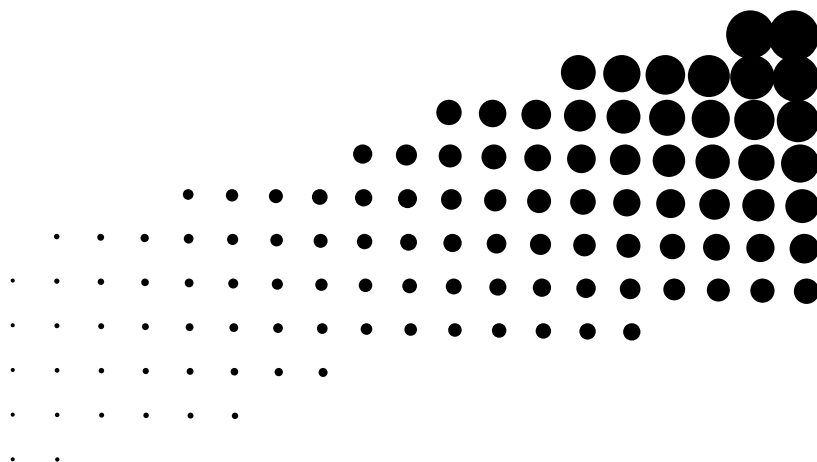




6th EICEFALA

International Meeting on Speech Sciences

19th to 21st September 2018 - Belo Horizonte - UFMG



ABSTRACT BOOKLET

Organizing Committee

Adriano Barbosa (UFMG)
Hani Camille Yehia (UFMG)
Leonardo Araujo (UFSJ)
Maria Cantoni (UFMG)
Rosana Passos (UFMG)
Thaís Cristófaró Silva (UFMG)



<http://www.lettras.ufmg.br/eicefala/>

EICEFALA 2018

19th to 21st September 2018

Plenary Session

Eric Vatikiotis-Bateson's View on Coordination in Human Communication

Hani Yehia and Adriano Vilela

In this talk, we try to present the late Professor Eric Vatikiotis-Bateson's view on coordination in human communication. Although Prof. Bateson was a linguist by training, he wasn't so much concerned with formal linguistics. Instead, he was more interested in looking at speech as a human communicative behavior: a complex, dynamic behavior that unfolds over time whenever people engage in verbal communication. This view had a profound influence on the questions he asked, the methodology he developed to address those questions, and the tools he created in trying to answer them. One of the main consequences of this view was that, in order to better understand the human communicative behavior, one needed to measure both the auditory and visual components of speech and try to make sense of how these two modalities work together to convey information. In this context, one of Prof. Bateson's main interests was coordination. Coordination is ubiquitous in biological systems and has been observed at various levels during verbal/visual interaction. It is characterized by temporal fluctuations in the coupling between measured domains. Conventional correlational techniques are not able to capture these fluctuations and therefore miss important aspects of the relationship between the domains of interest. In order to overcome this problem, a technique called Correlation Map Analysis (CMA) was developed at Prof. Bateson's lab which allows time-varying patterns in the correlation between measured signals to be identified. Although initially developed to study coordination in communicative interaction, this technique has been applied successfully to various other contexts. In this talk, we present a brief summary of Prof. Bateson's view on coordination, the methods and tools he developed in trying to characterize it, and some of the main results he obtained along the way.

Using correlation maps to analyze spoken communication: measuring the coupling between motion and acoustics during speech

Adriano Vilela

This talk presents a methodology for assessing the coupling between different modalities of speech (acoustic, visual) during spoken communication. The techniques presented can be equally applied to both intra- and inter-speaker scenarios. The talk focuses on two main points: i) the use of the FlowAnalyzer software to measure motion from video and ii) the use of CMA (Correlation Map Analysis) as a means of assessing the time-varying coupling between two domains of interest. We first show how to use FlowAnalyzer to extract motion signals from pre-recorded video sequences. We demonstrate the utility of the tool by showing how it allows the experimenter to rely only on ordinary video cameras to measure motion in a completely non-invasive way during speech production experiments. We then show how to use the CMA technique to quantify the time-varying coupling between the visual and acoustic components of speech. In our demonstration, the visual domain will be represented by the motion signals produced by FlowAnalyzer, whereas the acoustic domain will be represented by some parameterization of the acoustic waveform, such as the root mean square (RMS) value or Line Spectrum Pair (LSP) coefficients. We finish by discussing possible directions for future development of the tools and techniques presented.

Behavioral and Neurophysiological Representations of Speech Phonemic Units

Rafael Laboissière

Scientific studies have concluded about the possibility of humans perceive the world by categorizing the sensory input. This was showed, among others, by the perception of color, facial expressions and also the speech. This work aims to investigate the neurophysiological mechanisms involved in the categorization of the phonemic units of human speech. Electroencephalographic (EEG) measurements were performed in 17 healthy subjects during one passive task of phonemic categorization, compounded of a continuum of 100 phonemes which vary from the syllable /ba/ to /pa/, from which 4 phonemes were selected, in a customized way, according with the psychometric curve of each subject, and presented in a random way during the task. The phoneme continuum was constructed by the variation of the voice onset time (VOT) of the pure phonemes /ba/ and /pa/, which is a temporal characteristic of the phoneme. The main hypothesis is that the psychophysics representation of phonemes which are physically similar, like the /ba/ and /pa/ closer to the transition region of the psychometric

curve, will be similar to that representation of the pure phoneme located in the extremes of the continuum. A discrete wavelet transform (DWT) was performed over the EEG signals aiming for a time/frequency representation useful in the interpretation of the results and providing a reduced number of coefficients compared with other kind of transforms. Following, a linear regression by the Elastic Net (EN) technique was performed over the transformed data considering two cases, (i) physical response and (ii) psychophysical response; both found through the psychometric curve. This regression technique is interesting because it performs the regression together with a coefficient regularization and also a feature selection. Thus, just the wavelet coefficients that really explain the responses are selected, improving the results interpretability and also computational efficiency. Coefficients in each regression case were after separated according with the brain hemisphere they represent, right or left, and, in the end, four scalograms were plotted representing, for each subject, their physical and psychophysical responses and, for each one, both hemispheres. From the scalograms analysis made so far, it was concluded that there are significant differences between both hemispheres, showing a laterality effect in the perception of the presented phonemes. Comparing the results for the physical and psychophysical responses, it is observed an statistically significant effect in the phonemic categorization, confirming the proposed hypothesis.

Gradient elision and devoicing in Brazilian Portuguese

Eleonora Albano

Gradient elision and devoicing in Brazilian Portuguese This presentation is based on ongoing collaboration with my former student Francisco Meneses, now at Universidade Federal de São Carlos, and my friend and colleague Ryan Shosted, from the University of Illinois at Urbana-Champaign, together with his student Sarah Johnson. We started by taking note that traditional approaches to phonology claim that Brazilian Portuguese final vowels are deleted under two prosodically similar conditions. These are: a) final unstressed position preceded by a voiceless consonant; and b) final unstressed position in external sandhi with a following vowel. As we regard phonetic gradients as part of the phonology, we set out to investigate whether the first process is really apocope or just final vowel devoicing. For the same reason, we set out to investigate whether the second process is really elision or just extreme reduction with blending. Thus, the theoretical aim of the research is to contribute to the understanding of these two processes by clarifying whether the final unstressed vowels under each of the conditions are deleted or just hidden. However, the methodological issues involved are so complex that led us to devise a new procedure for processing ultrasound tongue images. The method is based on optical flow and was previously applied to larynx images only. The participants were four speakers of Brazilian Portuguese, who read sentences designed to trigger the two phonological processes. The data were collected at the phonetics laboratory of the

University of Illinois. The results show some very consistent trends, but vary as to the extent to which the labile vowel is detectable in the articulatory signal. While most images of the first process are interpretable as evidence for gradience, one speaker shows signs of apocope. On the other hand, gradience seems to be pervasive in the second process. No speaker shows unambiguous signs of elision. The data raise many questions about statistical treatment, since splines are inherently involved in the image analysis. The transdisciplinary audience of this meeting seems to constitute a good forum to discuss them.

