos a cada um dos esforços”.* (Gile, 1995)

Como podemos perceber, a fala do intérprete está comprimida entre todos os outros esforços necessários para a execução da atividade como um todo, e obviamente não passa incólume por tal compressão. Na espera pela fala do outro, o processamento da mensagem, a escolha da melhor solução para cada unidade de sentido, o intérprete produz uma fala que não pode ser chamada de espontânea nem tampouco guiada ou laboratorial – e que é absolutamente singular.

Embora tão notadamente específica, ainda há poucos estudos que investigam a fala do intérprete. Já existe evidência, no entanto, de que essa pressão de processamento mental interfere na produção verbal do intérprete, afetando sua produção prosódica (Schlesinger, 1994; Collados; 1998, 2001; Holub, 2010; Rennert, 2010).

O primeiro estudo com dados sobre as características específicas da fala do intérprete foi desenvolvido por Miriam Schlesinger (1994). Para investigar o papel da entoação na produção e percepção da interpretação simultânea, Schlesinger analisou a produção simultânea de oito intérpretes de conferência trabalhando do inglês para hebraico e do hebraico para o inglês, isolando dez extratos de aproximadamente 90 segundos, escolhidos aleatoriamente (6 extratos do inglês para hebraico) e (4 extratos do hebraico para o inglês).

Em ambos os idiomas, os resultados da análise de Schlesinger mostram um conjunto de características na fala do intérprete que trazem diferenças marcantes da fala espontânea. A primeira delas é a utilização de pausas e consequente distribuição do enunciado em unidades de informação (grupos entoacionais). Os dados parecem indicar que pausas no meio de estruturas gramaticais são um dos aspectos mais salientes da fala interpretada, ou seja, os intérpretes tendem a introduzir um número desproporcional de pausas em posições consideradas não “naturais”, quebrando as unidades de sentido dentro de um mesmo grupo entoacional. Também há um uso característico das proeminências, com sílabas “inesperadas” carregando o *pitch* accent, ou valor de *pitch* máximo, em cada grupo entoacional. Outro aspecto de destaque é a marcação do *pitch* final de fronteira, com uma tendência de os intérpretes terminarem suas frases com um *pitch* elevado, não concludente. O último aspecto investigado por Schlesinger foram as questões de duração e taxa de elocução, com a produção dos intérpretes sendo geralmente sem ritmo regular, com alterações fora do padrão tanto em duração quanto em taxa de elocução.

Os mesmos padrões prosódicos foram encontrados na produção de intérpretes brasileiros interpretando do inglês para o português (Penha, 2015), sugerindo que a descrição da fala do intérprete possa ser aplicada ao trabalho de interpretação simultânea em si, independentemente do idioma produzido. Mais estudos são necessários, no entanto, para embasar tal afirmação.

Os resultados entretanto são importantes porque nos apontam as principais direções para investigação da prosódia na fala do intérprete, os fatores que podem vir a atrapalhar a avaliação de qualidade e compreensibilidade desta fala e porquê. Para tanto, precisamos nos embasar na literatura para entender qual o papel de cada um destes aspectos prosódicos na produção da fala.

## 2.1 Pausa

A pausa é identificada por Schlesinger (1994) como a característica mais marcante da fala do intérprete. O intérprete tende a produzir um número maior de pausas e em posições consideradas não 'naturais'. O que seria, no entanto, a descrição do uso de pausas na fala espontânea e para que servem as pausas?

De acordo com Cruttenden (1986) as pausas tipicamente ocorrem em locais distintos em um enunciado:

**1)** Em um componente de fronteira importante (principalmente entre cláusulas e entre sujeito e predicado). Existe uma correlação entre o tipo de componente de fronteira e o tamanho da pausa, ou seja, quanto mais importante a fronteira, maior a pausa. Além disso as pausas tendem a ser mais longas onde os componentes de fronteira (neste caso geralmente fronteiras de sentença) envolvem um novo tópico. As pausas deste tipo são tipicamente silenciosas, mas podem ser preenchidas como ferramenta do discurso, ou seja, para evitar que outra pessoa interrompa o falante atual;

**2)** Antes de palavras de alto conteúdo léxico ou, colocando em termos da teoria da informação, em pontos de baixa probabilidade transacional. Assim, as palavras precedidas por uma pausa são geralmente difíceis de serem adivinhadas por antecipação. Este tipo de pausa geralmente ocorre antes de um componente de fronteira menor, frequentemente dentro de uma oração substantiva, oração verbal ou oração adverbial, isto é, entre um substantivo principal e um determinante;

**3)** Após a primeira palavra em um grupo entoacional. Esta é uma posição típica por outros 'erros de desempenho', por exemplo, correções de um falso início e repetições. Tanto a pausa do tipo 2 quanto a pausa do tipo 3 podem ser consideradas fenômenos de hesitação. As do tipo 2 indicam dificuldades em se encontrar a palavra adequada; as do tipo 3 parecem ter uma função de planejamento, isto é, servem essencialmente como uma operação de espera, enquanto o falante planeja o restante do enunciado.

Ao introduzir um número superior ou inferior de pausas em sua fala, em posições talvez diferentes do esperado em uma fala espontânea, o intérprete pode dificultar a compreensão de sua mensagem. Foram estes os achados de Kurz (2008) no estudo referido acima. Ao comparar a compreensão de um segmento primeiramente interpretado e depois lido pelos mesmos intérpretes, a autora identifica uma diferença significativa no nível de compreensibilidade da mensagem por parte dos ouvintes em detrimento da interpretação.

## 2.2 Entoação

Outro ponto apontado por Schlesinger (1994) como marcante na fala do intérprete é a entoação, com destaque para a marcação do *pitch* final de fronteira, com uma tendência de os intérpretes terminarem suas frases com um *pitch* elevado, não concludente.

Sabemos que o padrão entoacional de um falante é utilizado para diferenciar uma modalidade de discurso de outra (Madureira, 1999), e que diferentes padrões entoacionais sugerirão modalidades diferentes, sejam elas imperativa, declarativa ou interrogativa. A pista entoacional é importante para decodificarmos também a estrutura gramatical da fala, mas não só. Afinal, como afirma Bolinger (1958, p. 338) "A entoação é importante para quem está falando, para quem falará em seguida, para como o ato será compreendido (uma explicação, um pedido de desculpas, um

questionamento), para como o falante será avaliado (como indivíduo, como falante nativo, como membro de uma classe social) – apenas para mencionar alguns elementos que afetam nosso papel como Falantes e Ouvintes “.

Chun (2002) define quatro funções principais para a entoação: gramatical, atitudinal, de discurso e sociolinguística.

As funções gramaticais são aquelas que guardam uma maior aproximação com o nível sintático do discurso, ou seja, como já comentado anteriormente, a curva entoacional varia conforme a modalidade da oração (afirmativa, interrogativa, imperativa etc.). No entanto, como destaca Bolinger (1958), a relação entre gramática e entoação é casual e não causal. Existe uma certa escolha que vai depender de considerações pragmáticas. Isto leva Chun (2002) a subdividir a categoria de funções gramaticais em duas, onde uma concentra-se no alinhamento e associação de tons com estruturas sintáticas em particular (modalidades) e a outra na compartimentação do fluxo do discurso em várias unidades, ou seja, como função de segmentação da informação, que está associada a regularidade da fala.

As funções atitudinais são aquelas que ligam a entoação à expressão de emoções e atitudes. Brown et al (2001) sugerem que pode haver um pequeno número de padrões entoacionais que estão convencionalmente relacionados a uma série de atitudes. No inglês, por exemplo, um tom de fronteira final ascendente (na avaliação de Schlesinger (1994) não concludente no discurso do intérprete), poderá ser avaliado como característico de um falante que deseja ser agradável e encorajar o seu interlocutor a participar no diálogo.

As funções de discurso vão além do nível da oração e destinam-se a constituir coerência e continuidade com o discurso como um todo, independentemente de seu tamanho ou duração. São destinadas a marcar proeminências e fronteiras entre orações, parágrafos, tópicos a fim de dar continuação a um tópico estabelecido ou sinalizar um novo tópico a ser introduzido. Também evidenciam as pistas ao interlocutor, no caso de diálogo, de que é agora sua vez de falar.

As funções sociolinguísticas revelam diferenças na entoação de acordo com o grupo social a que o falante pertence, incluindo aí diferenças de gênero, idade, características socioeconômicas, geográficas e/ou ocupacionais. Através da fala, o intérprete revela informações sobre o grupo em que está inserido. Mais uma vez, mudanças entoacionais na fala do intérprete podem levar a uma percepção por parte do ouvinte de que o intérprete pertence a um contexto sociolinguístico diferente do seu.

### **2.3 Pitch Accent**

Sabemos que o intérprete faz um uso característico das proeminências, com sílabas ‘inesperadas’ carregando o *pitch* accent, ou seja, a sinalização de uma proeminência acentual por alteração de *pitch*.

De acordo com Laver (1994) uma das funções da entoação é sinalizar que algumas partes de uma estrutura entoacional são, de alguma forma, mais essenciais para a interpretação do sentido da mensagem do que outras partes, que vão necessariamente aparecer como menos proeminentes na fala. A gradiência da proeminência também sinaliza uma coordenação com a organização métrica de acento e ritmo de fala, imprimindo a esta uma regularidade que é importante para identificar ao ouvinte os blocos de informação que estão sendo produzidos. Orações sintaticamente idênticas com diferentes focos de *pitch* máximo vão transmitir ao ouvinte pistas de informação diferentes, e levarão, conseqüentemente, a diferentes interpretações da mensagem. Estas diferenças de interpretação de acordo com o ponto em que se percebe o *pitch* máximo (proeminência) da fala sugerem que a entoação pode ser tratada como uma forma de comportamento linguístico.

Cruttenden (1986) identifica que o *pitch* accent depende de algum tipo de obstrução do *pitch*

no ponto do acento a partir do *pitch* das sílabas circunvizinhas. Isto significa dizer que o *pitch* accent sempre aparece em comparação a um grupo entoacional como um todo e identifica um movimento de fala não isolado. Assim, o *pitch* accent depende de um movimento que vem ou se direciona às sílabas adjacentes e pode ser configurado como um patamar acima, um patamar abaixo, uma movimentação de baixo para cima ou uma movimentação de cima para baixo. Cada um desses movimentos e/ou troca de patamar leva a uma pista entoacional diferente, pois a marcação de proeminência (e, portanto, sinalização de importância da informação) irá variar de acordo com o local onde está o *pitch* accent.

Ao situar o *pitch* accent em um ponto de sua fala, o intérprete está necessariamente trazendo este ponto como informação mais importante de um dado segmento. *Pitch* accent inesperados podem levar à confusão por parte do ouvinte, pois interferem na regularidade da fala e deslocam o foco de atenção para estruturas sintáticas supostamente menos providas de sentido dentro de um grupo entoacional.

Como podemos perceber, as características prosódicas da fala não podem ser consideradas meramente uma questão pessoal de maior ou menor importância como parecem sugerir as pesquisas de qualidade na interpretação, uma vez que essas interferem no processamento da fala.

Ainda há muito a avançar nos estudos da fala no campo da interpretação, mas já há algo que podemos considerar como premissa básica dos estudos por vir – em um ofício que se comunica com base e pela fala, as questões prosódicas são absolutamente fundamentais para entendermos melhor o trabalho do intérprete.

