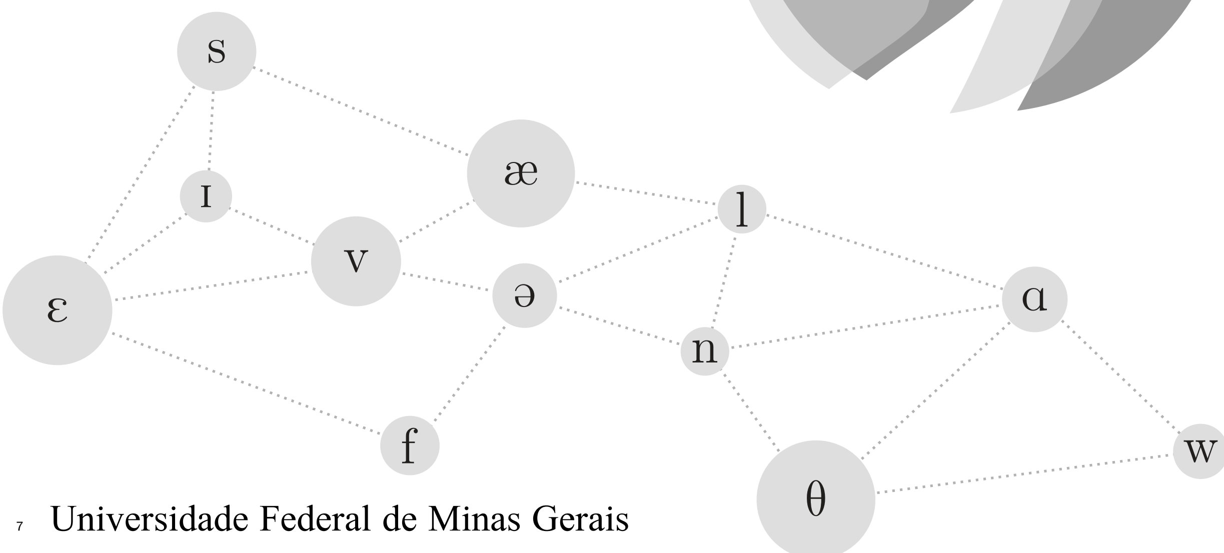
Thaïs Cristófato-Silva, Hani Yehia, Leonardo Araujo,
Maria Cantoni, Magnum Madruga and Adriano Vilela

EICEFALA 2021

International Meeting on Speech Sciences

Advances in speech and L2 processing

SEVENTH EDITION



Universidade Federal de Minas Gerais

PUBLISHED BY UNIVERSIDADE FEDERAL DE MINAS GERAIS

First printing, November 2021

Contents

Introduction

EICEFALA is an event promoted by UFMG laboratories from Faculdade de Letras (Laboratório de Fonologia) and Escola de Engenharia (CEFALA).

The comprehension of human communication involving speech acoustics and gestures requires knowledge from various fields of science, such as phonetics, phonology, linguistics, acoustics, mechanics, mathematics, physiology, neuroscience and computer science. The objective of the 7th EICEFALA is to present and discuss theoretical and methodological techniques to researchers from the several areas of knowledge working with speech science: linguists, engineers, physicists, speech therapists, musicians, etc. It is expected that the participants of the event find, at EICEFALA, a transdisciplinary forum to address questions related to spoken communication.

Part I:

Plenary Speakers

> Fairy tales are more than true: not because they tell us that dragons exist, but because they tell us dragons can be beaten.
>
> C.K. Chesterton

Using statistical learning techniques to determine Cantonese lexical tones from the acoustic and visual components of speech

João Vitor Possamai de Menezes[1], Hani Camille Yehia[1], Adriano Vilela Barbosa[1]
1 . Universidade Federal de Minas Gerais