Prosodic and gestural coordination and acquisition of French focus marking: what comes first?

Mariapaola D’Imperio

Prosody plays a fundamental role in acquiring language. Although the prosodic characteristics of language input provided to children and their early ability to discriminate and use them in the language acquisition process have been widely studied (for a review see, for example, Morgan and Demuth, 2014), the prosody of child productions has not been equally investigated. Most of these studies have also focused on early childhood productions, such as crying, vocalizations, babbling and first word productions (for a review see, for example, Snow and Balog, 2002), while only a few studies have investigated the prosodic characteristic of complex language, such as discourse or narration, in childhood (e.g., Redford et al., 2012; De Ruiter, 2014), and multi-word utterances. Very recent work on the prosody of Italian preschool children, conducted within the AM approach (D’Imperio et al. to appear), has shown that 3 year-old children use of prosodic breaks is already adult-like in broad focus, declarative utterances, despite being different from a melodic point of view. Children’s ability to use prosody for signaling information structure is not mastered until late in development. English-, German-, or Dutch-speaking children, for instance, do not use adult-like prosodic cues to focus before the age of 8-10 years (Ito 2014, Chen 2011, de Ruiter 2009). Children’s use of gestures to focus discourse elements has not been thoroughly investigated yet, though we know that adults can use head nods and eyebrow movements to mark new elements in the discourse (House et al. 2001, Kim et al. 2014, Kramer and Swerts 2007) and that these gestures usually co-occur with prosodic cues to focus (Esteve-Gibert et al. 2014). The study presented here aims examines how French children develop their control of prosodic cues for focusmarking, and whether and how they integrate gestures with prosody in doing so. Our findings seem to point a language-specific, prosodic behavior, which might be found in typologically similar languages (i.e. edge languages) but not in other.

Movement in Infant Language Research

Laurel Fais

Infant movement has long been used as a parameter in certain methodologies, for example, conditioned head-turn or preferential looking tasks. However, in those cases, the movement of the infant is in service of the infant directing her visual attention to a (usually multimodal) stimulus. The measure of infant movement as an analysis of interest in itself has been an extremely effortful and time-consuming task. Only recently has it become possible to measure infant movement easily and quickly using an optical flow algorithm, and thus to use the amount of motion exhibited by an infant itself as a measure of his performance in a task. Similarly, infant understanding of movement in communicative contexts, for example of pointing gestures made by a caregiver, have also been well studied. Recently, however, using this same type of analysis, much finer measures of more focused caregiver movements have been analyzed with respect to their information potential for infant language learners. This talk will present an overview of the use of methodologies measuring movement and gesture in infant research. In addition, it will highlight some important advances made possible by these methodologies in our understanding of infant speech perception; social referencing and interaction; the acquisition of syntactic knowledge of phrase boundaries; and word learning.

L2 speech Research Methodological Options and Challenges

Rosane Silveira

Deciding on the appropriate research design (e.g., quantitative/qualitative/mixed method; exploratory/descriptive/explanatory) and accompanying research methods is always a challenge. In the area of L2 speech, research interests have ranged from speech perception, speech production, L2 pronunciation, L2 intelligibility, and a few studies have addressed the role played by formal instruction and/or training. These different research topics demand specific research designs, instruments and procedures for data collection and analysis. Some research topics are more suitable for laboratory contexts, while others pose the researcher with the challenges of gathering naturalistic data, or data obtained in a classroom context. Different methodological options are available whether researchers are concerned with language description, theory testing, language development, pedagogical issues, or even language processing. In this presentation, an overview of the most common methodological options in the field of L2 speech research is provided, and considerations are made about the methodological challenges faced when conducting empirical research in the area.

Oral Session 1: Speech Technology

Fatigue and sleepiness detection based on speech analysis

Carla Vasconcelos, Maurílio Nunes Vieira, Hani Camille Yehia

Background: Mental fatigue and sleepiness are well recognized determinants of human-error related accidents and incidents in aviation. In Brazil, according to CENIPA (Center for Investigation and Prevention of Aeronautical Accidents), the rate of accidents in the aerial modal is of 1 per 2 days. Human factors are present in 90% of these accidents. Case Report: This paper describes a retrospective study of the communication between a pilot and an air traffic control tower just before a fatal accident. The objective was the detection of fatigue and sleepiness of a pilot, who complained of these signs and symptoms before the flight, by means of voice and speech analysis. The in-depth accident analysis performed by CENIPA indicated that sleepiness and fatigue most likely contributed to the accident. Speech samples were analyzed for two conditions: (i) non-sleepy data recorded thirty-five hours before the air crash (control condition), which were compared with (ii) data from samples collected about one hour before the accident and also during the disaster (sleepy condition). Audio recordings were analyzed for the extraction of objective measures of the temporal organization of speech, such as hesitations, silent pauses, prolongation of final syllables, and syllable articulation rate. Discussion: The results showed that speech during the day of the accident had a significantly low elocution and articulation rates compared to the preceding day, indicating also that the methodology adopted in this study is feasible for detection of fatigue and sleepiness through speech analysis. This seems to be the first such application in aeronautical accidents.

Articulatory Measures of Speech Production using NDI Wave Articulograph

Gabriela Chaltein, Melchior Augusto Syrio de Melo, Hani Camille Yehia

This project aims to elaborate and implement a methodology based on modern measurement techniques to analyze existing relations between processes of production and perception of human speech. Instead of investigating the process of perception through operations done directly over the speech signal, such as filtering and adding noise, this project focus on altering the articulatory parameters of speech production, in order to analyze the perception process in a **layer** above the acoustic signal. The methodology starts with articulatory measurements carried out with an NDI Wave articulograph. These measurements are the base to define articulatory gestures effectively used by speakers and

their relation with the speech sounds produced. The results obtained from articulatory measurements will be then combined with human speech perception measurements based on responses to complex sounds evoked by the auditory brain stem. This is done by using the articulatory parameters that have been measured to synthesize the complex sounds used as stimuli to obtain evoked potentials. Besides that, categorizing tests will be used to associate evoked potential measurements and speech production configurations with the conscious perception capacity of variations in the signals used as stimuli. This presentation has a focus on analyzing articulatory measurements using the NDI Wave articulograph.

Auditory perceived distance model for the ITU loudness algorithm

Leandro da Silva Pires, Hani Camille Yehia, Maurílio Nunes Vieira

The International Telecommunication Union, Radiocommunication Sector (ITU-R) Recommendation BS.1770 for loudness measurement in multichannel audio is established as a de facto standard for audio companies and de jure for digital broadcasters. Although its frequency weighting accounts for acoustic effects of the head, the model is insensitive to source distance. Listening tests were undertaken to investigate the effect of auditory distance perception on loudness of noise, speech, music and environmental sounds. Based on the variations found, an adaptation of the ITU-R algorithm is proposed and evaluated against subject responses. Resulting differences in loudness levels were within the confidence intervals of the level differences indicated by the subjects in source-receiver distances commonly found in living rooms.

Corpus CEFALA 1: Audiovisual database of speakers for biometric, phonetic and phonology studies

Arlindo Follador Neto, Adelino Pinheiro Silva, Hani Camille Yehia

Human speech has been studied in different areas of knowledge, which range from biometry to phonetics and phonology. In the research conducted in such areas, speech samples are necessary resources for obtaining results and validating hypotheses. For this, samples of different speakers and contents are stored in audio files and organized into databases. Such databases allow the continuity, practicality and reliability of studies, eliminating the difficult and time consuming step of data collection. Moreover, they allow consistent comparisons between different studies. However, free access databases in the Portuguese language or recorded in controlled environments are rarely found. The objective of this research is to construct a free and public database of Brazilian Portuguese,

named Corpus CEFALA-1. The database comprising 104 speakers guided by a specific protocol for the collection of audiovisual speech samples recorded in a studio. The study presents the methodologies of processing, segmentation and organization of speech samples, statistical analysis, application to biometric verification and preliminary phonetic-phonological analyses.