## Referências Bibliográficas

- BOLINGER, D. **Intonation and Grammar**. *Language Learning* 8, 31-117, 1958.
- BROWN, G., CURRIE, K., KENWORTHY, J. **Questions of Intonation**. Croom-Helm, London, 1980
- Cauldwell, R. **Streaming Speech**. *Speech in Action*, UK, 2001.
- BUHLER, H. **Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters**. *Multilingua*, 5 (4): 231-235, 1986.
- CHIARO, D. & NOCELLA, G. **Interpreters' perception of linguistic and non-linguistic factors affecting quality: A survey through the World Wide Web**, 2004, disponível em <https://www.erudit.org/revue/meta/2004/v49/n2/009351ar.pdf>. Último acesso em 09/07/2015.
- CHUN, D.M. **Discourse Intonation in L2: From theory and research to practice**, Benjamins Publishing Company, 2002.
- COLLADOS AÍS, A. **La evaluación de la calidad em interpretación simultânea. La importancia de la comunicación no verbal**. Granada: Comares, 2001.
- COLLADOS AÍS, A. **La evaluación de la calidad em interpretación simultânea. La importancia de la comunicación no verbal**. Granada: Comares, 1998.
- CRUTTENDEN, A. **Intonation**, Cambridge Textbooks in Linguistics, 1986.
- GILE, D. **Testing the Effort Models' tightrope hypothesis in simultaneous interpreting** – A contribution, *Hermes, Journal of Linguistics* no. 23, 1999 – disponível em [http://download2.hermes.asb.dk/archive/FreeH/H23\\_09.pdf](http://download2.hermes.asb.dk/archive/FreeH/H23_09.pdf). Último acesso em 09/07/2015.
- GILE, D. **Basic Concepts and Models for Interpreter and Translator Training**, Benjamins Translation Library, 1995.
- GILE, D. **L'évaluation de la qualité du travail par les délégués: une étude de cas**. *The Interpreters' Newsletter*, 3, 66-71, 1990.
- HARTLEY, A; MASON, Ian; PENG, G. PEREZ, Isabelle Anne; (2003) **Peer- and Self-Assessment in Conference Interpreting Training. / Research output: Book/Report** Commissioned report.
- HOLUB, E. "Does Intonation Matter? The impact of monotony on listener comprehension", in: *The*

Interpreters' Newsletter, 15, pp. 117-126, 2010.

KAHANE, E. (2000) **Thoughts on the quality of interpretation**, in <http://aiic.net/page/197/thoughts-on-the-quality-of-interpretation/lang/1>.

KURZ, I. The Impact of non-native English on students' interpreting performance, em *Efforts and Models in Interpreting and Translation Research*, 2008. Benjamins Translation Library.

KURZ, I. Quality and role: The Professionals' view, 2008. Disponível em <http://aiic.net/page/3044/quality-and-role-the-professionals-view/lang/1>. Último acesso em 09/07/2015.

KURZ, I. Conference Interpreting—User Expectations, Coming of Age. Proceedings of the 30th Annual Conference of the American Translators Association (D. L. Hammond, ed.), Medford (NJ), Learned Information, 1989.

LAVIER, J. **Principles of Phonetics**, Cambridge Textbooks in Linguistics, 1994.

MADUREIRA, S. **Estudos de prosódia**/Ester M. Scarpa, org – Campinas, SP: Editora da Unicamp, 1999.

MOSER P., 'Expectations of Users of Conference Interpretation'. *Interpreting*, 1(2), 145-178, 1996.

MOSER, P. Survey: **Expectations of Users of Conference Interpretation** [Final Report commissioned by AICC]. Wien: SRZ Stadt – und Regionalforschung GmbH, 1995.

PENHA L. **A Importância da Prosódia na Avaliação de Qualidade e na Compreensão e Compreensibilidade da Fala Interpretada Simultaneamente**. 2015, dissertação de mestrado, PUCSP.

PÖCHHACKER, F. The issue of quality- By a word in your year. *Interpreting quality: global professional standards?* In: Ren, W. (ed.), *Interpreting in the Age of Globalization: Proceedings of the 8th National Conference and International Forum on Interpreting Beijing: Foreign Language Teaching and Research Press*, 2012. <https://lourdesderioja.com/2013/04/24/the-issue-of-quality/>. Acesso em 09/07/2015.

PÖCHHACKER, F. ZWISCHENBERGER, **Survey on quality and role: conference interpreters' expectations and self-perceptions**, 2010, disponível em <http://aiic.net/page/3405/survey-on-quality-and-role-conference-interpreters-expectations-and-self-perceptions/lang/1>. Último acesso em 09/07/2015.

RENNERT, S. The impact of fluency on the subjective assessment of interpreting quality", in: **The Interpreters' Newsletter**, 15, pp. 101-115, 2010

SCHLESINGER, M. Intonation in the production of and perception of simultaneous interpretation, In Lambert and Moser-Mercer (Eds.) *Bridging the Gap. Empirical Research in Simultaneous Interpretation*. Benjamins, 1994.

SELESKOVITCH, D. Who should assess an interpreter's performance in *Multilingua - Journal of Cross-Cultural and Interlanguage Communication*, Volume 5, Issue 4, Pages 236–236, ISSN (Online) 1613-3684, ISSN (Print) 0167-8507, DOI: 10.1515/mult.1986.5.4.236, 2009.

SELESKOVITCH, D., *English, Interpreting For International Conferences*, Washington, DC: Pen and Booth, 1978.

ZWISCHENBERGER, C.; PÖCHHACKER, F. **Survey on quality and role: Conference interpreters' expectations and self-perceptions**. *Communicate!* Spring, 2010. <http://www.aiic.net/ViewPage.cfm/article2510.htm> (accessed 2014).

*On the Varying Reception of  
Speakers Expressivity Across Gender and  
Cultures, and Inference in their  
Personalities*

ALBERT RILLIARD  
DONNA ERICKSON  
JOÃO DE MORAES  
TAKA AKI SHOCHI

## 1. Introduction

Speakers' behaviour is interpreted by interlocutors in many ways. One can extract the intended speech act, the speaker's gender, age, regional or social origin, her/his personality or current mood and emotional state, etc. – cues to such kinds of judgements found in their face and other physical attributes, as well as in their voice and speech characteristics (e.g. Scherer 1972, 1978; Lippa, 1978; Ekman et al., 1980; Borkenau & Liebler, 1992).

Voice cues, i.e. cues that don't take part in the linguistic message (cf. the definition of "prosody" in Swerts & Kramer, 2005), are known to carry information about the speaker's emotional state and expressivity (Scherer, 2009; Goudbeek & Scherer, 2010). Emotion-related cues are said to be mostly linked to physiological changes, and thus largely coherent across languages and cultures (Scherer & Wallbott, 1994; Scherer et al., 2004). Conventional use of such cues, driven by symbolic and cultural choices (Léon, 1993; Ohala, 1994; Madureira & Camargo, 2010), are used together with speech production to help monitor the interpersonal interactions, or to produce more effective sentences in paralleling several sources of information (cf. the distinction made by Moraes et al., 2010, between social and propositional attitudes, that respectively express for example politeness or doubt). Conventionalized cues in voice are said to vary across languages (Delattre, 1963) – and the inventory of such conventional expressions indeed vary (e.g. Chang, 1958; Uldall, 1960; Martins-Baltar, 1977; Fujisaki & Hirose, 1993; Shochi et al., 2009, Gu et al., 2011). Some studies investigate what it takes to teach such conventionalized voice cues, and their contextual use, to foreign language learners (Shochi et al., 2010, 2014; Lu, 2015).

These works have underlined the similarities that some conventionalized expressions share across cultures (e.g. expressions of surprise), while others are linked to different conceptual backgrounds and/or are performed with varying strategies (Shochi et al., 2009). But comparisons of the vocal cues used in these expressions are hindered by the problematic translation of their concepts and contextual uses (Wierzbicka, 1992). To better understand what really differs in the voices (and not in the concepts) linked to expressions performed by speakers from different linguistic origin, a multilingual corpus was recorded (Rilliard et al., 2013), placing speakers in communication contexts, giving them a set of communication goals and an interlocutor with predefined hierarchical relationships. Dialogues have been set up, ending in the production by each speaker of the same two sentences produced with sixteen expressive variants.

Quality evaluations of the speakers' performances have been submitted to L1 listeners (Rilliard et al., 2013, 2014). They show variations in the mean performance scores attributed to groups with different linguistic backgrounds. If L1 speakers do perform better on average, some expressions, typically those that may be linked to the L2 speakers' cultures, received higher performance scores for L2 speakers than for L1 ones. Interpretation of such results is difficult. Recognition tests show that such expressions that received higher performance scores for L2 speakers are better recognized by L1 listeners when produced by L1 speakers. Thus performance judgements are partly biased by the listener's representation of speaker's cultural backgrounds.

To what extent did the expressive abilities of individual speakers, their expressive behaviours, their accents and/or language specificities influence the conceptualization speakers make from these individuals? This paper presents some results on the perception people have of the speakers' personality from such expressive performances, and then tries to link these perceptions to acoustic cues differentiating speakers and their expressive behaviours, as well as to how their individual performances were received. It aims at sorting out what cues can be used by listeners – how listeners may build an understanding of their interlocutor's communication aims, depending on their preconceptions of that person.

**Table 1** – ID of speakers in each language (US: United States of America, FR: France, JP: Japan) and gender group.

L1	Gender	
	<i>Male</i>	<i>Female</i>
<i>US</i>	<i>S12, S16, S7</i>	<i>S14, S18, S24, S3, S6</i>
<i>FR</i>	<i>S29, S32, S34</i>	<i>S37, S40</i>
<i>JP</i>	<i>S11, S15, S19</i>	<i>S17, S20, S25</i>

## 2. Data & Analyses

### 2.1. Corpus

The corpus is based on the English version of the recordings, and specifically, on one of the two sentences (“a banana”). This sentence was uttered in over sixteen situations by speakers – either L1 English speakers from the USA, or L2 speakers of English, from Japan or from France (cf. table 1). Audio-visual recordings were made in soundproof rooms at Waseda (Japan) and Bordeaux (France) universities. The sixteen situations correspond to the following labels (with their abbreviations): admiration (ADMI), arrogance (ARRO), authority (AUTH), contempt (CONT), doubt (DOUB), irony (IRON), irritation (IRRI), neutral declarative sentence (DECL), neutral question (QUES), obviousness (OBVI), politeness (POLI), seduction (SEDU), sincerity (SINC), surprise (SURP), uncertainty (UNCE), and “walking on eggs” (WOEG). The last label, “walking on eggs”, corresponds to a certain extent to the Japanese *kyoshuku* situation, adapted to an American English context. *Kyoshuku*, defined by Sadanobu (2004, p. 34), as “a mixture of suffering, ashamedness and embarrassment, which comes from the speaker’s consciousness of the fact his/her utterance of request imposes a burden to the hearer” has no lexical equivalent in English, but situations exist where an English speaker may feel something akin to *kyoshuku*. Thus, for each speaker, 16 “banana” utterances are available (i.e. a total of 304 stimuli for 19 speakers), containing a considerable variation in terms of expressive voice changes.

### 2.2. Measures of acoustic variation

An overview of this variation is given in this section. Acoustic parameters have been extracted to measure a set of possible changes in voice quality.

Laver (1980) proposed a description of voice quality dimensions, separating supralaryngeal from phonatory settings, which served as basis for a fully detailed description of voice qualities through the Vocal Profile Analysis Scheme (VPAS, adapted to Brazilian Portuguese by Camargo & Madureira, 2008). Supralaryngeal changes in voice quality include effect of lip spreading/rounding, tongue back or front position, more or less open jaw, nasalisation. Such changes have typical effects on formants’ values. Changes in phonatory settings lead to audible effects that may be described as register, noise (additive or structural), tension (lax to pressed), and effort (cf. d’Alessandro, 2006). These perceptive dimensions are linked by d’Alessandro (2006) to a set of acoustic parameters. From these possible measures a set of parameters have been selected for their availability and direct measure from the speech waveform (avoiding inversion procedures, that are rarely tested in expressive conditions):

The fundamental frequency (F0) relates to *pitch* and registers (d’Alessandro, 2006); measured on each vowel using Praat’s standard algorithm, and expressed in semitones (ST) relative to

1 Hertz. Intensity (and more specifically A-weighted intensity) is correlated to the perceived effort (Traunmüller & Eriksson, 2000; Liénard & Barras, 2013): it is measured on vowels using a Praat script, and expressed in decibels (dB). Note that the recordings were made using a calibrated microphone to correct for change in gain control – measures are thus comparable between speakers. Measures of Harmonic-to-Noise Ratio (HNR) in the speech signal give the amount of noise (noise originating in aperiodicities and in additive turbulences, d’Alessandro et al., 1998). HNR are measured on vowels, and expressed in dB.