This mini-course presents an introduction to the use of statistical learning techniques to speech processing problems. More specifically, we show how classification techniques can be used to predict lexical tones in Cantonese from the associated measurements of both the acoustic and the visual (to a lesser degree) components of speech. The acoustic and visual data we use were recorded during a speech production experiment where a native speaker of Cantonese produced a set of words spanning the full range of Cantonese tones. The visual data consists of 3D trajectories of markers on the subject's face and head recorded with an Optotrak. The acoustic component is represented by F0 trajectories extracted from the speech acoustics. The idea is to use the F0 and marker trajectories as input vectors to train classifiers to predict the lexical tones. However, these trajectories cannot be used directly because they have different durations for different tokens (utterances), whereas all input vectors to the classifiers must have the same dimension. In order to make all input vectors the same length, regardless of the duration of the utterances, all trajectories (both F0 and markers) are approximated by polynomials of a given order and represented by the corresponding coefficients. The polynomial coefficients are then used as input vectors to train different classification models (LDA, SVM, K-nearest neighbors, etc). The performance of the models is estimated

by means of k-fold cross validation. Although the statistical learning techniques we present are applied to a specific problem (estimating Cantonese lexical tones from the acoustical and visual components of speech), they are general and can be equally applied to a wide range of problems. All procedures presented in the mini-course are developed in the R language.

The multiple dimensions of speech: old questions and new challenges

DIDIER DEMOLIN[1]

1 . Laboratoire de Phonétique et Phonologie
CNRS-UMR 7018

Human speech, a product of the evolution of primates, can in essence be defined in terms of a signal. This is an acoustic wave varying over time with amplitude and frequency modulations, due to the articulatory movements of the vocal tract's organs. To perform these movements, motor controls are required, whose interactions with the aerodynamic parameters produce the acoustic signal. The main objective of research in this domain is to understand which primary principles,biological, physical and cognitive, to be based on to explain the production and perception of speech in the world's languages and to make the fundamental question: how does it work?

Among the main fields of activity involved in the study of sounds and sound systems of languages are the engineering sciences with the dimensions of automatic processing (speech recognition and synthesis);phonetics and phonology (the linguistic aspects); and pathological aspects(how to explain what doesn't work anymore or less well). This includes knowledge of similar fundamental principles. To these dimensions a readded physics, biology, cognition and neuroscience. These fields involves in-depth knowledge of various interconnected fields to explain how sounds and sound systems work. Therefore in addition to the symbolic dimension, anatomical, physiological, acoustic, aerodynamic, articulatory,auditory, proprioceptive, historical (phylogenies and diachrony),ecological, temporal, dynamic and self-organized aspects can, and should,be integrated in the explanation of the studied phenomena.

The complexity and interactions of these dimensions find new light in the paradigms resulting from the study of complex systems, which makes it possible to address old issues again, such as the search for a possible speech code, invariants and primitives. From these issues, others arise, such as the understanding of the open or closed nature of sound systems, which is far from being resolved. Explaining the diversity,complexity and dynamics of sound systems involves understanding the nature of variation in speech phenomena. How can we show that spontaneous speech, laboratory speech and pathological aspects are based on the same principles?

The evolution of theory, models, new statistical tools,computational, big data and deep learning tools, allow these issues to be addressed in a new light. New measuring instruments such as real-time magnetic resonance imaging, functional magnetic resonance, three-dimensional or four-dimensional ultrasound, digital endoscopy,electroencephalography (EEG) and many other recent tools make it possible to accurately observe, measure and quantify speech phenomena as well as bring to discussion fundamental issues still unresolved or poorly understood.

The lecture will discuss the controlled and automatic aspects involved in the control of breathing in speech, issues in speech embodiment, the quantal aspects of speech, the importance of thresholds values in aerodynamic and acoustic parameters, types of feedback (acoustic and proprioceptive) in speech phenomena and new ways to explain and formalize the source, the initiation and propagation of sound changes. This last point by using and adapting population ecology models to speech.

Some questions on L2 speech as related to colonialism

Eleonora Albano[1], Antonio Pessotti[1], Carla Diaz[1]

1 . LAFAPE-IEL-UNICAMP & CNPq

The first aim of this talk is to revisit the question of phonetic drift in L_2 speechin light of new data and theory. The new data consist of a sizeable set of acousticphonetic measures of the speech of Quechua and Spanish monolinguals and bilingualsresiding in Peru. The theoretical innovation draws on two sources: the relativelyfamiliar concept of accommodation, introduced by Giles et al. insociolinguistics, and the less familiar concept of coloniality, introduced by Quijano(2000) in sociology. At the same time, it aims at showing that phonetic analysis basedon gestural phonology can open new avenues for exploring the relationship betweenthese two concepts as explanans for L_2 pronunciation in an ethnically diverseenvironment.