Oral Session 2: Brazilian Portuguese

Phonatory and acoustic gradualness of devoiced vowels in Brazilian Portuguese

Maria Cantoni

Vowels, amongst other sounds called sonorants, are usually produced with vibrating vocal folds. Despite of this overall tendency, many languages across the world present both voiced and voiceless vowels, with or without phonological contrast (LADEFOGED; MADDIESON, 1996). Recent studies have shown that in Brazilian Portuguese (BP) devoiced vowels can occur as allophonic variations of regularly voiced vowels (MENESES, 2012 *inter alia*). This variation is usually favored by weak prosodic positions, such as posttonic syllables. Furthermore it has been argued that, as in many other languages, vowel devoicing may involve a progressive scale, rather than a well-defined opposition between voiced and voiceless segments. The present study advances this issue, by considering the phonatory characteristics of devoiced vowels. A dataset of four posttonic syllables with three different vowels in twelve BP words produced by five native speakers was evaluated, consisting of both acoustic and electroglottographic (EGG) signals. The EGG signal is obtained by electrodes positioned externally in the larynx during speech and represents the time variations of the contact area of the vocal folds (VIEIRA, 1997). It can be straightforwardly used to access voicing and phonatory properties. The results show that, regarding the span of voicelessness, devoiced vowels in BP can be indeed partially or fully devoiced. Additionally, devoiced vowels may be associated to articulatory weakening as well as phonatory adjustments such as breathiness. They may be further accompanied by coarticulation with the preceding consonant. The results presented suggest that devoiced vowels are involved in a gradual progression from fully voicing to voicelessness or deletion which can have an impact on the language's templates for syllable construction.

LADEFOGED, Peter; MADDIESON, Ian. The sounds of the world's languages. Oxford: Blackwell, 1996.

MENESES, Francisco de Oliveira. As vogais desvozeadas no Português Brasileiro: investigação acústico-articulatória. 2012. Dissertação de Mestrado. Campinas, UNICAMP, Instituto de Estudos da Linguagem.

VIEIRA, Maurílio Nunes. 1997. Automated measures of dysphonias and the phonatory effects of asymmetries in the posterior larynx. Doctoral Thesis. University of Edinburgh, Edinburgh.

Morphological awareness in writing acquisition by deaf students

Daniela Mara Lima Oliveira, Andreia Chagas Rocha Toffolo

The poor performance in reading and writing skills by deaf people has worried teachers and researchers working in this area (PEREIRA, 2014; ALMEIDA et al., 2015). Results of research on Portuguese language learning have shown that most of these students, even in more advanced grades, have a standard reading level typical of the initial grades (CAPOVILLA, 2004). The fact that they do not hear makes it difficult to familiarize them with language spoken in the country, which, in turn, directly influences the development of previous phonological analysis skills that allow the learner to discover the alphabetic principle. Therefore, it has been proposed the teaching / learning process focus on word recognition with emphasis on the exploration of orthographic tracks. Following this path, authors such as Nunes et al. (2006), Koppenhaver and Wollak (2014), Trussell et al., (2017) present satisfactory results considering the role of morphological awareness (CM) in reading and writing comprehension by the deaf student. Morphological awareness is the ability to manipulate and reflect about elements that carry meaning within the word (morphemes) intentionally. In this work, we aim to present a reflection on the already published analyzes about the subject in BP and to evaluate the possibility of creating an intervention program for Brazilian deaf readers, that can be used as a bilingual didactic strategy in the teaching of Portuguese Language as a second language for the deaf people.

Processual units in translation and comprehension tasks: an experimental approach

Matheus Freitas

This paper aims at presenting and discussing the results from an experiment which investigated the status of processual units in translation and comprehension tasks. Our main objective was to assess, during text processing, whether the whole text or individual words function as processual units. We thus contrasted two aspects of text processing: informational progression and individual

words. Jensen (2009) considers the concepts of textual complexity and processing difficulty, respectively, as an objective and a subjective approach to the text properties. Complexity is thus related to factual criteria concerning the text. Difficulty is related to the reader's/translator's perception of a text. Then, it is assumed that objective text properties affect subjective experience during translation and comprehension. Exemplar Models (JOHNSON, 1997; PIERRE-HUMBERT, 2001; BYBEE, 2002) consider that words function as processing units, being the locus of grammatical development. An experiment was carried out in order to examine how textual complexity variation affected text processing. Using a combination of eye-tracking and keystroke logging, we analysed translation and comprehension tasks performed by eight translation students from the Federal University of Ouro Preto – MG, Brazil. We consider that gaze (and typing) behaviour can serve as an indicator of cognitive effort during text processing (JUST; CARPENTER, 1980). The following data were collected and analysed: task duration; fixation count; fixation mean duration; pause frequency and pause duration. Texts labelled as less complex according to informational progress criteria presented higher difficulty levels than texts labelled as more complex. The analysed data also suggested the participants had more comprehension and translation problems with shorter units, such as words, instead of information progression problems, which are related to the text as a whole. For the translation tasks, the distribution of time and visual attention was higher to the source text area than to the target text area. With this research, we seek to collaborate on experimental research in language – especially when it comes to methodological development – with emphasis on works that use eye-tracking as a method for data collection. In view of the problem assessed in this analysis, we consider important to put more emphasis on other levels of linguistic processing, such as phonological patterns. For future work, we then propose an investigation which contrasts words and phonological patterns – such as the syllable or individual segments – as linguistic processing units.

Keywords: Experimental Linguistics; Translation Process; Cognitive Effort; Exemplar Models; Eye-tracking.

References:

- BYBEE, J. 2002. Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition*, v. 24, n. 2, 2002. p. 215-221.
- JENSEN, K.T.H. Indicators of text complexity. In S. Göpferich, A.L. Jakobsen and I. Mees, eds. *Behind the Mind*. Copenhagen Studies in Language 37. Copenhagen: Samfundslitteratur, 2009. p. 61-80.
- JOHNSON, K. Speech perception without speaker normalization: An exemplar model. In: JOHNSON, K.; MULLENIX, J. W. (eds.). *Talker Variability in Speech Processing*. San Diego: Academic Press, 1997. p. 145-165. JUST, M.A.; CARPENTER, P.A. A theory of reading: from eye fixations to comprehension. *Psychological Review* 87/4, 1980. p. 329-354.

PIERREHUMBERT, J. Exemplar dynamics: word frequency, lenition and contrast. In BYBEE, J.; HOOPER, P. (Ed.). *Frequency and the emergence of linguistic structure*. Amsterdam: J. Benjamins, 2001. p. 137-157.

Impoliteness as emergent behaviour: an analysis of the talk show Hell's Kitchen Brasil

Daniela Cristina Dias Menezes

The objective of this presentation is to analyse the emergence of impoliteness in the interaction between the chef and judge of the cooking television show Hell's Kitchen Brasil (season four) and its contestants. Impoliteness is sanctioned in this context, as viewers expect the chef to be a severe critic of contestants' dishes. Furthermore, as each episode involves the elimination of one participant, the designers of the show create situations that enhance individual flaws and maximize the rank of power between the chef and the contestants.