The difference between the energies of the first two harmonics (H1-H2, Hanson 1997) has been related to vocal fold tension. H1-H2 is measured using the COVAREP toolbox (Degottex et al., 2014) on vowels, and expressed in dB.

In addition to these five source parameters, measures of the first three formants were made using the Praat standard algorithm, with the parameters recommended for female or male voices. The median and range values (the range is measured as the difference between the 10<sup>th</sup> and 90<sup>th</sup> percentiles of the observed distribution) were recorded for each speaker.

A Principal Component Analysis (PCA) was run on the (normalized) acoustic measures. The three first dimensions explain more than 70% of the variance, and were kept on the basis of an elbow criterion. The distribution of acoustic variables on these dimensions is summarized in table 2. Abdi & William (2010) propose to keep variables that have a contribution over the mean contribution of one dimension. Under this criterion, the first dimension is based on the speaker’s median F0 – correlated to high HNR and wide variation of F2, and negatively to HNR range. The second dimension is constructed on range of variation in intensity, F0, F1 and H1-H2. The third dimension is based on F1 and F2 median values, correlated to intensity and H1-H2 medians.

**Table 2** – Coordinates ( $F_i$ ), contributions ( $ctr_i$ ) and squared cosines ( $\cos^2$  – multiplied by 100 and rounded for convenience) for the first three dimensions of the PCA run on acoustic measures taken on the 19 speakers.

Measure	$F_1$	$F_2$	$F_3$	$ctr_1$	$ctr_2$	$ctr_3$	$\cos^2_1$	$\cos^2_2$	$\cos^2_3$
F0 median	0.82	-0.07	0.30	<b>19</b>	0	5	72	1	9
F0 range	-0.09	0.79	0.15	0	<b>24</b>	1	1	66	2
INT median	-0.51	-0.42	0.50	7	7	<b>14</b>	28	19	26
INT range	-0.19	0.81	0.26	1	<b>25</b>	4	4	68	7
HNR median	0.75	-0.13	-0.33	<b>15</b>	1	6	59	2	12
HNR range	-0.82	0.10	0.16	<b>18</b>	0	1	70	1	3
F1 median	0.53	-0.33	0.64	8	4	<b>23</b>	30	11	43
F1 range	0.58	0.60	0.06	9	<b>14</b>	0	35	38	0
F2 median	0.55	0.13	0.49	8	1	<b>13</b>	32	2	25
F2 range	0.68	-0.16	-0.26	<b>13</b>	1	4	48	3	7
H1-H2 median	0.05	0.35	0.58	0	5	<b>19</b>	0	13	36
H1-H2 range	0.14	0.66	-0.44	1	<b>17</b>	11	2	47	20

**Table 3** – Coordinates ( $F_i$ ), contributions ( $ctr_i$ ) and squared cosines ( $\cos^2$  – multiplied by 100 and rounded for convenience) for the first four dimensions of the PCA run on quality measures judged for the 19 speakers.

	$F_1$	$F_2$	$F_3$	$F_4$	$ctr_1$	$ctr_2$	$ctr_3$	$ctr_4$	$\cos^2_1$	$\cos^2_2$	$\cos^2_3$	$\cos^2_4$
ADMI	0.42	0.46	-0.41	0.50	3	9	11	<b>18</b>	18	21	16	25
ARRO	0.88	-0.05	-0.01	-0.16	<b>12</b>	0	0	2	77	0	0	3
AUTH	0.92	-0.17	-0.13	-0.02	<b>13</b>	1	1	0	84	3	2	0
CONT	0.76	0.05	0.22	0.10	9	0	3	1	58	0	5	1
DECL	0.60	0.47	-0.14	-0.04	6	<b>10</b>	1	0	35	22	2	0
DOUB	0.31	0.04	0.40	-0.50	2	0	11	<b>19</b>	10	0	16	25
IRON	0.45	0.59	-0.08	-0.44	3	<b>15</b>	0	14	20	34	1	19
IRRI	0.77	-0.34	-0.10	-0.07	<b>9</b>	5	1	0	60	11	1	0
OBVI	0.79	-0.24	0.21	-0.29	<b>10</b>	3	3	6	63	6	5	8
POLI	0.43	0.30	0.36	0.47	3	4	9	<b>17</b>	18	9	13	22
QUES	0.25	0.59	-0.52	-0.27	1	<b>15</b>	<b>18</b>	5	6	35	27	7
SEDU	0.60	-0.63	-0.06	0.22	6	<b>17</b>	0	4	36	39	0	5
SINC	0.27	0.49	0.21	0.18	1	<b>10</b>	3	2	7	24	4	3
SURP	0.74	-0.22	-0.27	0.33	<b>9</b>	2	5	8	55	5	7	11
UNCE	0.42	0.37	0.67	0.20	3	6	<b>31</b>	3	18	14	44	4
WOEG	0.86	-0.19	-0.10	-0.13	<b>12</b>	2	1	1	74	4	1	2

### 2.3. Measure of the performances' quality

The quality of the speakers' performances was rated by L1 USA English speakers. Listeners were first presented the sixteen contexts, with examples of prototypical situations. Then, for each speaker, the name of the context was given, and then the speaker's performance was audio-visually displayed. Judges rated on a 1 to 9 scale. Each of the three linguistic groups of speakers was rated in separate experiments (details in Rilliard et al., 2013, 2014). Quality scores given by each listener were thus centred and scaled for normalizing for individual usage of the scale. A PCA was run on the scores obtained by each speaker for each expression. The first four dimensions were kept, explaining more than 70% of the variance. The distribution of the speaker's performance on these dimensions is summarized in table 3. The type of expression regrouped on each dimension, according to their contribution, is as follow: The first dimension is positively based on arrogance, authority, contempt, irritation, obviousness, surprise and "walking-on-egg". The second dimension is based on declaration, irony, question, sincerity (positively), and seduction (negatively). The third dimension is based on question (positively) and uncertainty (negatively). The fourth dimension is based on admiration, politeness (positively), and doubt (negatively). Note that these groupings of expression are based on the speakers' performances in each of them, not necessarily on their semantic (or otherwise) similarity.

The first dimension is mostly based on the overall performance score received by each speaker (S3 and S34 being judged as the best performers; S29 and S14 as the worst), and this overall judgement also corresponds to the scores given to highly activated expressions, expressing an imposition onto the addressee (i.e. most of the attitudes well represented on the first dimension). Other dimensions show more specific performances. The second dimension is linked to the expressions of irony and seduction; speakers receiving high performance for only one of those two. Dimensions 3 and 4 separate the best performers in one or two of these expressions: uncertainty, politeness, admiration, question, or doubt.

## 2.4. Measure of speaker's personality

The personality of the speakers was rated by people watching the expressive performances. The test consists of a presentation of all 16 “banana” expressive performances recorded from each of the 19 speakers. For each speaker, subjects had to rate a list of characteristics they thought the speaker might have. Listeners then answered the same questionnaire for themselves. 36 subjects, all L1 speakers of USA English took the test – 15 were presented with the audio-only performances, 21 with the audio-visual performances. The questionnaire subjects had to fill in for speakers (and themselves) is presented in Nettle (2007: the “Newcastle Personality Questionnaire”). It contains twelve assertions that have to be rated for their adequacy of the speaker, on a 5-point scale ranging from “very uncharacteristic” to “very characteristic”. The assertions are given in table 4. A combination of these answers to the assertions (cf. table 3, and Nettle, 2007, for details), allows rating the five personality traits proposed in Costa & McCrae (1992): extraversion, neuroticism, conscientiousness, agreeableness, openness. The scores are then transformed into four categories (low, low-medium, high-medium, high) according to the spread of population on these scores, as described in Nettle (2007: pp. 272ff).

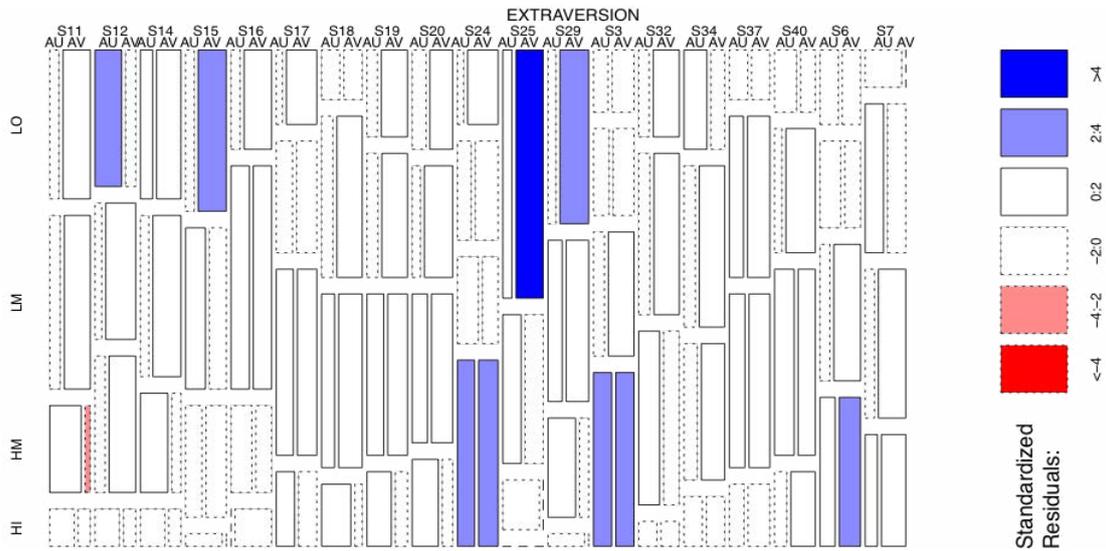
**Table 4** – the twelve questions composing the questionnaire subjects had to fill in for rating the speakers' personality, with the trait each contributes to rate (E: extraversion, N: neuroticism, C: conscientiousness, A: agreeableness, O: openness).

Assertion	Trait
Starting a conversation with a stranger	E
Making sure others are comfortable and happy	A
Creating an artwork, piece of writing, or piece of music	O
Preparing for things well in advance	C
Feeling blue or depressed	N
Planning parties or social events	E
Insulting people	A
Thinking about philosophical or spiritual questions	O
Letting things get into a mess	C
Feeling stressed or worried	N
Using difficult words	O
Sympathizing with others' feelings	A

For each personality trait, contingency tables are built on the basis of the subjects' answers, counting how many times each of the four levels of the scale was selected, for each speaker in each modality. These tables are analyzed with a log-linear model allowing the tiles representations of figures 1 to 5. These representations show in blue the speakers who are significantly perceived more often than average in a given scale level for a personality trait and in red the speakers who are significantly perceived less often than average in a given level for a personality trait.

## 2.5. Extraversion

Figure 1 shows that three speakers are judged with high levels of extraversion: S24, S3, S6. The first two are judged so in both modalities, S3 only in the AV modality. Four speakers (S12, S15, S25, S29) are judged with low level of extraversion – all but S12 in the AV modality (S12 in the audio only modality). Extraverted speakers are L1 English speakers; the AV introverted speakers are L2 speakers – thus a potential language/cultural effect.



**Figure 1** – Tile representation of the contingency matrix giving, for each speaker, in each modality (in columns), a representation of the number of judgements in the four levels of Extraversion (in lines: Low, Low-medium, High-medium, High). The levels significantly departing from an average distribution are shown in blue (for excess), or in red (for fewer ratings).

## 2.6. Neuroticism

Results for neuroticism are spread over the scale (cf. figure 2), thus few tendencies emerge.