Accommodation refers to a “constant movement toward and away from others,by changing one’s communicative behavior” (GILES & OGAY, 2007). It encompassesspeech and various other communicative behaviors. Moreover, it has a convergent side – enhancing similarities between interlocutors – and a divergent one – enhancingdifferences between interlocutors. Both can occur between two or more people or withinand across speech communities. Some acoustic phonetic parameters have been useful totap such shifts (e.g., VOT, as in the pioneering work of SANCIER & FOWLER, 1997).

The concept of coloniality refers to “how colonial patterns of power andinequality exceed the spatial and temporal boundaries of empire and colony” (ROCHE,2019). It aims at dealing with the epistemology underlying the pervasive replication ofcolonial social, economic, and cultural practices in postcolonial societies (QUIJANO,2000).

We will start by revisiting earlier work on phonetic drift in L_2 conducted at ourlab – Laboratório de Fonética e Psicolinguística (LAFAPE). Ramirez et al. (2011)showed that contact situations may exhibit intralinguistic phonetic drift in both L_1 and L_2. In turn, Albano et al. (2020) reported preliminary observations of intralinguistic driftattributable to language attrition in Quechua/Spanish bilinguals residing in Brazil.

We believe that the understanding of the results of both of these works can beconsiderably improved by reference to the above-defined concepts. In particular, someintriguing signs of partial loss of Quechua stop distinctions shown by the expatriatedPeruvians can be interpreted as mistiming of articulatory gestures converging towardsthose of the two hegemonic languages (namely, Spanish and Portuguese).

Then we will move on to inquire how the study conducted in Peru can elucidateour questions about Quechua/Spanish relations. All data collection on this topic waspart of Carla Diaz's requisites for completing her bachelor and master's degrees inlinguistics (DIAZ, 2018; 2021).

Carla recorded 10 Spanish monolinguals and 10 Quechua/Spanish bilinguals inLima in August 2019. Then she travelled to Cuzco to record 11 monolingual Quechuaspeakers, with the help of a bilingual friend specializing in Quechuan literature.

The corpus, similar to that of Albano et al. (2020), focused on Quechua andSpanish stop contrasts. The analysis, likewise, employed measurements that have beenused in the description of Quechua: VOT, amplitude of the stop burst, f_0, and $H1 - H2$.

The results show that, unlike the residents of Brazil, the residents of Lima haveno trouble distinguishing stops within the Quechua series or differentiating them fromSpanish stops. The remarkable fact is that divergence from Spanish was more frequentthan convergence. Moreover, certain distinctions were enhanced by shifting the acousticparameters beyond the values of the monolingual group.

After Carla defends her thesis, we are planning to conduct finer-grained analysesconsidering the linguistic values and attitudes captured by our sociolinguisticquestionnaire. May we succeed in helping unravel the coloniality issues behind thesubtle attempts of Peruvian Quechua speakers at resisting diglossia and language loss.

Bibliography

Howard Giles, Donald M. Taylor, and Richard Bourhis. Towards a theory of interpersonal accommodation through language: some canadian data. Language in Society, 2(2):177–192, October 1973. DOI: 10.1017/s0047404500000701. URL https://doi.org/10.1017/s0047404500000701.

How to model the influence of orthography on L2 representationswith BiPhon Neural Networks

SILKE HAMANN[1], CHAO ZHOU[1]
1 . University of Amsterdam and University of Lisbon

Many studies have shown that written forms influence the acquisition of a second language. This influence can be helpful, as is the case of the English /æ/-/ɛ/ contrast that is notoriously difficult for Dutch learners but where the written form can aid in the creation of the distinction [?Escudero and Wanrooij, 2010]. But orthography can also cause the creation of so-called ghost contrasts, which do not exist in the L2, as is the case with the intervocalic singleton/geminate contrast in the L2 English of Italian speakers [Bassetti, 2017, Hamann, 2018].

In this talk, we illustrate how such orthographic influences on the creation of L2 representations can be formalized, by this yielding theoretical predictions that can be tested again in experimental studies. Our formalization is performed with a symbolic neural network based on the Bidirectional Phonetics-Phonology model [Boersma, 2007] and its extension by a reading grammar [Hamann and Colombo, 2017].