Intervocalic Tap Lenition in Brazilian Portuguese

Wellington Araujo Mendes Junior

This investigation reports on findings of tap lenition in several states of Brazil. Examples of intervocalic tap lenition include the variation observed in para [pa. ə] ~ [pa] and in brasileiro [b a.zi.lej.] ~ [b a.zi.lej.]. The corpus was collected through the website www.fonologia.org and it is made up of speech recordings resulting from the reading of a text from the project "Sibilants and Rhotics of Brazilian Portuguese" (CNPq, 484590/2013-8). The research presents insights on the consequences of segmental loss for phonological representations. Investigations were mainly based on the assumptions of the Exemplar Theory (JOHNSON, 1997; PIERREHUMBERT, 2001, 2003). Since tap lenition is an optional phenomenon in Brazilian Portuguese, this investigation sought to assess which conditions favor or inhibit such phenomenon, while trying to acoustically assess whether tap loss is gradient or abrupt. In order to carry out these tasks, two data analyses were performed: the categorical and the acoustic one. The categorical analysis evaluated the variables (1) tonicity, (2) number of syllables, (3) adjacent vowels, (4) lexical frequency, (5) lexical item, (6) sex, (7) individual and (9) geographical border. The data was submitted to the statistical program R (R CORE TEAM, 2013) and the following factors were significant in the statistical model: tonicity, number of syllables, following vowel, lexical frequency and geographical boarder. Furthermore, it was reported that intervocalic tap lenition occurs among speakers of, at least, 12 states of Brazil. This provides evidences that intervocalic tap lenition is an emergent and recurrent phenomenon

in Brazilian Portuguese. As for the gradient or abrupt implementation of tap lenition, results suggest that such phenomenon is gradient. In addition, results indicate that tap lenition implies gradual loss of lexical duration. As for the impact of the intervocalic tap lenition on mental representations, evidence was presented for the relevance of the phonetic detail in the implementation of lenition, corroborating principles of the Exemplar Theory. Key words: intervocalic tap, lenition, gradiency, Exemplar Theory.

Oral Session 3: Second Language

Non-native speech perception: reanalyzing PAM-L2 from the perspective of cognition, acoustic-articulatory gesture, and indirect realism

Reiner Vinicius Perozzo, Ubiratã Kickhöfel Alves

This presentation addresses the phonic perception of non-native languages and aims to review the principles of the Perceptual Assimilation Model for Second Language Speech Learning [PAM-L2 (BEST; TYLER, 2007)] with respect to its cognitive, phonological and philosophical tenets. According to this model, (i) the perception of non-native speech disregards any cognitive mechanisms related to mental representations or inferential processes, since our senses act as our own perceptual systems; (ii) the unit for analyzing perceptual events, in terms of speech, corresponds to the articulatory gesture (BROWMAN; GOLDSTEIN, 1989, 1992), which develops from the notion that coordinated articulations leads to contrastive lexical knowledge; and (iii) we have direct access to the information available in the world, bearing in mind that the surrounding environment affords invariant structures that are directly detected by the perceiver. We argue that these premises are limited and inconsistent with the research object of the authors, considering their own theoretical framework: (i') if mental representations and inferential processes are not involved in the perception of non-native speech, then we are not able to capture perceptual phenomena as relying upon an intellectual basis, a statement that intrinsically collides with the latest body of research conducted in neurosciences and neuropsychology; (ii') if speech perception assumes gestures as linguistic primes, then a sharp boundary between phonetic and phonological information should not be posited, since phonetic and phonological properties are brought together in a common mental-physical domain; and (iii') if non-native speech crucially depends on native speech categories, and native linguistic knowledge acts as a filter for non-native sound contrasts, then non-native speech elements cannot not be accessed directly. We raise these arguments as a counterpoint to the principles of PAM-L2 and state that the model should be glimpsed from an alternative point of view. Firstly, we argue that the perceptual event is essen-

tially a cognitive phenomenon (KANDEL, 2014; GAZZANIGA et al., 2012), created and managed by the brain, which involves abstractions, mental representations and inferences about the objects of the world. Secondly, we state that an acoustic-articulatory treatment (ALBANO, 2001) to the gestural unit is more appropriate to the perception of non-native phonic elements, differing from the articulatory treatment (BROWMAN; GOLDSTEIN, 1989, 1992) that is originally conveyed by the model. Thirdly, opposing direct realism (J. GIBSON, 1966, 1986; FOWLER, 1986), we adopt the indirect realist position (JACKSON, 1977, 2010; LOWE, 1981) as the one that encompasses the perception of non-native phonic units more adequately. Such an alternative version of the tenets of PAM-L2 also have an impact on issues such as the design, the type of knowledge tested (either native or nonnative language) and the purpose of each perceptual task to be used in a laboratory or similar environment. By revising these three domains of PAM-L2, we aim to propose a new version of the model that seems more coherent regarding the compatibility of cognitive, phonic and philosophical aspects of speech perception and language development. **KEYWORDS:** acoustic-articulatory gesture; cognition; indirect realism; nonnative phonic perception.

REFERENCES: ALBANO, E. O gesto e suas bordas: Esboço de fonologia acústico-articulatória do português brasileiro. Campinas: Mercado de Letras, 2001. BEST, C.; TYLER, M. Nonnative and second-language speech perception: Commonalities and complementarities. In: BOHN, O.; MUNRO, M. (Orgs.). Language experience in second language speech learning: In honor of James Emil Flege. Philadelphia: John Benjamins Publishing Company, 2007, p. 13-34. BROWMAN, C.; GOLDSTEIN, L. Articulatory gestures as phonological units. Haskins Laboratories Status Report on Speech Research, v. 100, p. 69-101, 1989. BROWMAN, C.; GOLDSTEIN, L. Articulatory phonology: An overview. Haskins Laboratories Status Report on Speech Research, v. 112, p. 23-42, 1992. FOWLER, C. An event approach to the study of speech perception from a direct-realist perspective. Journal of Phonetics. v. 14, p. 3-28, 1986. GAZZANIGA, M.; HEATHERTON, T.; HALPERN, D.; HEINE, S. Psychological Science. Nova Iorque: W.W. Norton & Company, 2012. GIBSON, J. The ecological approach to visual perception. Nova Iorque: Psychology Press, 1986. GIBSON, J. The senses considered as perceptual systems. Boston: Houghton Mifflin, 1966. JACKSON, F. Perception. Nova Iorque: Cambridge University Press, 1977. JACKSON, F. Representative realism. In: DANCY, J.; SOSA, E.; STEUP, M. (Eds.). A companion to epistemology. 2. ed. Malden: Blackwell, 2010. p. 702-705. KANDEL, E. Das células nervosas à cognição: As representações internas de espaço e ação. In: KANDEL, E.; SCHWARTZ, J.; JESSEL, T.; SIEGELBAUM, S.; HUDSPETHET, A. (Eds.). Princípios de neurociências. 5. ed. Porto Alegre: Artmed, 2014. p. 327-344. LOWE, J. Indirect perception and sense data. The Philosophical Quarterly, v. 31, n. 125, p. 330-342, 1981.

The development of speech rhythm of English by Brazilian learners: a research proposal

Ronaldo Manguiera Lima Júnior

There are several prosodic differences between English and Brazilian Portuguese that pose difficulties for Brazilian learners of English both in production and perception. Concerning rhythm, the impression that these two languages have distinct patterns has long been reported (Lloyd James, 1940), initially as the former having similar interstress intervals and the latter similar syllable intervals (Pike, 1945). However, despite the unchallenged impression that these two languages present distinct rhythmic patterns, such isochrony has never been found in acoustic measurements (e.g. DEMANRIQUE & Signorini, 1983; Lea, 1974; O'Connor, 1965; Shen & Peterson, 1962). This has led scientists to look for other acoustic cues to justify such impression, with three current models displaying different levels of success in arranging the so-called stress-timed and syllable-timed languages (such as English and Brazilian Portuguese, respectively) within a gradient classification – as opposed to the initial dichotomic proposals. The first one was Ramus, Nespor and Mehler's (1999) proposal in which they plotted the data of eight different languages using the relative duration of vocalic intervals in one axis and standard deviation of consonantal intervals in the other, being able to position languages according to their impressionistic stress-timing or syllable-timing classification, yet in a gradient, more realistic, manner. A similar result was achieved by Grabe and Low (2002) as they analyzed the PVI's (Pairwise Variability Indices) of vocalic and intervocalic intervals in 18 languages. Finally, both O'Dell and Nieman (1999) and Barbosa (2002, 2006) have used a general mathematical model, namely the Coupled-Oscillator Model, to successfully classify the so-called stress-timed and syllable-timed languages in a gradient fashion using one single parameter: the coupling strength. From a dynamic perspective of linguistic rhythm, this model assigns vowel duration to the syllabic oscillator, and stress to the stress oscillator. It is possible to mathematically estimate the coupling strength from the speech signal, and thus infer which oscillator has more influence over the other both among different languages and among different varieties (dialects, elocution rates, recording tasks, etc.) within the same language. Even though all these models have already been used to account for the rhythm of several languages, very little work has been done with the two first models and none with the third one on the speech of second or foreign languages. Therefore, the research project herein proposed aims at using these three models (Ramus et al's 1999, Grabe & Low 2002, Barbosa 2006) to investigate the development of the rhythmic patterns of English as a Foreign Language as spoken by Brazilian undergraduate students of English. To do so, three corpora of the oral production of 10 Brazilian undergraduate students of English and one corpus of oral production of 10 native speakers of English will be used. The three corpora of the learners will be: oral production in English at the beginning of their second semester of college studies, oral production in En-