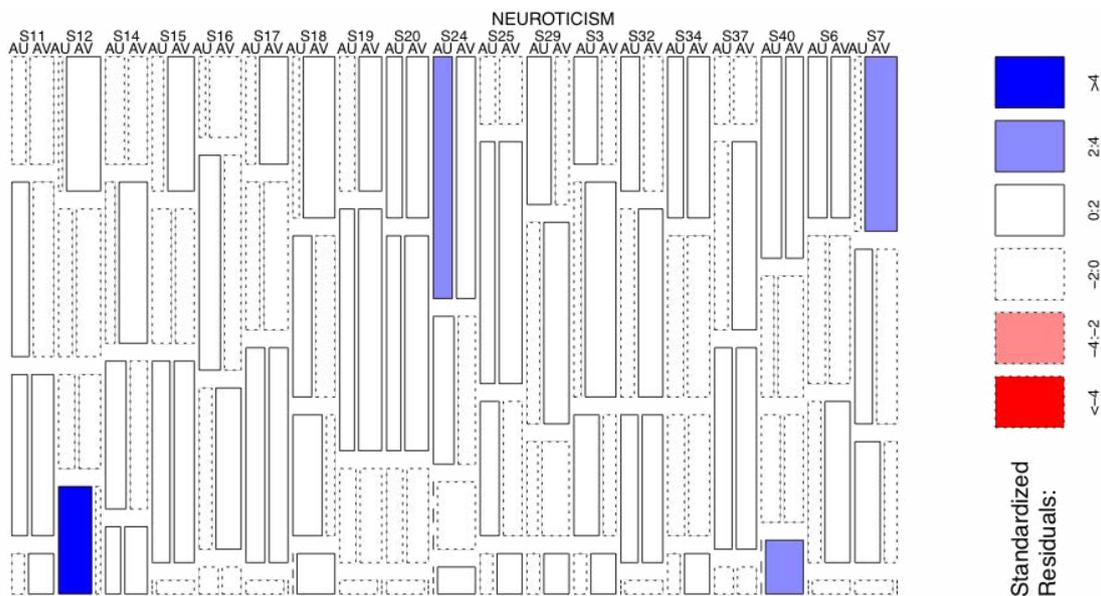
Four speakers received significantly high or low scores, two at each side (high: S12, S40; low: S24, S7), and half on each modality. Two of these speakers (S12, S24) had opposite ratings in extraversion.

## 2.7. Conscientiousness

Ratings of conscientiousness varied across speakers (cf. figure 3). Three speakers (L1 females) received high scores (S18, S3, S6), the first two in both modalities, S6 in AV only. Four speakers received low scores (S11, S29, S32, S7), all in the AV modality. All of them are male, and most L2 speakers

## 2.8. Agreeableness

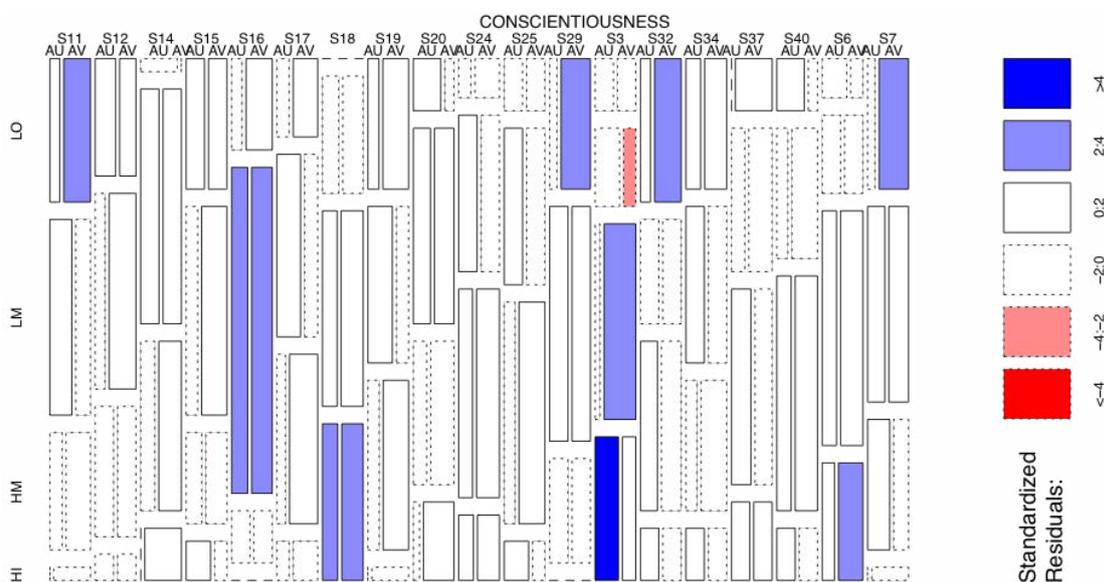
For agreeableness (cf. figure 4), four speakers (L1 females) received high or high-medium agreeableness scores (S18, S19, S24, S6); and one speaker (L2 male) was judged as having a disagreeable personality (S34). All the positive judgements were made in the AV modality, the negative one being done in the audio modality.



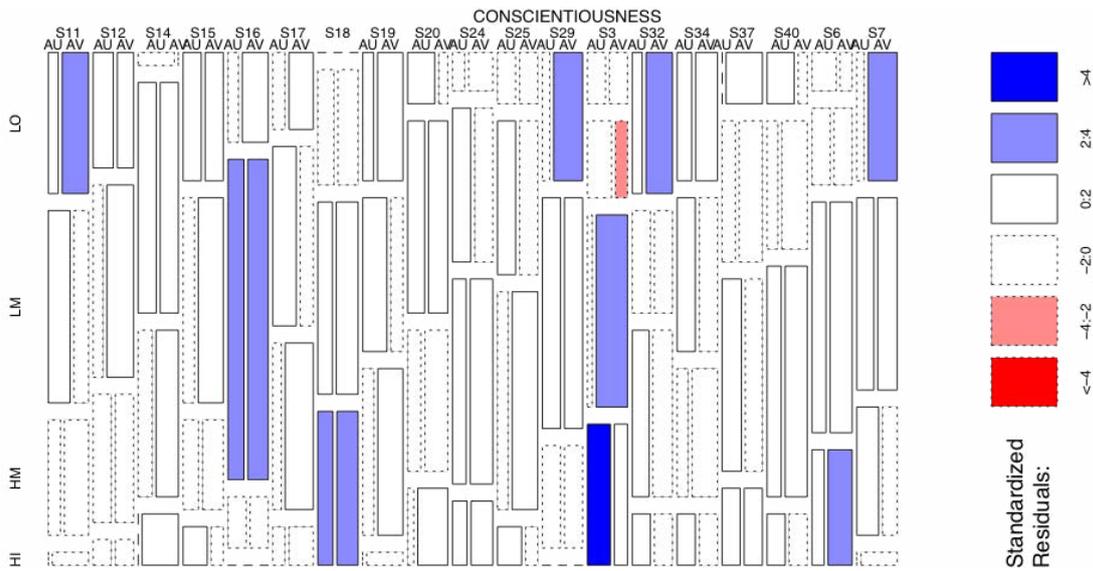
**Figure 2** – Tile representation of the contingency matrix giving, for each speaker in each modality (in columns), a representation of the number of judgements in the four levels of Neuroticism (in lines: Low, Low-medium, High-medium, High). The levels significantly departing from an average distribution are shown in blue (for excess), or in red (for fewer ratings).

## 2.9. Openness

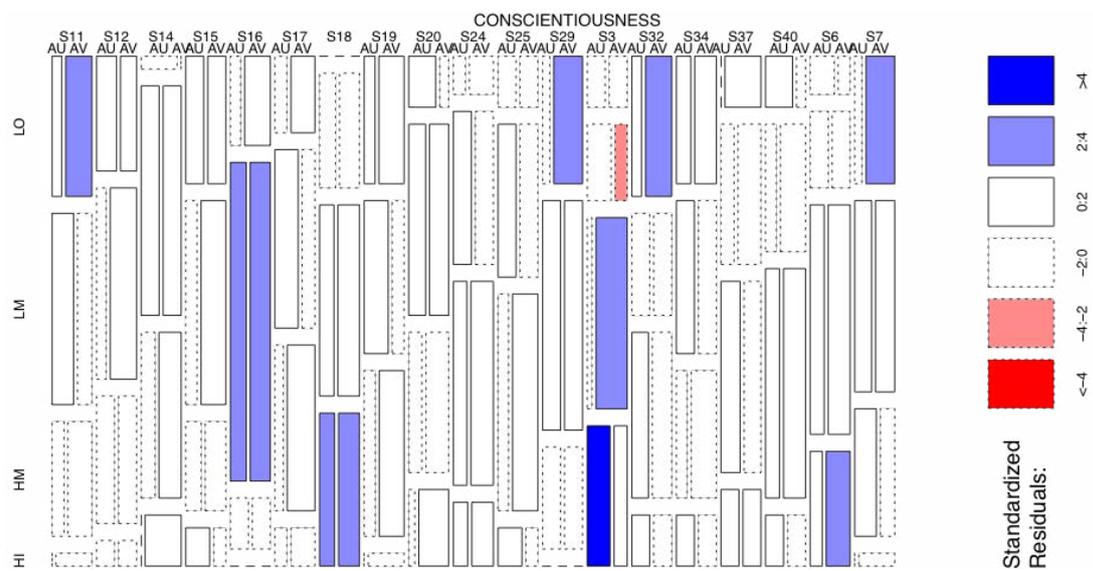
Judgements of openness (cf. figure 5) are spread across the scale for most speakers. One speaker (S32) is rated high in AV; another (S14) is rated high-medium in AV, but low in audio-only.



**Figure 3** – Tile representation of the contingency matrix giving, for each speaker in each modality (in columns), a representation of the number of judgements in the four levels of Conscientiousness (in lines: Low, Low-medium, High-medium, High). The levels significantly departing from an average distribution are shown in blue (for excess), or in red (for fewer ratings).



**Figure 4** – Tile representation of the contingency matrix giving, for each speaker in each modality (in columns), a representation of the number of judgements in the four levels of Agreeableness (in lines: Low, Low-medium, High-medium, High). The levels significantly departing from an average distribution are shown in blue (for excess), or in red (for fewer ratings).



**Figure 5** – Tile representation of the contingency matrix giving, for each speaker in each modality (in columns), a representation of the number of judgements in the four levels of Openness (in lines: Low, Low-medium, High-medium, High). The levels significantly departing from an average distribution are shown in blue (for excess), or in red (for fewer ratings).

In order to further analyze the changes in personality trait ratings made on the basis of observations of audio or audio-visual performances, PCA were run on each trait scales in each modality (complete results cannot be presented here). Most PCA have their first dimension correlated with the high-low dimension. For neuroticism and openness, where the spread is large, many answers are grouped toward the lower bound, and the second dimension is used to differentiate the remaining high scores.

### 3. Comparison between Analyses

The previous section gives three different views on the performances of a set of 19 speakers performing expressive behaviours on the same “banana” sentence: an acoustic description of the prosodic variation in each speaker, judgements on their performances, and judgements of the personality traits perceived through these behaviours. In order to compare these three sets of views and check how they may be related, and to what extent the factors of gender and language background have an influence on them, a grouped analysis of this data is given here. To compare such mixed data, we opted for a Multiple Factor Analysis (MFA) that allows comparing the various multidimensional analyses presented here and linked their various dimensions and variables (cf. Bécue-Bertaut & Pagès, 2006, for details and mathematical details on the method; R’s FactoMineR library was used for the computation, Husson et al., 2010). The MFA took the following set of tables as input to describe the 19 speakers:

- The mean and range of acoustic measures (12 measures, scaled and centred);
- The performance scores obtained on each attitude (16 scores, scaled and centred);
- The number of high, high-medium, low-medium and low ratings in the 5 personality traits, for audio-only and for audio-visual modalities (5 scores for 5 traits in 2 modalities);
- The supplementary categorical variables of speaker’s gender and speaker’s L1 (note: supplementary variables do not participate in the analyses, but are used to interpret results).

Table 5 presents the variables that are significantly associated with the first three dimensions of the MFA (that explains more than 40% of the variance). No significant associations are observed for the other dimensions.

The MFA first dimension is mostly associated with the perception of high extraversion, conscientiousness, and agreeableness (and negatively to low measures of extraversion and conscientiousness). This percept may be related to the important range of F0, F1 and intensity, and a high F0 register – which relates to the perception of good performance of irritation. The first dimension is related to the speaker’s gender, with females having higher F0 register, and being rated as more extraverted than males.