Our main data comes from an experimental study on Mandarin [Zhou and Hamann, 2020]: 23 L1-Mandarin speakers with no prior knowledge of EP (naïve listeners), representing the initial stage of L2 acquisition, performed a delayed-imitation task. They were presented with EP nonce words containing /ɾ/ in intervocalic onset (e.g., parafa) or word-internal coda (e.g., parfa), first auditorily, and then with accompanying orthography. Our results show 1) that participants only produced L1 [ɻ] when exposed to orthography, confirming that the use of Mandarin rhotic in L2 speech is orthographically driven; and 2) that even at the initial stage the substitution with Mandarin [ɻ] occurs almost exclusively in coda position, reminiscent of L2 learners [Zhou, 2017, Liu, 2018].

Bibliography

Bene Bassetti. Orthography affects second language speech: Double letters and geminate production in English. Journal of Experimental Psychology: Learning, Memory, and Cognition, 43(11): 1835–1842, November 2017. ISSN 1939-1285, 0278-7393. DOI: 10.1037/xlm0000417. URL http://doi.apa.org/getdoi.cfm?doi=10.1037/xlm0000417.

Paul Boersma. Some listener-oriented accounts of h-aspiré in French. Lingua, 117(12):1989–2054, December 2007. ISSN 00243841. DOI: 10.1016/j.lingua.2006.11.004. URL https://linkinghub.elsevier.com/retrieve/pii/S0024384106002191.

Paola Escudero and Karin Wanrooij. The Effect of L1 Orthography on Non-native Vowel Perception. Language and Speech, 53(3): 343–365, September 2010. ISSN 0023-8309, 1756-6053. DOI: 10.1177/0023830910371447. URL http://journals.sagepub.com/doi/10.1177/0023830910371447.

Silke Hamann. Ghost phonemes in second languages – How orthography can create contrasts without perceptual correlates, 2018.

Silke Hamann and Ilaria E. Colombo. A formal account of the interaction of orthography and perception: English intervocalic consonants borrowed into Italian. Natural Language & Linguistic Theory, 35(3):683–714, August 2017. ISSN 0167-806X, 1573-0859. DOI: 10.1007/s11049-017-9362-3. URL http://link.springer.com/10.1007/s11049-017-9362-3.

Wen Liu. Aquisição da Vibrante Simples [r] pelos Alunos Chineses Aprendentes de Português como Língua Estrangeira, 2018.

Chao Zhou. Contributo para o estudo da aquisição das consoantes líquidas do português europeu por aprendentes chineses, 2017.

Chao Zhou and Silke Hamann. Cross-Linguistic Interaction Between Phonological Categorization and Orthography Predicts Prosodic Effects in the Acquisition of Portuguese Liquids by L1-Mandarin Learners. In Interspeech 2020, pages 4486–4490. ISCA, October 2020. DOI: 10.21437/Interspeech.2020-2689. URL https://

www.isca-speech.org/archive/interspeech_2020/zhou20h_interspeech.html.

Part II:

Advances in speech and L2 processing

Fairy tales are more than true: not because they tell us that dragons exist, but because they tell us dragons can be beaten.

C.K. Chesterton

An Analysis of the Development of the Rhythm of English-L2 by Brazilian Learners through Rhythmic Metrics and Acoustic Parameters

LEONARDO ANTONIO SILVA TEIXEIRA[1], RONALDO MANGUEIRA LIMA JR.[1]

1 . Universidade Federal do Ceará

Abstract

The aim of this study is to describe and discuss the development of L2 English rhythm by Brazilian learners through rhythmic metrics and prosodic-acoustic parameters that characterize the oral production of these learners at different stages of L2 development. Five Brazilian learners of English-L2 were recorded reading a text in English at the beginning of their college studies in English Language Teaching, and again four semesters later, after having taken two English phonology courses. They were also recorded reading a version of the text translated into Portuguese. Besides the learners, five native speakers of North American English were recorded reading the same text in English. Data were manually segmented into vowel units (V), consonant (C), vowel-vowel (VV), sentences (S) and higher prosodic units - chunks (CH) in PRAAT [Boersma and Weenink, 2019], and the parameters were automatically by means of a script. Data were statistically treated via R [?]through the implementation of mixed-effects regression models. Results placed Brazilian Portuguese and English-L1 in different rhythmic spaces, as predicted by the literature; in the durational dimension, the metrics positioned the English-L2 of the first recording far from both English-L1 and Brazilian Portuguese; in the f0 and intensity dimensions, however, the acoustic parameters placed the English-L2 of the first

recording closer to Brazilian Portuguese. In both dimensions, the English-L2 of the subsequent recording was closer to English-L1, suggesting a developmental route towards the target language. The results also suggest positive effects of the explicit teaching of pronunciation.