English in their fifth semester of college studies, and oral production in Portuguese. This way, the data of English-L1, Brazilian Portuguese-L1 and English-L2 (in two longitudinal moments) will be used to run the three aforementioned models. Besides its main goal of investigating the development of rhythmic patterns of English by Brazilian learners, this study can potentially also be used to compare and evaluate the three current models of speech rhythm. The data haven't been collected yet, and the opportunity to present this project proposal and receive suggestions and feedback before it begins will be of paramount importance to its success.

REFERENCES Barbosa, P. A. (2002). Explaining cross-linguistic rhythmic variability via a coupled-oscillator model of rhythm production. In *Speech Prosody 2002, International Conference*. Barbosa, P. A. (2006). *Incursões em torno do ritmo da fala*. São Paulo: Pontes. DEMANRIQUE, A. M. B., & Signorini, A. (1983). Segmental duration and rhythm in Spanish. *Journal of Phonetics*, 11(2), 117–128. Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology*, 7(515–546). Lea, W. A. (1974). Prosodic aids to speech recognition: IV A general strategy for prosodically-guided speech understanding. . St. Paul, Minnesota: Sperry Univac. Lloyd James, A. (1940). *Speech signals in telephony*. London: Sir I. Pitman & sons, Limited. O'Connor, J. D. (1965). *The perception of time intervals*. London: University College London. O'Dell, M., & Nieminen, T. (1999). Coupled oscillator model of speech rhythm. In *Proceedings of the XIVth international congress of phonetic sciences* (Vol. 2, pp. 1075–1078). Pike, K. L. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press. Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265–292. Shen, Y., & Peterson, G. G. (1962). Isochronism in English. *Occasional Papers*, (9), 36.

Language Attrition in Initial Plosives in Bilingual and Trilingual Speakers in an L1-dominant Environment

Laura Castilhos Schereschewsky, Ubiratã Kickhöfel Alves; Felipe Flores Kupske

This study aims to discuss language attrition in the Voice Onset Time (VOT) production patterns of Brazilian Portuguese (BP) (L1) initial voiceless plosives by bilingual (BP and English) and trilingual (BP, English and German) speakers in an L1-dominant environment. In order to discuss this phenomenon, we analyzed the VOT (the time interval between the release of a plosive consonant and the beginning of the following vocal vibration) of the bilabial, alveolar and velar voiceless plosives (/p/, /t/, /k/), in word-initial position, in the three referred languages. The study analyzed data from 40 participants. The experimental groups consisted of 10 bilinguals (BP - English) and 10 trilinguals (BP, English, German), all native speakers of Portuguese, residents of the city

of Porto Alegre (Southern Brazil). Control data were obtained from 10 monolinguals of Portuguese, also residents of Porto Alegre, and 10 monolinguals of English, residents of London (cf. KUPSKE, 2016). The participants individually took part in reading tasks of carrier sentences in their known languages, such as “Eu diria “, *in Portuguese*, “*I would say*”, in English, and “Ich sage —“, in German, followed by words beginning with the target consonants. As results, we found significant differences between the L1 and the additional languages of each of the plurilingual groups; however, we did not find differences between the L2 and L3 of the trilingual group. When compared to the English monolingual control group, the Brazilian participants did not show significant differences for /t/ and /k/ in English, thus indicating a native-like pattern. When compared to the BP monolingual control group, the two plurilingual groups showed significantly different productions. These results suggest the occurrence of attrition even in non-dominant L2 or L3 environments, also providing evidence in favor of the multidirectionality of language transfer and the importance of typology in the development of additional languages. Furthermore, these results reinforce the dynamic character of language, predicted by the view of language as a Complex Adaptive System (CAS) (LARSEN-FREEMAN & CAMERON, 2008; BECKNER et al., 2009; DE BOT et al., 2013).

Oral Session 4: Prosody

Prosodic description of the short parenthetical information unit in spoken Brazilian Portuguese

Saulo Mendes Santos, Giulia Bossaglia

This work presents the first results of a prosodic description of a subtype of the information unit of Parenthesis (PAR), the short modal PAR, based on a representative sample of the Informal section of the C-ORAL-BRASIL corpus (Raso & Mello 2012) of spoken Brazilian Portuguese (PB). According to the adopted theoretical framework, the Language into Act Theory (Cresti & Moneglia 2005; Moneglia & Raso 2014), PAR is defined as an inserted semantic information, compositional to the rest of the utterance, with metalinguistic, metanarrative or modal value, restricting and clarifying how a part of the locutive content should be interpreted by the addressee. Previous research carried out on the Italian C-ORAL-ROM corpus indicates that the prosodic profile of PAR is characterized by changes in fundamental frequency (f0), intensity, and speech rate with respect to the rest of the utterance, and that PAR can be found in any position of the utterance, except for the initial one (Tucci 2010). Santos & Bossaglia (2018, in preparation) have also found out that short PAR tokens have a slope to bear a modal function and that several tokens are similar to what has been described in the literature as “reduced parenthetical clauses” (Schneider 2007). For this study, PARs were extracted from a sample of the

C-ORAL-BRASIL corpus, searching for typical fillers of reduced parenthetical clauses – expressions such as “I think”, “I would say”, “I suppose”, “I hope”, “it seems” – enclosed by two prosodic breaks, thus fulfilling an exclusive tone and information unit. Then, 30 tokens bearing modal function and distribution of PAR were randomly chosen. In order to describe the profile of short modal PAR, the following prosodic parameters were measured and calculated: a) mean f0 of the final or initial part of the information units neighbouring PAR (from the last stressed syllable included to the prosodic break or vice versa); b) mean f0 of PAR; c) ratio of speech rate (phonological content) of PAR to the ones of neighbouring units; d) ratio of speech rate (phonetic content) of PAR to the ones of neighbouring units; e) ratio of f0 range of PAR to f0 range of the host utterance; f) ratio of f0 range of PAR to f0 ranges of neighbouring units; g) ratio of intensity range of PAR to intensity range of the utterance. Furthermore, two non-prosodic variables were considered: a) speaker’s sex; and b) the distribution of PAR within the utterance (medial and final). Partial results suggest that short modal PAR has a prosodic profile with a higher speech rate and a lower f0 range with respect to its neighbouring units and that mean f0 of PAR also varies depending on the position of the token within the utterance, being lower in final position.

REFERÊNCIAS

- CRESTI, E.; MONEGLIA, M. (Ed.). (2005). C-ORAL-ROM. Integrated reference corpora for spoken romance languages. Amsterdam/Philadelphia: John Benjamins.
- MONEGLIA, M.; RASO, T. (2014). Notes on the Language into Act Theory. In: RASO, T.; MELLO, H. (Ed.), Spoken corpora and linguistics studies. Amsterdam/Philadelphia: John Benjamins, p. 468-494.
- RASO, T.; MELLO, H. (2012). C-ORAL-BRASIL: corpus de referência do português brasileiro falado informal. I. Belo Horizonte: UFMG.
- SANTOS, S.; BOSSAGLIA, G. (2018). Morphosyntactic, prosodic, functional and distributional description of the information unit of Parenthesis in spoken Brazilian Portuguese. In preparation.
- SCHNEIDER, S. (2007). Reduced parenthetical clauses as mitigators: A corpus study of spoken French, Italian and Spanish. Amsterdam/Philadelphia: John Benjamins.
- TUCCI, I. (2010). Obiter dictum: la funzione informativa delle unità parentetiche. In: LEONI, F. A. et al. (Ed.). Atti del Convegno internazionale G.S.C.P. La comunicazione parlata, Naples: Università degli Studi di Napoli L’Orientale, 2010, v. I, p. 635-654.