The second dimension is inversely related to judgements of agreeableness and openness; it is also linked to voices with a large range of noise, large intensity variation, small F2 variations, and low median HNR. These perceptions are linked to good performances in authority, surprise, seduction, contempt and irritation. A relation with speakers’ gender is also observed (weaker than the one observed for the first dimension, but significant), with females having higher ratings than males on the side linked to high agreeableness, and less noisy voices.

The third dimension is correlated to the perception of neuroticism. It is correlated to the perception of high performances in sincerity, declaration, authority and “walking-on-eggs” – most expressions having strong social values and constraints. No specific relations with the measured acoustic variables have been found.

For the two supplementary variables linked to speakers (gender and language background), only gender shows significant relation with the three sets of measures: there is thus no observed link between the speakers’ linguistic origin and personality judgement by interlocutors. On the contrary, females in our set of speakers are rated with higher levels on the extraversion, conscientiousness, agreeableness and openness traits (no effect on neuroticism is observed). For agreeableness and openness, the video display may play a role, as most correlates are only observed within the audio-visual group (but for high agreeableness). The relatively limited number of speakers of both genders did not allow us to draw many conclusions on gender, given the complex nature of these

observations.

**Table 5** – measures of association (r – correlations for acoustic, performance and personality variables, regression coefficient for supplementary variable) between the variables and the first three dimensions of the MFA. The personality measures are referred to according to the first letter of the personality trait (E: extraversion, N: neuroticism, C: conscientiousness, A: agreeableness, O: openness), the modality (au: audio, av: audio-visual), and the ratings (high, high-medium, low-medium, low). Negative correlations are set in red font.

1st dimension		2nd dimension		3rd dimension	
Name	r	Name	r	Name	r
<i>Acoustic</i>					
F1 range	0.76	HNR range	0.72		
F0 range	0.64	INT range	0.59		
INT range	0.60				
F0 median	0.60				
		HNR median	-0.70		
		F2 range	-0.70		
<i>Performance</i>					
IRRI	0.59	AUTH	0.58	SINC	0.51
		SURP	0.58	DECL	0.48
		SEDU	0.56	AUTH	0.47
		CONT	0.46	WOEG	0.47
		IRRI	0.46		
<i>Personality</i>					
<b>E, au, high</b>	<b>0.85</b>	A, av, low	0.67	<b>N, au, hi-med</b>	0.69
C, av, high	0.82	E, av, hi-med	0.56	<b>O, au, lo-med</b>	0.66
<b>C, au, high</b>	<b>0.71</b>				
C, av, hi-med	0.71				
<b>A, au, high</b>	<b>0.68</b>				
E, av high	0.68				
E, av, low	-0.56	A, av, lo-med	-0.57	<b>N, au, low</b>	-0.60
<b>C, au, lo-med</b>	<b>-0.59</b>	E, av, low	-0.61		
C, av, lo-med	-0.64	<b>A, au, hi-med</b>	-0.62		
C, av, low	-0.65	O, av, hi-med	-0.73		
<i>Supplementary variable</i>					
Gender	0.53	Gender	0.30		

## 4. Discussion & Conclusions

The acoustic analysis shows a first dimension primarily related to *pitch* register – which mostly correlates to speaker’s gender, i.e. static cues linked to the physical characteristic of individuals; this dimension also correlated to the noise measure (high voices having less noise in our data). The second dimension is linked to the range of changes in the measured cues (intensity, F0, F1, H1-H2), and thus is related to each speaker’s vocal expressivity, to her/his dynamic capabilities. The third dimension is related to median values in intensity, F1 and H1-H2 – this is related to amount of vocal

effort exerted by the speaker during these expressive performances. A-weighted intensity is known to covary with effort (Traunmüller & Eriksson, 2000; Liénard & Barras, 2013), as H1-H2; F1 is known to increase with wider jaw movement (Erickson et al., 2012) that are likely to happen to express increased linguistic prominence. These acoustic dimensions are related to the judgments listeners give for the speakers.

The main link observed between acoustic cues and perception relates expressive cues to extraversion and conscientiousness, and to the perception of higher irritation performances. The irritation expression was set up in the recordings as the overt expression of one's annoyance towards a colleague's behaviour – thus requiring some sort of extraversion. It therefore makes sense both are linked in the results. Moreover, irritation was consistently produced with a strong vocal effort, which is coherent with Gussenhoven's (2004) effort code prediction, and to measures of wide intensity, F0 and F1 ranges. It also recalls Scherer's "loud voice of extraversion" (Scherer, 1978).

The second MFA dimension links the perception of low agreeableness and openness to high performances in authority, surprise, seduction, contempt and irritation. All these expressions expressed to some extent an imposition on the interlocutor (imposing the speaker's authority, lack of understanding, sexual desire, rejection, or norm), which may link the expressivity to a more disagreeable and close-minded personality. Such expressions and personality judgements are related to increased noise level in the voice.

The third dimension links perception of high neuroticism to efficient behaviour for the sincerity, declaration, authority and "walking-on-eggs" expressions. As already mentioned, these expressions are strongly related to social norms of behaviour, that define how one can behave, with who, and when. Hill et al. (1986) discuss the notion of discernment for the expression of politeness in Japanese, where social expectations are described as higher than in the American cultures. The sincerity and "walking-on-eggs" expressions were set up here on the basis of Japanese social situations (cf. Rilliard et al., 2013 for details), and described socially normed relationships – so does authority, expressed by a person in charge of an official responsibility that allows her/him to behave that way. Such highly normed behaviours are here linked to neuroticism (which may be described as a tendency to feel negative emotions). The complex feeling expressed in "walking-on-eggs", that includes notions of suffering and embarrassment, may fit with such a personality trait.

The results of these two perceptual experiments thus bring coherent results, and acoustic cues have been detailed for the first two dimensions of behavioural evaluations. One has to stress that there are no visual cues available until now to bring more information on this channel. Visual cues are certainly of a primary importance to judge both personality and behavioural expressivity (cf. de Moraes & Rilliard 2014 for an example).

There seems to be a gender effect on the judgements made by listeners. The acoustic cues did not point toward more expressivity in the female speakers – if expressivity can be limited to the mean values of a few acoustic measures. Visual information is partly responsible for such a potential effect, typically for agreeableness.

On the contrary, no effect of the linguistic background emerges from this data. Meanwhile, some personality traits are linked to expressions being socially constructed – linked to the role of individual in the society, and their expected behaviour in such a role helps understand such changes in the perception of personality. On that point, one may relate the constrained behaviours of Japanese speakers to their cultural training in behaving in well-defined situations, to the notion of character in Sadanobu's view (2015) – in between personality and style: the taboo nature of playing a character in social situation may explain why it is interpreted in such an experimental context as "personality". Creation of various such characters to fit socially constrained situations is coherent to Hill et al. (1986)'s notion of high/low discernment in societies – Japan being a high discernment society. Listeners may thus have both prototypical views on some linguistic groups that could also fit

the personality judgements of some of these speakers. This co-occurrence may explain the higher performance scores received by some expressions, that did not fit with the lowest recognition scores: foreign speakers have behaviours that fit with an expression, but their expressive strategies did not necessarily fit the USA English prototypes for such a situation.

## Acknowledgements

The authors are deeply indebted to the many speakers and listeners from Waseda, Bordeaux, New Mexico, and Black Hills State Universities, who participated in these recordings and perception experiments. Also, we acknowledge the French National Research Agency (ANR) PADE and SEDUCTION CPU (ANR-10-IDEX-03-02) grants, and the Japan Society for the Promotion of Science (JSPS), Grants-in-Aid for Scientific Research (A) #25240026 to the second author.

## References

- ABDI, H.; WILLIAMS, L. J. Principal component analysis. **Wiley Interdisciplinary Reviews: Computational Statistics**, 2, v. 04, p. 433–459, 2010.
- D'ALESSANDRO, C.; YEGNANARAYANA, B. D. Effectiveness of a periodic and aperiodic decomposition method for analysis of voice sources. **IEEE Transactions on Speech and Audio Processing**, 6, v. 01, p.12–23, 1998.
- D'ALESSANDRO, C. Voice source parameters and prosodic analysis. In SUDHOFF, S. et. al. (Eds). **Methods in empirical prosody Res. Berlin**, Walter de Gruyter, p. 63–87, 2006.
- BÉCUE-BERTAUT, M. PAGÈS, J. Multiple factor analysis and clustering of a mixture of quantitative, categorical and frequency data. **Computational Statistics & Data Analysis**, 52, v. 06, p.3255–3268, 2008.
- BORKENAU, P., LIEBLER, A. The cross-modal consistency of personality: Inferring strangers' traits from visual or acoustic information. **J of Res in Personality**, 26, v.02, p. 183-204, 1992.
- CAMARGO, Z.; MADUREIRA, S. Voice quality analysis from a phonetic perspective: Voice Profile Analysis Scheme (VPAS) Profile for Brazilian Portuguese. In: **Speech Prosody 2008**, Campinas. **Proceedings...** Campinas: p.57–60, 2008.
- CHANG, N-C. T. Tones and intonation in the Chengtu dialect. **Phonetica**, Szechuan, China, 2, v. 1-2, p.59-85, 1958.
- COSTA, P. T.; MCCRAE, R. R. Four ways five factors are basic. **Personality and individual differences**. 13, v. 06, p. 653–665, 1992.
- DEGOTTEX, G. et. al. A collaborative voice analysis repository for speech technologies. In: **IEEE Int. Conf on Acoustics, Speech and Signal Processing (ICASSP)**, 2014. Florence. **Proceedings...** Florence: 960–964, 2014.
- DELATTRE, P. Comparing the prosodic features of English, German, Spanish and French. In: **IRAL-Int. Review of Applied Ling. in Lang. Teaching**. 1, v. 01, p. 193–210, 1963.
- EKMAN, P.; et. al. Relative importance of face, body, and speech in judgments of personality and affect. **J of Personality and Social Psychology**, 38, v. 02, p. 270–277, 1980.
- ERICKSON, D.; et. al. Metrical structure and production of English rhythm. **Phonetica**. 69, p. 180–190, 2012.
- FUJISAKI, H.; HIROSE, K. Analysis and perception of intonation expressing paraLing. information in spoken Japanese. In: **Proceedings of the ESCA Workshop on Prosody**, Sweden: Lund, p. 254-257, 1993.
- GOUDBEEK, M.; SCHERER, K.R. Beyond arousal: Valence and potency/control cues in the vocal expression of emotion. **The J of the Acoustical Society of America**. 128, v. 03, p. 1322–1336,

2010.