Introduction

Regarding research in non-native language (L2) development, there seems to be greater emphasis on segmental aspects rather than prosodic ones [Li and Post, 2014, Thomson and Derwing, 2015]. This tendency is also reflected in L2 acquisition models [Flege et al., 2021, Best and Tyler, 2007], which emphasize segmental aspects, providing little support to the understanding of L2 prosodic development. There is also evidence that rhythm can influence the communication process in a global way, affecting degrees of perceived foreign accent and intelligibility [Silva Junior and Barbosa, 2020].

The scarcity of studies on the acquisition of L2 rhythm may be related to the difficulty of establishing the physical reality of such construct. There are at least three trends on research regarding linguistic rhythm. Lloyd James (1940), as cited in Abercrombie [1971], relied on the dichotomy Morse code versus machine gun to illustrate the perceptual difference between English and Spanish, respectively. Pike [1945] formalized that difference by proposing a rhythmic approach based on the type of units that stood up in such languages: stress-timed languages, for which interstress intervals would be the most prominent units, and syllable-timed languages, for which syllables would be such units. Later, Abercrombie [1971] proposed that those rhythmic units would be isochronous, that is, of the same duration. However, the isochrony paradigm proved to be empirically unsustainable since intervals of the same duration are not found in the acoustic signal [Cumming, 2010].

From the mid-90s, a second trend of studies in linguistic rhythm emerges, in which the rhythmic patterns are investigated by means of the durational characteristics of the reference intervals (vowels, consonants, syllables, etc.), which can be computed by statistical indexes called rhythmic metrics. Ramus et al. [1999] proposed the standard deviation of the duration of consonantal intervals (ΔC) and the percentual of the total duration of the utterance composed of vowel intervals (%V). Those metrics were able to spatially discriminate languages considered syllable-timed (French, Spanish, Italian and Catalan), stress-timed (English, Polish and Dutch) and mora-timed languages

(Japanese) [Ladefoged, 1975] on a plane with ΔC and %V on each axis, as can be seen in Figure 1, reproduced from the original paper:

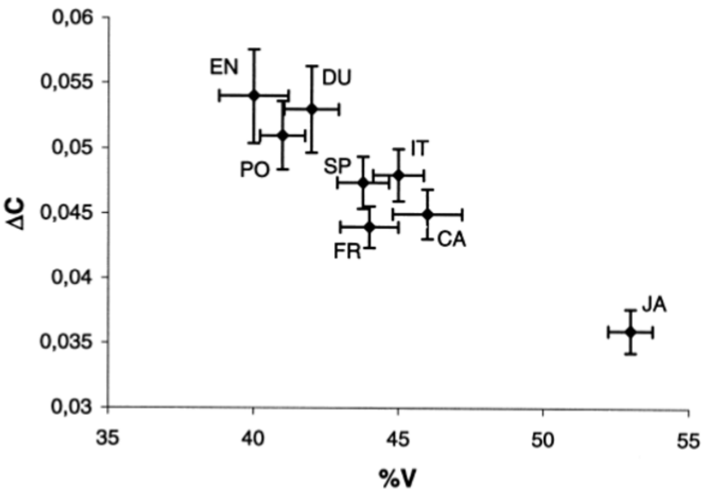


Figure 1: Distribution of languages over the (%V, ΔC) plane. Error bars represent 1 standard error. Source: Ramus et al. [1999, p. 273].