Prosodic boundaries and automatic intonation analysis

Waldemar Ferreira Netto, Marcus Vinicius Moreira Martins

The purpose of this research is to establish criteria for the automatic phrasal segmentation of Brazilian Portuguese. In previous works we have defined that the intonation can be understood as a system composed of two fundamental elements: the middle tone, extracted from time-accumulated averages and two lateral intervals of maximum and minimum F0 values, which establish references for emphatic and emotional intonation analyzes. Values below or beyond these thresholds are considered Emphasis/Focus (E/F). Although the method is functional for the segmentation and analysis of phenomena at the syllabic level, the replication of the same method is not functional for segmentation of lexical or syntactic items, therefore prosodic boundaries or sentence boundaries are not detected by the automatic analysis proposed by us. For the analysis in this paper, only phonetic and phonological criteria were considered to determine the prosodic boundaries of segments, regardless of whether they behave as whole sentences or not, so the syntactic criterion was not considered at this stage. Thirty samples of spontaneous speech, from male voices, taken from headlines of radio transmissions are analyzed in 7 parameters: intonation (movement), direction of intonation (direction), target of intonation (target), interruption of tonal rhythm (pause > 0.05ms), UBI elongation (stretching), correlation between mean tone and total mean tone (correlation), difference between mean and total mean tone (difference) and intonation dispersion above the mean pitch (F/E). The prosodic breaks would be defined: (i) by the occurrence of only one significant parameter ($p < 0.05$); (ii) by a set of parameters that are also significant. The analysis of these parameters demonstrated that for the definition of boundaries it would be necessary: (a) the occurrence of an upward interval with a value greater than F/Esup between the moments $z(t-1)$ and $z(t)$, where the moment $z(t-1)$ is less than or equal to F/Einf (t) - significant parameters: tonal target, intonational movement, difference between mean and total mean tone, and intonation dispersion above mean pitch (F/E); (b) interruption of tonal rhythm with duration greater than or equal to twice the duration defined for a pause - significant parameters: interruption of tonal rhythm; (c) interruption of the tonal rhythm with duration greater than or equal to the lowest value defined for pause followed by a tonal range greater than or equal to F/Esup or less than or equal to F/Einf - significant parameters: interruption of tonal rhythm and intonation dispersion above the midtone (F/E). Further research is necessary to characterize the types of boundaries determined by the parameters presented here, however, it is noted that only the combination of parameters was able to determine prosodic boundaries.

Oral Session 5: Speech Technology

Phonetic Spell Checker

Leonardo Araujo

In text editing, words that are not present in a given dictionary are considered unknown or misspelled. Those strings are flagged by a spell checker which should also provide a correction or a list of suggestions. Essentially, a string is not found in a given dictionary due to the possible factors: typing errors or spelling error, when regarding human's source of errors (what might be understood as a noisy channel); the presence of interfering noise in a communication channel, whether regarding transmission, storage, printing errors or an automatic character recognition system, among other possible non human sources; or a given word does indeed exist but it simply is not listed on the system's dictionary. Spelling errors might be classified into whole word errors, faulty graphemes or faulty phonemes. In general, spell checkers intend to provide correction suggestions for the last two classes of errors, since both of them, in written form, might be a combination of the following grapheme edition: deletion, insertion, substitution or transposition. Automatic correction is indeed part of larger problem, concerning the recognition and matching among a universe of erroneous data.

When an erroneous word is found, spell checkers usually consider a list of word candidates that might be generated from the faulty word applying a combination of the four type of grapheme edition, as pointed before. To choose between the potential candidates, we need distance metric between two strings (the potential source string and the observed erroneous string). A non probabilistic approach is the minimum edit distance algorithm, which solves the problem by means of dynamic programming. The minimum edit distance is the minimum number of edit operations (insertion, deletion, substitution and transposition) necessary to transform one string into another. Each operation might have a different cost. The Levenshtein distance is the simplest measure, where only three operations are allowed (insertion, deletion, substitution) and each of them has the same cost of 1. Levenshtein also proposed another metric in which substitutions are not allowed (or equivalently, they have a cost of 2, since any substitution might be represented by a deletion followed by an insertion). The Damerau-Levenshtein distance includes the four edit operations and assign a cost of 1 to all of them (substitution will cost nothing if the characters involved are the same).

A large dictionary has proven important to avoid making rare words as errors. It is specially true when using word frequency as a factor in spelling correction, what is used in modern spell checking systems. Peter Norvig (2007) proposed a spell checker which uses the probability of the edit candidates from a erroneous string. A first improvement over his spell corrector is to consider both word probability and minimum edit distance. Faulty graphemes spelling errors are the major target of this approach. Still some spelling errors might not

be efficiently corrected by this approach, such as those created by insufficient linguistic knowledge or faulty phonemes, when the misspelled word does not sound like the target word, when a phoneme is inserted, removed, substituted or transposed.

To take in account those phonological errors that lead to misspelling, it is here proposed a spell checker, based on the Norvig's approach and the minimum edit distance considering each word as a string of phonemes, rather than a string of characters. The sole use of this approach decreases the spell checker performance, but using it in conjugation with the original Norvig's spell checker proves to increase in 10% the correction performance.

Robustness in forensic speaker comparison: The great quest

Adelino Pinheiro Silva, Adriano Vilela Barbosa, Maurílio Nunes Vieira

the latter has more control of some variables while the former deals with recordings from different origins. In FSC, the rule is to compare a sample pattern with a crime vestige. Also, the stakes are higher in FSC, because of the risk of wrongly convicting an innocent person or acquitting a guilty person.

Within this scenario, two dilemmas can be pointed out. The first is how to make a reliable, transparent, and reproducible speaker comparison, considering the new paradigms of forensic science and human rights. The second is how to test and evaluate existing speaker comparison techniques. In this work, we considered the effects of noise, communication channel coding, microphone type, and time lapse between the compared voices.

The hunt begins by tracing the steps of the forensic sciences and speaker comparison. It was necessary to discuss the interface of forensic science and the Brazilian legal environment, the main acoustic measurements in use in forensic speaker comparison, as well as the techniques for statistical classification.

Three approaches have been used. The first is based on the separation of the speech signal in blocks with similar signal-to-noise ratios (SNR). The second approach applies the Full Bayesian Significance Test (FBST) to find the credibility interval of forensic speaker comparison. The third approach applies the combination of different acoustic features in order to reduce equal error rate (EER). This approach comes from a baseline a study of different features used in speaker comparison in noisy conditions, such as GSM (Global System for Mobile communications) channel contamination.

The main objective of block separation is to allow comparisons between more homogeneous data to be made. The challenge is to realize automatic segmentation

and comparison. Results show that blocking by acoustic similarity improved accuracy in 5.5 %. Blocking by SNR further decreased the EER from 2.2% to 1.7%, on average. The method can potentially improve forensic speaker comparison.

In the FBST approach, the main challenge was to apply the test to a large number of observations and to formulate an equation to solve the test faster. Comparisons with other interval inference methodologies indicated that FBST is more conservative in relation to confidence intervals. This helps to reduce the risk of wrong association of an innocent person.

The study of speaker comparison in GSM and noise conditions include different approaches to calculate the likelihood ratio and the acoustic features. So far, mel-frequency cepstrum coefficients (MFCC) and universal background model with gaussian mixture models (UBM-GMM) methodology have shown the best results. When used to compare GSM encoded samples at an SNR of 17 dB, the MFCC produces an EER of less than 10%, on average.