- GU, W.; ZHANG, T.; FUJISAKI, H. Prosodic Analysis and Perception of Mandarin Utterances Conveying Attitudes. In: **Interspeech**, 2011. Firenze **Proceedings...** Firenze: p.1069-1072, 2011.
- GUSSENHOVEN, C. **The phonology of tone and intonation**. Cambridge: Cambridge U P, 2004.
- HANSON, H.M. Glottal characteristics of female speakers: Acoustic correlates. **The J of the Acoustical Society of America**. 101, v. 01, p. 466–481, 1997.
- HILL, B.; et. al. Universals of Linguistic politeness: Quantitative evidence from Japanese and Am English. **J of pragmatics**, 1, v. 03, p.347-371, 1986.
- HUSSON, F.; LE, S.; PAGES, J. **Exploratory multivariate analysis by example using R**. London: Chapman & Hall/CRC, 2010.
- LAVER, J. **The phonetic description of voice quality**. Cambridge: Cambridge U P, 1980.
- LÉON, P. **Précis de phonostylistique: parole et expressivité**. Paris: Nathan, 1993.
- LIÉNARD, J.S.; BARRAS, C. Fine-grain voice strength estimation from vowel spectral cues. In **INTERSPEECH**, 2013. Lyon. **Proceedings...** Lyon: p.128-132, 2013.
- LIPPA, R. The naive perception of masculinity-femininity on the basis of expressive cues. **J of Res in Personality**, 12, v. 01, p. 1-14, 1978.
- LU, Y. **Etude contrastive de la prosodie audio-visuelle des affects sociaux en chinois mandarin vs. français: vers une application pour l'apprentissage de la langue étrangère ou seconde**. PhD. U Grenoble Alpes, 2015.
- MADUREIRA, S.; CAMARGO, Z. A. Exploring sound symbolism in the investigation of speech expressivity. In: 3rd ISCA Workshop ExLing, 2010. Athens. **Proceedings...** Athens: p.105–108, 2010.
- MARTINS-BALTAR, M. **De l'énoncé à l'énonciation: une approche des fonctions intonatives**. Paris: Didier, 1977.
- MORAES, J.; et. al., Multimodal perception and production of attitudinal meaning in Brazilian Portuguese. In: **Fifth Int. Conf on Speech Prosody**, 2010. Chicago. **Proceedings...** Chicago: ISCA Archive, p 10-14, 2010.
- \_\_\_\_\_.; RILLIARD, A. Illocution, attitudes and prosody: A multimodal analysis. In Raso, T. and Ribeiro De Mello, H. (Eds.). **Spoken Corpora and Ling. Studies**. John Benjamins Pub Co, p. 233-270, 2014.
- NETTLE, D. **Personality: What makes you the way you are**. Oxford: Oxford U P, 2007.
- OHALA, J.J. The frequency codes underlies the sound symbolic use of voice *pitch*. In HINTON, L, NICHOLS, J.; OHALA, J.J. (Eds.), **Sound symbolism**. Cambridge: Cambridge U P, p.325–347, 1994.
- RILLIARD, A.; et. al. Social face to face Com - Am English attitudinal prosody. In: **Interspeech**, 2013, Lyon. **Proceedings...** France: 2013.
- \_\_\_\_\_. US English attitudinal prosody performances in L1 and L2 speakers. In: 7th Int. Conf on Speech Prosody, 2014. Dublin. **Proceedings...**Dublin: p. 895–899.2014.
- SADANOBU, T. A natural history of Japanese pressed voice. **J of the Phonetic Society of Japan**. 8, v.01, p. 29–44, 2004.
- \_\_\_\_\_. **“Characters” in Japanese Com and Lang.: an overview**. Acta Ling.a Asiatica, 5, v.02, p. 9–28, 2015.
- SCHERER, K. R. Judging personality from voice: A cross-cultural approach to an old issue in interpersonal perception. **J of Personality**, 40, v.02, p. 191–210, 1972.
- \_\_\_\_\_. Personality inference from voice quality: the loud voice of extroversion. **European J of Social Psychology**, 8, v. 04, p. 467–487, 1978.
- \_\_\_\_\_.; WALLBOTT, H. G. Evidence for universality and cultural variation of differential emotion response patterning. **J of personality and social psychology**, 66, v.02, p. 310–328, 1994.
- \_\_\_\_\_.; et. al. Emotions in everyday life: probability of occurrence, risk factors, appraisal

- and reaction patterns. **Social Sci Information**, 43, v.04, p. 499-570, 2004.
- SCHERER, K.R. Emotions are emergent processes: they require a dynamic computational architecture. **Philosophical Transactions of the Royal Society B: Biological Scis**, 364, v.1535, p. 3459–3474, 2009.
- SHOCHI, T. et al. Learning effect of French prosodic social affects for Japanese learners of French Lang., In: Int. Conf on Speech Prosody, 2010. Chicago. **Proceedings...** Chicago: 2010.
- SHOCHI, T.; et. al. Intercultural perception of English, French and Japanese social affective prosody. In: S. Hancil (Ed.). **The role of prosody in affective speech. Ling. Insights 97**, Peter Lang AG, Bern, p. 31–59, 2009.
- \_\_\_\_\_. Apprentissage des expressions prosodiques et gestuelles de politesse par des francophones. In: **M. Andro-Ueda & J.-M. Butel (Eds.), Japon pluriel 9 — Histoires d’amour : quelques modalités de relation à l’autre au Japon**, Paris: Philippe Picquier, p. 415-424, 2014.
- SWERTS, M.; KRAHMER, E. Audiovisual prosody and feeling of knowing. **J of Memory and Lang.**, 53(1), p. 81–94, 2005.
- TRAUNMÜLLER, H.; ERIKSSON, A. Acoustic effects of variation in vocal effort by men, women, and children. **The J of the Acoustical Society of America**, 107, v.06, p. 3438–3451, 2000.
- ULDALL, E. Attitudinal meanings conveyed by intonation contours. **Lang. and Speech**. 3, v.04, p. 223–234, 1960.
- WIERZBICKA, A. Defining emotion concepts. **Cognitive Sci.** 16, p.539–58, 1992.

*Scripts for the Acoustic Analysis of speech  
data*

PLÍNIO A. BARBOSA  
ZULEICA A. CAMARGO  
SANDRA MADUREIRA

# 1. Introduction

Due to volume of data, prosody research can enormously benefit from the automatisisation of procedures for describing prosodic form closely related to prosodic functions such as prominence, boundary and discursive relations marking. Automatisisation is advantageous because the same procedures can be applied to analyze the entire corpus and that is useful for preparing data for the statistical analysis as well.

This paper presents five scripts running on Praat (Boersma and Weenink, 2013) which generate prosodic descriptors for speech prosody research. Assuming from early research that duration is a crucial parameter for signalling stress, prominence and boundary in languages such as English, Portuguese, French, German and Swedish, the scripts are able to detect prosody-related acoustic salience, as well as to generate a 12-parameter vector of duration, intensity and F0 descriptors for prosodic analysis. The usefulness of the scripts is discussed regarding the results obtained from their application in previous research.

## 2. Semi-automatic detection of acoustic salience via duration

The SGdetector script for Praat was implemented in 2004 and improved in 2009 and 2010 for allowing the semi-automatic detection of local peaks of normalised syllable-sized durations. Languages that use duration to signal both stress and prosodic boundary such as Brazilian Portuguese (henceforth BP) and Swedish (Barbosa et al., 2013), English (Wightman et al., 1992; Fry, 1958), German (Dogil, 1995; Sluijter, 1995) and French (Barbosa, 1994) are well suited to take advantage of such a tool.

Thoroughly tested since 2004 with BP (see, for instance, \cite{barbosa08,barbosa10}), the script was also used to do analyses in the aforementioned languages and has potential to be applied to additional languages, at least those genetically related to the ones already tested. It was also in studies of speech expressivity (Madureira, 2008; Madureira e Camargo, 2010).

The input files for running the script are a TextGrid file containing a phone-sized or syllable-sized segmentation and a broad phonetic transcription of the corresponding Sound file, as well as a TableOfReal file containing a table listing the means and standard-deviations in milliseconds of the phone durations of the language under study. This latter file is delivered as part of the script and is available for BP, European Portuguese, British English, German, Swedish and French.

Manual and semi-automatic segmentations and transcriptions of audio files were repeatedly tested for BP along the years (see Barbosa, 2010), confirming the usefulness and correction of this method for detecting potential perceived prominence and boundary based on syllable-sized duration, as shown by experimental research presented below.

Syllable-sized segmentation is meant as a first step to capture prosodic-relevant duration variation along the utterances (Barbosa, 2006) and is understood here as an interval between two consecutive vowel onsets. This unit constitutes a phonetic syllable called a VV unit. Besides the crucial importance of vowel onset detection for speech signal processing (Dogil and Braun, 1988), a clear advantage of a segmentation based on vowel onsets is its potential for automatic detection (Barbosa, 2010) even under moderately noisy conditions. Automatic detection of vowel onsets can be carried out by using Praat script BeatExtractor, also developed in 2004 (Barbosa, 2006; 2010), which is explained in more details in the next section.

In the SGdetector script, detection of peaks of prosodic-relevant VV duration is carried out by serially applying a technique of normalisation followed by a smoothing technique.

For normalising VV duration, the script uses the z-score transformation given in equation 1, where  $dur$  is the VV duration in ms and the pair  $(\mu_i, var_i)$  is the reference mean and variance in ms

of the phones within the corresponding VV unit. The reference descriptors for the aforementioned languages are included in the file `TableOfReal` delivered with the script.

$$z = \frac{dur - \sum_i \mu_i}{\sqrt{\sum_i var_i}} \quad (1)$$

For smoothing, the script applies a 5-point moving average filtering technique given by equation 2 to the sequence of z-scores ( $z_i$ ).

$$z_{smoothed}^i = \frac{5 \cdot z^i + 3 \cdot z^{i-1} + 3 \cdot z^{i+1} + 1 \cdot z^{i-2} + 1 \cdot z^{i+2}}{13} \quad (2)$$

The two-step procedure just described aims at minimising the effects of intrinsic duration and number of segments in the VV unit, as well as minimising the effect of the implementation of stress irrelevant for the prosodic functions of prominence and boundary marking. Local peaks of smoothed z-scores are detected by tracking the position of the VV unit for which the discrete first derivative of the corresponding smoothed z-score changes from a positive to a negative value.

At the output, the script generates two text files, a new TextGrid object and an optional trace of the syllable-sized smoothed/normalised duration along the time-course of the Sound file under analysis. The first text file is a 5-column table displaying the following values for each VV unit: (1) its given transcription recovered from the TextGrid itself, e.g., “eNs”, “at” (even for the case where the segmentation is made phonewise), (2) its raw duration in milliseconds, (3) its duration z-score, (4) its smoothed z-score and (5) a binary value indicating if its position corresponding to a local peak of smoothed z-score (value 1) or not (value 0).

The second text file is a 2-column table containing (1) the raw duration in milliseconds of the acoustically-defined stress groups, which are delimited by two consecutive peaks of smoothed z-scores and (2) the corresponding number of VV units in these stress groups. These corresponding values were often used to do linear regressions in order to evaluate the degree of stress-timing of a speech passage, for instance in (Barbosa et al., 2009).

The TextGrid generated by the script contains an interval tier delimiting the detected stress group boundaries, synchronised with the input TextGrid, which allows, when selected with the corresponding Sound file, to listen to the chunks that end with a duration-related salience. The optional feature, implemented when the option `DrawLines` is chosen in the input parameters windows, plots a trace of the smoothed z-scores synchronised with the VV unit sequence: each value of smoothed z-score is plotted in the y-axis in the position of each vowel onset along the plotted original TextGrid. The advantage of this choice for integrating intonation and rhythm descriptions is discussed below.

The correspondence between smoothed z-scores peaks and perceived salience, which refers to both prominence and prosodic boundary, is striking. In \cite{barbosa10}, we demonstrated an accuracy varying from 69 to 82 % between perceived and produced salience, as shown in Table 1 for the semi-automatic algorithm described here.

**Table 1** – Precision, recall, and accuracy in percentage (%) for semi-automatic detected salience against perceived salience for the Lobato corpus read by a female (F) and a male (M) speaker at slow (s), normal (n) and fast (f) rates.

Sp/rate	precision	recall	accuracy
F/n	90	74	82
F/f	73	57	69
M/s	88	67	73
M/f	61	70	70

Perceived salience was determined by asking two groups of ten listeners to evaluate two readings of a passage by two BP speakers (a male and a female) at two distinct rates. The listeners in both groups were lay undergraduate students in Linguistics. They were free to listen to the four readings as many times as they wanted.

In the first group, each listener was given a handout with the orthographic transcription of the recording and was instructed to circle all the words s/he considered highlighted by the speaker.

The second group was instructed to circle the words that preceded a boundary. In each group, the percentage of listeners that circled each word in the text for each reading was initially used to define three levels of salience, according to a one-tailed z-test of proportion. Since the smallest proportion significantly distinct from zero is about 28 % for  $\alpha = 0.05$  and  $N=10$ , words circled by less than 30 % of the listeners were considered non-salient. For  $\alpha = 0.01$ , the threshold for rejecting the null hypothesis is about 49 %. Thus, words circled by 50 % of the listeners or more were considered strongly salient. Words salient by between 30 and 50 % of the listeners were considered weakly salient. For the purpose of computing the performance measures in the table, weakly and strongly salient words were both considered as “salient” *tout court*.

The relatively high correspondence between perceived and produced salience allowed us to evaluate the degree of stress-timing in two different speaking styles for two varieties of Portuguese (Barbosa et al., 2009) This work revealed that the speech rhythm of Portuguese speakers differs remarkably from the rhythm of Brazilian speakers when both groups narrate but not when both groups of speakers read. This was possible to demonstrate through the linear correlation between interval durations delimited by smoothed z-score peaks and number of VV units in the same interval. These two series of values were recovered from one of the tables generated by the SGdetector script.