The present study follows a third trend on research on linguistic rhythm, which define it as a function of the distribution of prominent elements in the acoustic signal, which involves several acoustic dimensions – duration, fundamental frequency (f0) and intensity, and may be influenced by the native language of the speaker [Cumming, 2010, Fuchs, 2016, Silva Junior and Barbosa, 2020]. Thus, this study was guided by the following questions: (i) how do the metrics and acoustic parameters place North American English-L1, Brazilian English-L2, and the Brazilian Portuguese (BP)-L1 in the rhythmic space? (ii) What is the influence of the rhythm of BP-L1 on the development of English-L2 of learners? (iii) What is the effect of explicit pronunciation teaching on learners' English-L2 rhythm development? The following hypotheses were raised: (i) PB-L1, English-L1 and English-L2 are rhythmically different systems; (ii) there will be rhythmic differences between the English-L2 of the speakers in the two different stages of development whose recordings were analyzed; (iii) the English-L2 of the first recording should be more dissimilar to English-L1 due to L1 transfer and lack of explicit instruction.

Methods

As for the participants, the experimental group was composed of five BP-L1 speakers, who were also learners of English-L2. They were all college students of English Language Teaching, being four men and one woman, aged between 18 to 24. The control group comprised five English-L1 speakers, all Canadians, being one man and four women, aged 23-34. Four corpora of oral production were analyzed in this study: English-L1, PB-L1, English-L2 (1) and English-L2 (4). The data of English-L2 were obtained by means of recordings of the Brazilian learners reading the first paragraph of a text in two different moments, before and after completing courses in English Phonetics and Phonology. Those were the first and fourth recording made so they are referred to as English-L2 (1) and English-L2 (4). The data of English-L1 resulted from the reading of the same text by the control group. Finally, the Portuguese-L1 data came from the reading of the Portuguese version of the text by the Brazilian learners. The recordings took place in a silent room with a cardioid Shure MX150B lapel microphone connected to a Zoom 4HnSP recorder. The audio was captured in mono, with a sampling rate of 44.1 kHz, and saved in wav format.

Data were manually segmented into vowel units (V), consonant (C), vowel-vowel (VV), that is, the interval between the acoustic onset of a vowel and the onset of the adjacent one, sentences (S) and higher prosodic units - chunks (CH) in PRAAT [?], and the script Metrics & Acoustics Extractor [Silva Junior and Barbosa, 2020] was used to extract the parameters. Following Silva Junior and Barbosa [2020], the term metric(s) is used in this research to refer to the duration-based parameters, and the term acoustic parameter(s) refers to the f0, speech rate and intensive-related ones. The table below presents a summary of the metrics and acoustic parameters analyzed in this study and the types of segments they compute:

Data were then statistically treated via R [?]through the implementation of mixed-effects regression models, adopting language and semester as predictor variables, and rhythmic metrics and acoustic parameters as response variables.

Results

In this section, we present some of the significant results, based on the mixed-effects regression models adjusted for each metric and acoustic parameter, and the boxplots and bidimensional planes, in which the effect of each corpus (independent variables) on the significant metrics and acoustic measures (the dependent variables) can be visually inspected.

[1] See Fuchs [2016] for a comprehensive account of rhythmic metrics.

[2] Standard deviation of the segment duration divided by the mean, multiplied by 100.

[3] Mean of the differences between successive segments.

[4] Mean of the differences between successive segments divided by their sum, multiplied by 100.

[5] Mean of pairwise quotients of adjacent segment durations, where the duration of the shorter is divided by the duration of the longer one and multiplied by 100.

[6] Mean of the differences between successive segments where the duration of each segment is normalised through division by the mean of all segments' durations.

[7] Mean of the differences between successive segments where the durations are normalised by z-transformation.

METRICS[1]		ACOUSTIC PAREMETERS	
Parameters	Segment of application	Parameters	Segment of application
Percentual (%)	V, C	f0 median	S, CH
Standard-deviation (Δ)	V,C (V ou C), VV	f0 peak	S, CH
Variation coefficient (Varco)[2]	V,C (V ou C), VV	f0 minimum	S, CH
Raw pairwise variability index (r-PVI)[3]	V,C (V ou C), VV	f0 standard deviation	S, CH
Normalized pairwise variability index (n-PVI)[4]	V,C (V ou C), VV	f0 skewness	S, CH
Rhythm ratio (RR)[5]	V,C (V ou C), VV	Mean of f0 first derivative (μΔ1-f0)	S, CH
Variability index (VI)[6]	V,C (V ou C), VV	Standard deviation of f0 first derivative (σΔ1- f0)	S, CH
		Skewness of f0 first derivative (skΔ1- f0)	
Yet another rhythm determination (z-score duration) (YARD)[7]	V,C (V ou C), VV	Speech rate (SR)	VV, S, CH
		f0 rate (f0-R)	S, CH
		Spectral emphasis	S, CH
		Mean of normalized syllable-peak duration (μdur- Sil)	VV, S, CH
		Mean duration of pauses (μdur-#)	S, CH