Quantifying the Multivariate Instantaneous Association between the Acoustic and Visual Domains during Speech

Luciano Bruno Domingos Neves, Adriano Vilela Barbosa

This work presents a method for quantifying the instantaneous association between the acoustic and visual domains in audiovisual speech processing. An important limitation of previous approaches is eliminated in that the instantaneous association can now be computed between two groups of signals as opposed to just two signals.

Two measures of association are defined. The first one (v) captures the shared variance between the groups of variables by mapping the linear relationship between them. In order to establish each group's total variance, Principal Component Analysis (PCA) and Canonical Component Analysis (CCA) are used to remove redundant information by diagonalizing the covariance matrix.

The use of CCA provides another definition of association: one that estimates the probability of the two groups being independent (h). Time-varying fluctuations are captured by using an exponential moving average filter to estimate the instantaneous covariance between variables. The method was applied to two databases collected during speech production experiments.

The first database contains measurements of vocal tract motion of two speakers collected by electromagnetic articulography (EMA) while they were interacting with each other.

The second database contains vocal tract motion, face motion and speech acoustics collected for a single speaker during a non-spontaneous speech task.

Once again, vocal tract motion was measured by electromagnetic articulography, whereas face motion was captured by optical sensors and the speech acoustics was represented by Line Spectrum Pairs (LSP).

The method was able to detect how the association between domains changes over time. For the first experiment, it was found that the association between the groups of variables is closely related to the first principal component within each group; it was also possible to identify which variables were most descriptive of the overall association between the groups and how this behavior changed over time. In the second experiment, a strong relationship was observed between the instantaneous association and the energy of the speech acoustics, something that could potentially lead to future improvements to audiovisual coding algorithms.

Poster Session 1

Estimating a psychometric function for breathiness perception

João Pedro Hallack Sansão, M. N. Vieira, H. C. Yehia

The main objective of this work was to estimate a psychometric function for vocal breathiness perception. For this purpose, an adaptive procedure was used to adjust the parameters of the psychometric functions, the QUEST method implemented by Palamedes toolbox.

In this method, a model is estimated after each psychophysical trial, which serves to select the next stimulus to be presented. Two experiments were carried out, with different types of stimuli:

- (a) pure tone with additive pink noise;
- (b) synthetic voice with pink noise added to glottal source.

Data for calculating perceptual thresholds was obtained through psychometric tests with alternative forced choice. Stimuli were presented to the judges by headphones connected to a computer sound card. Judge's basic task in each essay, after being presented to two stimuli, was to define which one presented the lower noise level.

As levels were previously set by the method, it was possible to assess the input correctness and calculate the next step. For each experiment, different signal-to-noise (SNR) ratio levels have been tested, since the psychometric functions are defined locally (i.e.: noise perceptual threshold levels may be different under different SNR levels).

Thus, for each level a fixed reference stimulus was used in the comparisons and variable one was selected by the method. Presentation order of each sample in the pair was random.

For the pure tone experiment, the sinusoid had a frequency of 220 Hz, 1.5 seconds long. SNR levels of 10, 20, 30, 40 and 50 dB were tested.

For the synthetic speech experiment, a 220 Hz glottic signal was generated and filtered by a vocal tract simulating the vowel /a/. The noise was added in the pre-filtering phase. SNR levels were placed in 10, 20, 30 and 40 dB.

Three subject participated in each experiment.

For the pure tone experiment, breathiness perceptual threshold varied between 0.8 ± 0.3 dB (for a 50 dB SNR level) and 1.9 ± 0.6 dB (for a 10 dB SNR level).

In the synthetic voice experiment, detected thresholds were 1.7 ± 0.5 dB, 1.3 ± 0.5 dB, 1.5 ± 0.5 and 2.3 ± 0.3 for SNR levels of 40, 30, 20 and 10 dB, respectively.

The experiments also determined parameters for each psychometric function.

Phonological path in language development by twins: segmental, syllabic and phonotactics aspects

Jéssica Caroline Souza Aguiar, Vanessa Cordeiro de Souza Mattos, Laís Rodrigues Silva Bockorni, Maria de Fátima de Almeida Baia

This paper aims at analyzing the phonological path in language development by one pair of twin girls, considering segmental, syllabic, and phonotactics aspects, in order to investigate the mismatch regarding the development of twins as pointed out by the literature (LEONARD et al., 1980; SMITH, 2011). For the discussion, we follow the Complex Adaptative Systems (CAS) (LARSEN-FREEMAN, 1997; BAIA, 2013), perspective which comprehends language development as gradual and emergent. Studies about twins' linguistic development are still few, however there is already evidences of a mismatch in the linguistic development of twin children, generally due to the fact that one child tends to receive less directed speech, namely, input (SMITH, 2011). In addition, the literature also observes some tendencies in twins' speech, such as i) they tend to complete sentences from each other; (ii) they tend to use less their own name and; iii) to have their own linguistic system (idioglossia/cryptophasia). In this study, we analyze longitudinal data of Bg and Mg, aged 1-2 years old, dizygotic twins, in the process of development of Brazilian Portuguese spoken in Vitória da Conquista - BA. In the analysis, we investigated the phonological path of Bg and Mg and some differences were found as reported by the literature. For instance, whereas Bg realized in her first session the segments [i, u, â, n, d], Mg presented only the vowel [u]. Nevertheless, the mismatch did not represent a significant delay in Mg development as at 2;0 years both of them presented the same vocalic system [i, e, , u, o, , a, â, õ] and consonantal segments. In relation to the syllabic path, Bg presented $V > CV > VV > CVV > CV(V)C$ and Mg, $V > CV, VV > CVV$. In general, Mg produced more V and CV whereas

Bg attempted to produce different syllabic structures. Regarding the phonotactic development, Bg and Mg presented similar combinatory paths in some aspects, and different in others. Mg and Bg preferred vowels, bilabial stops and nasals combinations over their development. However, differently from Mg, Bg used dental segments in her initial phonological trajectory. In conclusion, we claim that there is, indeed, a mismatch in the phonological path of twins. Considering language as a complex adaptative system, and the fact that language development occurs based on stability and instability that affect the system itself and transform it constantly (THELEN; SMITH, 1994), this variability in the phonological path is expected, since each system is unique (BAIA, 2013).

High back vowel weakening in word boundaries

Amanda Fernandes Ivo, Matheus Freitas

This study is part of an undergraduate research project that aims at investigating the reduction of the highvowel [ɨ] between stop consonants at word boundaries in Brazilian Portuguese (BP) (e.g.: *muit[ɨ]populosos*). In a majority of languages, high vowels, such as [i] and [u], exhibit lower intrinsic duration than low vowels, such as [a], due to articulatory and aerodynamic factors. Some works (LEITE, 2006; SOUZA, 2012; ASSIS, 2017) indicate that high vowels are reduced in BP. Assis (2017) shows that, at word boundaries, the non-stressed high front vowel [ɨ] is reduced, especially when it is between voiceless consonants. Thus, this study intends to investigate the vowel reduction for the non-stressed high back vowel [ɨ] at word boundaries. This work objective is to assess the vowel reduction as a variable and gradient phenomenon. The theory adopted in this study is the Exemplar Theory (JOHNSON, 1997; JOHNSON; MULLENIX, 1997; PIERREHUMBERT, 2001, 2003). According to the Exemplar Theory, we consider that phonetic detail, in this case vowel duration, has impact on the phonological representation. The methodology adopted in this study is based on the Laboratory Phonology paradigm (BROWMAN; GOLDSTEIN, 1990; BYBEE; CHAKRABORTI; JUNG; SCHEIBMAN, 1998), since we use empirical data to corroborate our analysis. To study the reduction of [ɨ], we used the recordings of 14 speakers (7 male, 7 female). The analysed data were recorded in the following seven Brazilian states: Maranhão, Minas Gerais, São Paulo, Rio de Janeiro, Paraná, Santa Catarina, and Rio Grande do Sul. For the recordings, the speakers read the text *O Brasil em 2012* (project: *Sibilantes e róticos do Português Brasileiro*, process: 484590/2013-8, CNPq). The data are composed by 199 words with a non-stressed final high back vowel [ɨ] following a word that starts with a stop consonant. The data were analysed through a speech analysis software, Praat (BOERSMA; WEENINK, 2018). We determined the production of a [ɨ] vowel and its duration. Results indicate that vowel reduction is both a variable and gradient phenomenon. The vowel was completely reduced in 21.6% of the occurrences. The reduction is favoured in the state of São Paulo. Vowel duration is lower in vowels following a voiceless stop than

in vowels following a voiced stop. For future work, we intend to analyse vowel reduction of [] in more Brazilian regions and states.