### 3. Making the script completely automatic

For helping detecting produced salience in large corpora, the SGdetector script was modified into a SalienceDetector script for which phone labelling and manual vowel onset marking was made unnecessary. For this we associated the BeatExtractor script with the SGdetector script described above.

The BeatExtractor script implements Cummins’ Beat Extractor (Cummins and Port, 1998; Barbosa, 2006) with some modifications. It generates a TextGrid containing intervals between consecutive vowel onsets. It runs according to five steps: (1) the speech signal is filtered by a default second-order Butterworth (or Hanning) filter; (2) the filtered signal is then rectified; (3) the rectified signal is low-pass filtered using 20 Hz (see step 4a) or 40 Hz (see step 4b) as cut-off frequencies. This signal is normalised by dividing all points by the maximum value. This normalised, band-specific amplitude envelope is called the beat wave, a technique also applied by Cummins and Port (1998) as well as Tilsen and Johnson (2008); (4) a vowel onset is set either (a) at a point where the

amplitude of the beat wave local rising is higher than a certain threshold, or (b) at a local maximum of the normalised first derivative of the beat wave, provided this maximum is higher than a certain threshold; (5) a Praat TextGrid is generated that contains all vowel onsets as interval boundaries. More details in Barbosa (2010).

After obtaining the vowel onset positions, the SaliencyDetector script proceeds by computing duration z-scores by using fixed values for the reference mean (Refmean = 193 ms) and standard-deviation (RefSD = 47 ms) duration according to equation 3, where  $m$  estimates the actual number of VV units between each interval generated by the BeatExtractor algorithm, which may miss vowel onsets (up to 20 % from all vowels effectively present in the Sound file).

$$z = \frac{\frac{\sqrt{m}}{m} \cdot dur - \sqrt{m} \cdot Refmean}{RefSD} \quad (3)$$

Smoothed z-scores are determined in the same way as before, by using the 5-point moving average filter. The output files are the same of the semi-automatic SGdetector script. The performance of this algorithm is a little lesser than that of the semi-automatic algorithm shown in Table 1, as it can be seen in Table 2, for which accuracy varies from 53 to 80 %.

**Table 2** - Precision, recall, and accuracy in percentage (%) for automatic detected salience against perceived salience for the Lobato corpus read by a female (F) and a male (M) speaker at slow (s), normal (n) and fast (f) rates.

Sp/rate	precision	recall	accuracy
F/n	80	69	74
F/f	61	53	61
M/s	75	57	62
M/f	78	67	79

There is the possibility of enhancing the performance of the script by manually changing the input parameters or by using a gradient-descent technique to find the input parameters that achieve the better performance in a limited set of utterances of a particular language, since this script is language-independent. Nevertheless, its usefulness depends on the relevance of syllable-sized duration to signal both boundary and prominence in the language under analysis. As an additional feature, the SaliencyDetector script indicates the occurrence of silent pauses in the corresponding TextGrid interval.

#### 4. Describing the relations between F0 trace and syllable-sized duration trace

The normalised syllable-sized duration trace obtained with the *DrawLines* option of the SG-Detector script was conceived in such a way as to give the value of normalised duration along the vowel onsets of the utterance. This feature allows the possibility of plotting the F0 contour of the utterance against the evolution of normalised duration and examining the VV units for which *pitch* accents and boundary tones coincide with normalised duration peaks. This was presented in Barbosa (2008).

A priori probabilities are the proportion of *pitch*-accents,  $p(F0)$ , and normalised duration peaks,  $p(dur)$ , considering the total number of phonological words. Conditional probabilities, on the other hand, consider the co-occurrence between a duration peak with a *pitch* accent over the total number of duration peaks,  $p(F0/dur)$ , or the total number of *pitch* accents,  $p(dur/F0)$ .

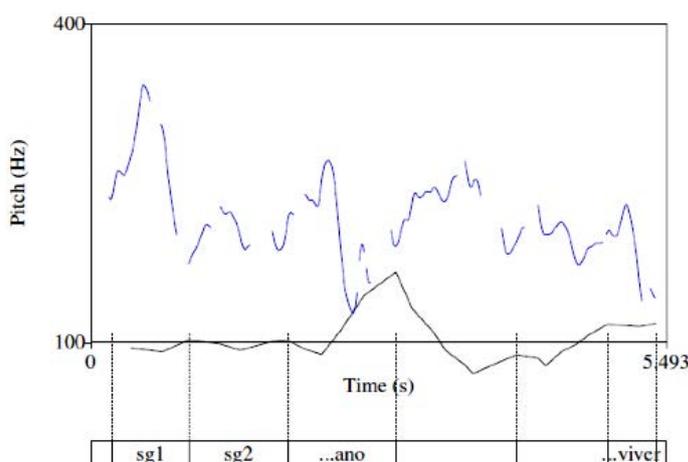
A significant difference, computed from a test of proportions with  $\alpha = 0.02$ , between a priori and conditional probabilities signals a dependence between *pitch* accent and duration peak.

The table shows that there is dependence between duration peak and *pitch* accent for the female speaker in spontaneous speech, as well as for speaker AC in read speech: for the latter, a *pitch* accent implies 76 % of chance of a duration peak. For the female speaker both are inter-related.

**Table 3** – A priori probability of *pitch* accent  $p(F0)$  and duration peak  $p(dur)$  in percentage (%) of number of phonological words. Speaker and speaking style are indicated. Stars signal significant differences between a priori and conditional probabilities  $\alpha= 0.02$ ).

sp (sp.sty)	$p(F0)$	$p(F0/dur)$	$p(dur)$	$p(dur/F0)$
F (spont.)	63*	79*	49*	63*
M (spont.)	73	80	48	56
AC (read)	54	66	56	76*
AP (read)	70	83	65	74

This inter-relation is confirmed when the analysis is restricted to major prosodic boundaries in read speech (utterance boundaries, clause and subject-predicate boundaries): 98 % (speaker AP), and 100 % (AC) of the time, both *pitch* accent and duration peak occur in the same lexical item, usually in the stressed vowel for *pitch* accents, and in the stressed or pre-pausal VV unit for duration peaks. Fig. 1 illustrates how both traces can be visualised. This was possible with the use of the *DrawLines* option of the SGDetector script. In this figure, the labels *sg1* and *sg2* signal the first two stress groups. The first rising F0 contour during *sg1* signals a prominence not accompanied by a duration peak. The two low boundary tones inside the stress groups ending in *ano* and *viver* occur during a VV unit with a duration peak.



**Figure 1** – F0 contour superposed on the VV normalised duration contour of read utterance “Manuel tinha entrado para o mosteiro há quase um ano, mas ainda não se adaptara àquela maneira de viver.”

## 5. Semi-automatic extraction of global prosodic parameters

The ProsodyExtractor script delivers 12 prosodic descriptors for whole utterances or chunks of the same utterance in order to allow research on the link between prosody production and perception.

This script has as input parameters the names of the Sound and corresponding TextGrid files. The TextGrid file must be composed of two interval tiers, one with the labelling and segmentation of the VV units (VV tier), and the other with the delimitation of the chunks of the audio file for analysis (Chunk tier). The number of intervals in the Chunk tier can vary from one to any number of units corresponding to any kind of phrasing needed for the intended analysis (e.g., syntactic phrases, prosodic constituents like stress groups, content-based chunks, among others). F0 contour is also computed, thus, it is necessary, as for the *Pitch* buttons in Praat, to inform minimum and maximum *pitch* range.

For each chunk in the corresponding interval tier, the algorithm generates (a) 6 duration-related measures computed from the metadata obtained by using the algorithm of the previously described SGdetector script, (b) 5 descriptors obtained from the *Pitch* object computed by the script and (c) a measure of spectral emphasis as defined by (Traunmüller and Eriksson, 1995).

The six duration-related measures computed in each chunk are: speech rate in VV units per second (sr), maximum of smoothed VV duration z-score, mean of smoothed VV duration z-score, standard-deviation of smoothed VV duration z-score, rate of smoothed VV duration z-score local peaks (pr), and rate of non-salient VV units. The five F0 descriptors are F0 median, range, maximum, minimum, as well as F0 peak rate. For computing the latter measure a smoothing function (with cut-off frequency of 1.5 Hz) followed by a quadratic interpolation function are applied before the F0 peak rate computation.

The 12 measures generated per chunk can be used both to study the evolution of these prosodic parameters throughout a speech signal, as well as to correlate prosody production and perception. As regards the latter, we used the difference of these values between paired utterances as predictors of the degree of discrepancy between perceived manner of speaking (Barbosa and Silva, 2012). The experimental design consisted in instructing 10 listeners to evaluate two subsets of 44 audio pairs combining 3 different speakers of BP and two speaking styles, storytelling and reading. The instruction was “Evaluate each pair of excerpts as to how they differ according to the manner of speaking given a scale from 1 (same manner of speaking) to 5 (very different manner of speaking)”. After testing more than 50 models of multiple linear regression, results showed that the best model was the one which explained 71 % of the variance of the listeners’ responses ( $lr$ ), as given in equation 4 with p-value of at least 0.009 for all coefficients ( $F_{3,11} = 12.4$ ,  $p < 0.0008$ ).

$$lr = -1.5 + 10.4pr + 2.65sr - 10.75pr * sr \quad (4)$$

This reveals that the significant production parameters that explain the listeners’ performance are speech rate in VV units/s and normalised duration peak rate, which can be associated with the syllable succession and salient syllable succession.

A similar version of this script, ExpressionEvaluator, extracts five classes of acoustic parameters and four statistical descriptors, producing 12 acoustic parameters. The five classes of acoustic parameters comprise the fundamental frequency (F0) with the extraction of the following descriptors: median, inter-quartile semi-amplitude, skewness, and 0.995 quantile; the fundamental frequency first derivative (dF0) with the descriptors mean, standard-deviation and skewness; global intensity skewness, spectral tilt (SpTt), and Long-Term Average Spectrum (LTAS) standard-devia-

tion. Spectral tilt is a correlate of vocal effort and was set to the difference of intensity in dB between the bands 0-125 Hz and 1250-4000 Hz.

The F0 first derivative was used to detect abrupt changes in the melodic contour. The values of the F0 and dF0 descriptors were z-scored by using F0 mean and standard-deviation reference values for adult males (136 Hz, 58 Hz) and females (231 Hz, 120 Hz). Spectral tilt was normalised by dividing its value by the complete-band intensity median, whereas LTAS standard-deviation was normalised by dividing its value by 10.

The ExpressionEvaluator script was applied to analyze voice quality settings in clinical and expressive speech samples (Barbosa, 2009). In this line of investigation, the acoustic measures derived by the script were correlated to perceived voice quality settings identified by the application of the Vocal Profile Analysis Scheme-VPAS by Laver, Wirs, Mackenzie, Hiller (1981). The application of the ExpressionEvaluator script made it possible to distinguish between neutral (modal voice) and non-neutral phonatory settings (Camargo, Madureira, 2010; Rusilo, Camargo, Madureira, 2011; Camargo, Queiroz, Rusilo, Madureira, 2013). The identification of the modal voice setting was related to little variation in F0 and to first F0 derivative (df0) values across speech samples of the same speaker. Furthermore, f0 and df0 measures were found to be useful to detect intermittent occurrences of voice quality settings, such as phonatory settings with irregular vocal fold vibration: breathy voice, harsh voice and creaky voice.

Camargo, Madureira, Pessoa, Rusilo (2012) described, from the perceptual and acoustic (via ExpressionEvaluator script) points of view, differences in voice quality settings related to gender. The corpus was composed by speech samples recorded by 38 subjects, aging from 20 to 58 years-old. The data showed the discriminatory capability of f0 and spectral tilt measures in differentiating speaker's gender.

Long term measures obtained by means of the ExpressionEvaluator script in Camargo, Queiroz, Rusilo, Madureira (2013) were able to predict laryngeal and tension voice qualities. The results showed the discriminant power of long-term acoustic measures to predict neutral (79 %) and non-neutral (75 %) laryngeal and tension settings. The relevant measures to detect neutral setting (modal voice) were f0 e df0, with high discriminant scores (82 %). On the other hand, for the non-neutral settings in laryngeal and tension domains, the combination of spectral tilt and LTAS acoustic measures was influential, with discriminant power of 66 %, especially for laryngeal hyperfunction and raised larynx settings.