Table 1: Rhythm metrics and prosodic-acoustic parameters analyzed in this study

Metrics

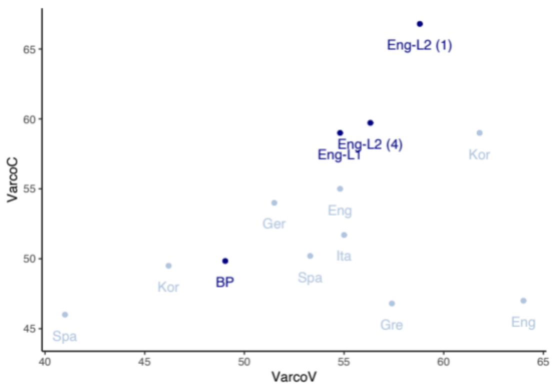
Twenty out of the thirty employed metrics reached statistical significance for at least two of the (inter) languages. One example of the mixed-regression models that were implemented via R can be seen in table 2, which were adjusted for the standard deviation of the duration of consonantal intervals (deltaC) and the percentual of vocalic intervals (%V).

	ΔC	%V				
Predictors	Estimates	CI	p	Estimates	CI	p
(Intercept)	46.48	36.17 – 56.79	< 0.001	48.78	46.02 – 51.55	< 0.001
Lang [Eng-L1]	21.94	7.35 – 36.52	0.004	-9.66	-12.97– -6.35	< 0.001
Lang [Eng-L2 (1)]	58.71	44.13 – 73.30	< 0.001	-11.92	-14.23 – -9.61	< 0.001
Lang [Eng-L2 (4)]	37.61	23.02 – 52.19	< 0.001	-9.28	-11.47 – -7.09	< 0.001

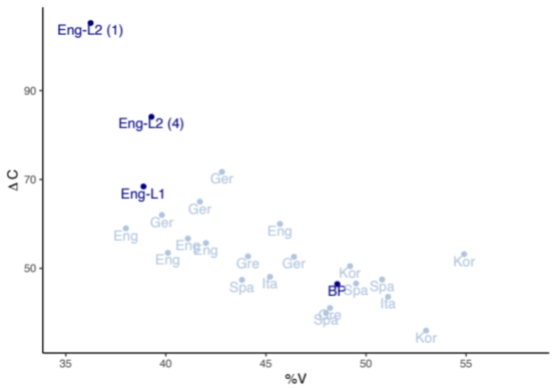
Table 2: Coefficients, confidence intervals (95%) and p-Values for the two linear mixed-effect regression models adjusted for ΔC and %V. models: deltaC Lang + (1|Chunk) + (1|Speaker) and percV Lang + (1|Chunk) + (1 / Speaker).

Figure 2 shows the distribution of the 4 corpora over the planes formed by the pairs ΔC-%V and VarcoC-VarcoV in comparison to the data reviewed and obtained by Arvaniti (2012).

As can be seen in figure 2 (a), English-L1 presented greater standard deviation of consonantal intervals duration (ΔCEng-L1 = 68.41) compared to BP (ΔCBP = 46.48); and BP presented greater proportion of the utterance composed of vowel intervals (%VBP = 48.56) compared to English-L1 (%VEng-L1 = 38,88). The data of English-L2 (1) were positioned far from the two native languages, scoring ΔC values quite high (ΔCEng-L2 (1) = 105.19) and the lowest proportion of vowel segments (%VEnglish-L2 (1) = 36.24). On the other hand, English-L2 (4) values were much closer to English-L1 in relation to both axis (ΔCEngl-L2(4) = 84.08; %VEng-L2(4) = 39.28). As for VarcoC-VarcoV (Figure 2), English-L1, English-L2 (1), English-L2 (4) and BP were distributed analogously to the plane ΔC-%V, with the BP data recording the lowest values for both the VarcoC (49.84) and VarcoV (49.04) axis, and English-L2 (1) presenting the highest scores both in the VarcoC axis (66.8) and in relation to the VarcoV axis (58.8). The fact that English-L2(1) assumed values far from the L1 (BP) indicates no objective transference of durational prosodic patterns to the learners' interlanguages. On the other hand, the approximation between English-L2(4) and English-L1 indicates a possible effect of explicit instruction, among other factors, that may have influenced the temporal (re)organization of the learners' speech towards the prosodic patterns of the target language.