Keywords: Vowel reduction; Brazilian Portuguese; Exemplar Theory; Laboratory Phonology.

REFERENCES ASSIS, A. M.A emergência de consoantes finais no português brasileiro na microrregião de Araguaína/Tocantins.2017.162f.Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal de MinasGerais, Faculdade de Letras, 2017.

BOERSMA, P.; WEENINK, D. Praat– 32-bit edition for Windows. Disponível em: <http://www.fon.hum.uva.nl/praat/> Accessed in: 05 de abril de 2018. BROWMAN, C.; GOLDSTEIN, L. Representation and reality: Physical systems and phonological structure. In: Journal of Phonetics 18, p.411-424. 1990.

BYBEE, J.; CHAIRABORTI, P.; JUNG D.; SCHEIBMAN, J. Prosody and segmental effect - some paths of evolution for word stress. In: Language Studies 22.2. p.267-314, 1998.

JOHNSON, K.; MULLENIX, J. W. Complex representations used in speech processing. In: JOHNSON, K.; MULLENIX, J. W. (Eds). Talker variability in speech processing. San Diego: Academic Press, 1997.

LEITE, C. T. Sequências de oclusiva alveolar + sibilante alveolar como um padrão inovador no português de Belo Horizonte. 2006. 126f. Dissertação (Mestrado em Estudos Linguísticos) – Faculdade de Letras, Universidade Federal de Minas Gerais, Faculdade de Letras, 2006.

PIERREHUMBERT, J. B. Exemplar dynamics: word frequency, lenition and contrast. In: BYBEE, J.; HOPPER, P. (Eds), Frequency effects and emergent grammar. Amsterdam: John Benjamins, 2001.

SOUZA, R. F. Redução de vogais altas pretônicas no português de Belo Horizonte: uma abordagem baseada na gradiência. 2012. 146f. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade Federal de Minas Gerais, Faculdade de Letras, 2012.

Poster Session 2

Brazilian Portuguese and English Rhotic Interphonology

Miriam Gurgel Da Silva, Clerton Luiz Felix Barboza

The objective of this research is to discuss the typology of Brazilian Portuguese (BP) and English rhotics concerning their specific place and manner of articulation, as well as possible pathways to interphonology construction. According

to Ladefoged & Maddieson (1996) rhotics are particularly associated to much acoustic and articulatory variation. Owing to their phonetic and phonological nature, both between and within languages, rhotics are described as a complex class to be investigated. This study is the first step in order to carry out an acoustic experimental research, aiming at better understanding of rhotic interphonology by Brazilian learners of English as a foreign language (EFL), in Northeastern Brazil. Although many researches have been devoted to rhotics interphonology, it is still an open area of investigation in literature, since no study involving Brazilian EFL learners from Rio Grande do Norte has ever been conducted. This study tries to answer the following research question: which are possible pathways for BP and English rhotics to construct the interphonology of Brazilian EFL learners? We hypothesize BP and English phonology connect in a network to build EFL interphonology. This study discusses results presented by by Ladefoged e Manddieson (1996), Brandão (2007), Vegini (2007), Bisol (2010), Gregio (2012), Carmargos (2013) and Rennie (2015). The poster is organized as follows: rhotic articulatory and acoustic characteristics; BP rhotics and R-variation; English rhotic and non-rhotic accents; BP and English rhotics interphonology; and conclusions. The poster emphasizes the need of a minute study of the rhotic systems associated with each language, besides hypothesizing various possibilities for Brazilian EFL learners constructing their interphonology.

Variability in ejective production in Quechua/Spanish bilinguals in Brazil

Carla Diaz; Eleonora Albano

Ejectives are aerodynamically complex stops occurring in only 20% of the world's languages. They occur in Quechua, but not in Spanish. In addition, their differences from the rest of the Quechua plosive inventory are still understudied. In a pilot acoustic-phonetic study conducted with Quechua-Spanish bilinguals in Brazil, we observed that: (1) the bilinguals produced the difference between the ejective and the aspirated stop with difficulty; (2) VOT was highly variable, and did not suffice to convey all the stop distinctions at stake. We sought to elucidate this phenomenon by analyzing the burst and aspiration durations, but found no clues. Variability was very high in both measures, pointing to a complex situation of language contact. The study measured the VOT of voiceless unaspirated, aspirated and ejective stops in five Quechua/Spanish bilinguals, as well as the VOT of voiceless and voiced stops in five Spanish-monolinguals. We also investigated how bilinguals produce these phonemes in L1 and L2; and evaluated our data against models of acquisition of L2 (Speech Learning Model and Perceptual Assimilation Model) that claim that phonemes that do not exist in L1 are easier to learn in L2 (Flege & Eefting, 1987; Best & Tyler, 2007). Sixty words were selected and included in carrier sentences. The Quechua corpus con-

sists of triads or minimal or analogous pairs presented in 36 sentences, including distractors. In the Spanish corpus, the words were distributed in 24 sentences, including distractors. The results raised the question of language attrition due to Spanish dominance in the bilinguals. The statistics partly confirmed the theoretical predictions about the greater ease of producing L2 (Spanish) phonemes that do not exist in L1 (Quechua): the bilinguals easily produced the pre-voiced stops of Spanish that do not exist in Quechua (Stewart, 2015). However, the bilinguals' difficulty in distinguishing the ejective from the aspirated stops in L1 cannot be explained by the L2 acquisition models. Aspirated and ejective stops have similar variation coefficients. The statistics ($F = 17.84$, $p = .013$) revealed that the bilinguals only differentiate VOT between the following pairs: voiceless vs. voiced stops in Spanish, and Spanish voiceless stops vs. Quechua aspirated stops. Thus, VOT does not seem to be the only acoustic cue to distinguish these consonants, since the ejectives require several articulatory adjustments (Lindau, 1984). Possibly, other differences were masked by a pause participants tended to introduce in the carrier sentence. In a post-hoc analysis of the interviews conducted for participant selection, we found that the bilinguals unanimously declared that they speak Spanish very fluently in all situations, but the same does not apply to Quechua. There is no doubt that language contact was intense and that Spanish exerts a great deal of pressure. The high variability of the data and the difficulty of the bilinguals to distinguish the aspirated from the ejective stops suggest that Spanish is affecting their glottal settings for ejective production. If this is confirmed by data yet to be collected, the concepts of language attrition (Ribes & Llanes, 2015) and diglossia (Wölck, 2008) may become central to the next steps of this research.

Early prosodic development or early musical development? Findings from Psycholinguistics and Music Psychology

Maria de Fátima de Almeida Baia

In this study, I present a literature review on the (dis)encounters between studies of phonological development (prosodic) and studies of musical perception / learning. The aim of this ongoing study is to establish a dialogue between the two areas through a joint analysis of the prosodic development literature and musical development. For instance, prosodic development studies point out to universals in early intonational patterns (SCARPA, 1997) as the alternation between high and low tones in intonation, short melodies, binary rhythmic sequence of (un)stressed syllables, etc, whereas studies on perception and musical production of babies find specific musical intervals of each stage of development (PARIZZI, 2006). However, psycholinguistics studies, in general, observe linguistic aspects in early production/perception (SCARPA, 1997; DEMUTH, 1996; JUCSZYK, 1997) whereas studies on early musical development claim there would be “pure” musical aspects only at the beginning of development

as language aspects would appear by the first year. The findings will be relevant for the practice of the music therapist who deals with both linguistic and musical aspects.