In clinical environment with adults patients, some investigations using the ExpressionEvaluator Script have been carried out (Camargo, Medina, Rusilo, Gorinchteyn, 2013; Camargo, Lima-Silva, Rusilo, Madureira, 2013).

Camargo, Medina, Rusilo, Gorinchteyn (2013) found some correspondences between perceived voice quality and acoustic measures in AIDS patients speech samples. The relevant measures were F0 (especially for female), spectral tilt and LTAS measures (specially for male).

Camargo, Lima-Silva, Rusilo, Madureira (2013) investigated the acoustic and perceptual correlates of voice quality and vocal dynamics in speech samples recorded by teachers with voice complaints and laryngeal disorders. In agglomerative hierarchical cluster analysis-AHCA, tendencies towards clustering spectral tilt (mean and SD) and LTAS in one group and F0 (median and quantile 0.995) measures in another one were found. Canonic correlation analysis showed the relevance of spectral tilt and F0 measures in acoustic-perceptual correlations: spectral tilt (skewness) and the following perceptual findings: breathiness (40%), harsh voice (32 %), high habitual *pitch* (28,8%) and supralaryngeal hyperfunction (26 %); spectral tilt (mean) and the following perceptual findings: high habitual loudness (30 %), laryngeal hyperfunction (30 %) and harsh voice (30 %); spectral tilt(SD) and the following perceptual findings:pharyngeal constriction (29 %) and high habitual loudness (27 %); f0 (99,5% quantile) and harsh voice (26 %); f0 (median) and inadequate breath

support (25 %); f0 (interquartile semi-amplitude) and inadequate breath support (21 %). Linear regression analysis reinforced the relevance of acoustic measures-f0 (median: 59 %) and spectral tilt (skewness: 43 % and mean: 39 %) and perceptual data (harsh voice: 48 %, fast speech rate: 47 %, retracted tongue body: 41 %, breathiness: 40 % and pharyngeal constriction: 39 %) in relation to laryngeal disorders. These findings reinforce the multidimensional aspect of voice and the importance of acoustic and perceptual profiles to provide a detailed description of teacher vocal behavior.

In clinical studies of children with voice and speech disorders, some explorations were also carried out (Pessoa, Novaes, Kuhn, Camargo, 2011; Pessoa, Novaes, Madureira, Camargo, 2012). In a case report, Pessoa, Novaes, Kuhn, Camargo (2011) applied the ExpressionEvaluator script to speech samples, recorded by a unilateral cochlear implant (UCI) user at the ages of 5 and 6. Data related to f0 measures were influential in discriminating the different speech samples collected in the period of a year.

Pessoa, Novaes, Madureira, Camargo (2012) explored the speech characteristics of a hearing impaired child, a three-year and ten-month-old bilaterally implanted user. These findings indicated the influence of laryngeal hyperfunction setting and aperiodicity on *pitch* extension and variability.

The investigation of voice quality based on the measures extracted by means of the ExpressionEvaluator script made it possible to correlate acoustic to perceptual dimensions and to identify instances of co-occurrence of voice quality settings. (Rusilo, Camargo, Madureira, 2011). Those aspects are relevant not only to consider compensatory strategies to voice and speech disturbances in the clinical context (patients with AIDs, voice disorders, including teachers with laryngeal disturbances and children with cochlear implants) but also but also to foster research on speech expressivity (Madureira, 2011; Madureira, Camargo, 2010).

Madureira (2011) discusses specific uses of sound symbolism concerning segmental and prosodic properties in the reading of a poem by a professional actor. The duration values of the V-V units in the five repetitions of the stanza chosen for the purpose of analysis were compared by means of ANOVA. No differences were found among repetitions 1, 2 3 and 4. Repetition 5 was found to differ from the others  $p = 0.000$ . The fifth repetition was also found to differ from the others in relation to f0 (median, 0.995 quantile, skewness and its first derivative mean, standard deviation and skewness) and in relation to the long term average spectrum (LTAS). These differences in LTAS correlate with differences in voice quality identified by means of the VPAS.

## **Summary and availability of the tools**

The tools presented here were used to conduct research on speech rhythm analysis and modelling either in a single language or crosslinguistically, on the relation between intonation and rhythm both *stricto sensu*, as well as on the link between speech rhythm production and perception. They were tested in French, German, Brazilian and European Portuguese, Swedish and English, the latter two less systematically.

All scripts, cited in this article, are available freely from the author. They are licensed under the terms of the GNU General Public License as published by the Free Software Foundation; version 2 of the License and have been tested in French, German, Brazilian and European Portuguese, Swedish and English, the latter two less systematically.

## **Acknowledgments**

The authors thank research grants from CNPq: first author, CNPq grant 301387/2011-7), second author CNPq grant 303227/2012-5 and third author, CNPq grant 306818/2010-8.

## References

- BARBOSA, P.A. **Caractérisation et génération automatique de la structuration rythmique du français**. PhD thesis, France: ICP/Institut National Polytechnique de Grenoble, 1994.
- \_\_\_\_\_. At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration. In: **Proceedings of the 1st ETRW on Speech Production Modeling**. 1996, AuTrans. AuTrans: 85-88, 1996.
- \_\_\_\_\_. **Incursões em torno do ritmo da fala**, Campinas: RG/Fapesp, 2006.
- \_\_\_\_\_. Prominence- and boundary-related acoustic correlations in Brazilian Portuguese read and spontaneous speech, **Proc. Speech Prosody**. Campinas: 257-260, 2008.
- \_\_\_\_\_. Detecting changes in speech expressiveness in participants of a radio program. In: **Proceedings of the Interspeech - Speech and Intelligence**, 2009, Brighton. Londres: Causal Productions, p. 2155-2158, 2009.
- \_\_\_\_\_; VIANA, M. C.; TRANCOSO, I. Cross-variety Rhythm Typology in Portuguese. In: **Proceedings of the Interspeech - Speech and Intelligence**, 2009, Brighton. Londres: Causal Productions, p. 2155-2158, 2009.
- \_\_\_\_\_; DA SILVA, W. A New Methodology for Comparing Speech Rhythm Structure between Utterances: Beyond Typological Approaches. In: H. Caseli et al. (Eds.), **PROPOR 2012, LNAI 7243**. Springer: p. 329-337, 2012.
- \_\_\_\_\_; ERIKSSON, A.; ÅKESSON, J. Cross-Ling. similarities and differences of lexical stress realisation in Swedish and Brazilian Portuguese. In: E.L. Asu and P. Lippus (Eds.), **Proceedings of the Nordic prosody. XIth Conference**, 2013, Tartu. Frankfurt: Peter Lang, p. 97-106, 2013.
- \_\_\_\_\_. Automatic duration-related salience detection in Brazilian Portuguese read and spontaneous speech. In: **Proceedings of the Speech Prosody**, Chicago: 100067:1-4, 2010. Disponível em: <<http://www.speechprosody2010.illinois.edu/papers/100067.pdf>>
- BOERSMA, P.; WEENINK, D. **Praat: doing phonetics by computer (Computer program)** Disponível em: <<http://www.praat.org>> Versão 5.2.44, 2013.
- CAMARGO, Z.; MADUREIRA, S. The acoustic analysis of speech samples designed for the Voice Profile Analysis Scheme for Brazilian Portuguese (BP-VPAS): long-term f0 and intensity measures. In: **Tutorial and Res Workshop on Experimental Ling.**, Grécia. University of Athens. Grécia: ISCA Int. Speech Com Association, 1, p. 33-36, 2010.
- \_\_\_\_\_.; et al. Voice quality and gender: some insights on correlation between perceptual and acoustic dimensions. In: **Proceedings of the 6th Int. Conference on Speech Prosody**, Shanghai: Tongji U P, v. 01, p. 115-118, 2012.
- \_\_\_\_\_. Voice disorders in elementary and secondary school teachers: some reflections on acoustic and perceptual profiles. In: **Proceedings of the 10th Pan-European Voice Conf. 10th Pan-European Voice Conf Prague**, Praga: Pan European Voice Association, v. 01, p. 411-411, 2013.
- \_\_\_\_\_. Voice quality in Acquired Immunodeficiency Syndrome (AIDS): some perceptual and acoustic findings. In: **Proceedings of the 10th Pan-European Voice Conf. 10th Pan-European Voice Conf Prague**, Praga: Pan European Voice Association, v. 01, p. 371-371, 2013.
- \_\_\_\_\_. Acoustic and perceptual correlates of voice: trends between short and long-term analysis. In: **Proceedings of the 10th Pan-European Voice Conf**, Praga: Pan European Voice Association, v. 01, p.377-377, 2013.
- CUMMINS, F.; PORT, R. **Rhythmic constraints on stress timing in English**, J. Phon., v. 26, p.145-171, 1998.
- DOGIL, G. Phonetic correlates of word stress. In: Van der Hulst, (Ed.), **Word Prosodic System of**

- European Languages**, Berlin: p. 371-376, 1995.
- \_\_\_\_\_.; BRAUN, G. **The PIVOT model of speech parsing**, Verlag, Wien, 1988.
- FRY, D. B. Experiments in the perception of stress, **Lang. and Speech**, 1:126-152, 1958.
- LAVIER, J et al. **A perceptual protocol for the analysis of vocal profiles**. Edinburg U Department of Ling. Edinburg: v. 14, 139-155, 1981.
- MADUREIRA, S. The investigation of speech expressivity. In: Mello, H.; Panunzi, A.; Raso, T. (Eds.). **Illocution, modality, attitude, information patterning and speech annotation**. Firenze: Firenze U P, v. 01, p.101-118, 2011.
- \_\_\_\_\_. Reciting a sonnet: production strategies and perceptual effects. In: **Proceedings of the Speech Prosody**, São Paulo: RG, v.01, p. 697-700, 2008.
- \_\_\_\_\_.; CAMARGO, Z. Exploring sound symbolism in the investigation of speech expressivity. In: **Proceedings of the ISCA Tutorial and Res Workshop on Experimental Ling.**, Greece: University of Athens, v. 01, p. 105-108, 2010.
- PESSOA, A. et al. Perceptual and acoustic correlates of a speech in a bilateral cochlear implant user. In: **Proceedings of the 6th Int. Conf on Speech Prosody**, Shanghai: Tongji U P, v. 01, p. 51-54, 2012.
- \_\_\_\_\_. Dados de dinâmica e qualidade vocal: correlatos acústicos e perceptivo-auditivos da fala em criança usuária de implante coclear. **J of Speech Sciences**, v. 01, p. 17-33, 2011.
- RUSILO, L. C.; CAMARGO, Z.; MADUREIRA, S. The validity of some acoustic measures to predict voice quality settings: trends between acoustic and perceptual correlates of voice quality. In: **Proceedings of the 4TH Isca Tutorial and Res Workshop on Experimental Ling.**, Athens: University of Athens, v. 01, p. 115-118, 2011.
- SLUIJTER, A. M.C. **Phonetic Correlates of Stress and Accent**, Ph.D. Thesis. Leiden: Holland Inst of Generative Ling., 1995.
- TILSEN, S.; JOHNSON, K. **Low-frequency Fourier analysis of speech rhythm**, JASA Express Letters, v. 124(2), EL34, 2008.
- TRAUNMÜLLER, H.; ERIKSSON, A. **The frequency range of the voice fundamental in the speech of male and female adults**, Unpublished Manuscript. Disponível em: <<http://www.ling.su.se/staff/hartmut/aktupub.htm>>
- WIGHTMAN, C.W. et al. Segmental durations in the vicinity of prosodic boundaries, **J of the Acoustical Soc of America**, v. 91(3); p. 1707–1717, 1992.

Sonoridades é uma publicação que aborda conteúdos relacionados à análise da expressividade oral, ao trabalho com a expressão oral, ao enfoque da gestualidade vocal e visual e à expressão de sentidos na fala, no canto e na declamação.