(a)



(b)

Figure 2: Present study data (dark blue) amid all the data reviewed and obtained by Arvaniti (2012) (light blue) for ΔC - %V (Figure 2 (a) and VarcoC-VarcoV (Figure 2 (b)), in which Eng = English, Ger = German, Gre = Greek, Spa = Spanish, UI = Italian, Kor = Korean. Source: Teixeira and Lima Jr. (2021).

Regarding the data from Arvaniti (2012), BP grouped with languages considered more syllable-timed, that is, with more durational regularity among the segments of reference, such as Spanish and Italian. English-L1 results were also consistent with the literature, gathering with the results for English and German from other studies, which are considered languages with more stress-timing tendency.

The hierarchy of values for VarcoV-VarcoC and ΔC-%V illustrates the dominant positioning pattern for the significant metrics, as can be seen in table 3: ([+stress-timed] English-L2 (1) > English-L2 (4) > English-L1 > BP [+ syllable-timed]), except for %V and RR, whose higher values indicate a tendency towards syllable-timing.

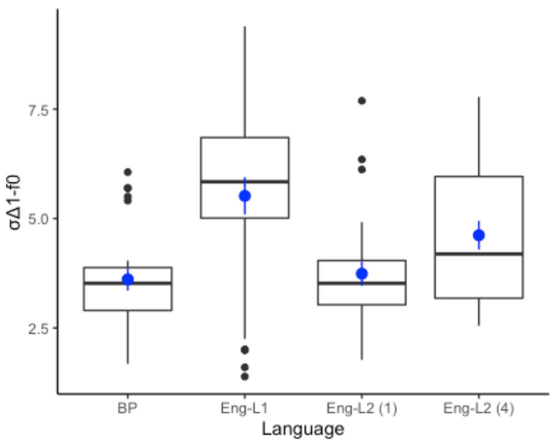
Metric	BP	English-L1	English-L2(1)	English-L2(4)
%V	48.56 (3.16)	38.88 (4.96)	36.24 (5.36)	46.48(8.02)
%C	51.44 (3,16)	61.12 (4.96)	63.76 (5.36)	68.416(14.55)
ΔV	40.08 (10.81)	41.16 (11.82)	51.81 (12.79)	105.192(32.6)
ΔC	46.48 (8.02)	68.41 (14.55)	105.192 (32.6)	84.088(36.51)
ΔS*	133.4 (45.27)	198.53 (77.44)	217.46(97.65)	184.75 (56.72)
VarcoV	49.04 (8.49)	54.80 (12.52)	58.80 (11.30)	56.32(14.81)
VarcoC	49.84 (7,00)	59 (11.89)	66.80 (13.69)	59.72(22.02)
rPVI-V	65.1 (11.71)	70.74 (14.84)	96.98 (19.27)	81.21(15.75)
rPVI-C	48.22 (8.91)	86.16 (20.23)	116.84 (46.64)	88.7(21.69)
rPVI-VC	64.73 (18.72)	83.58 (11.12)	114.4 (38.18)	89.1(14.53)
rPVI-S	102.85 (40.24)	130.66 (33.59)	176.27 (90.03)	137.96(44.62)
nPVI-C	53.96 (7.21)	68.56 (11.72)	72.36 (12.14)	64.84(11.46)
nPVI-VC	59.76 (7.69)	68.96 (11.09)	72.6 (9.40)	65.08(8.12)
RR-C	61.17 (4.36)	53.13 (6.29)	50.97 (6.59)	54.59(6.42)
RR-VC	58.07 (4.35)	52.8 (5.65)	50.91 (4.97)	54.55(4.59)
VI-V	0.818 (0.166)	0.981 (0.322)	1.128 (0.302)	0.894(0.188)
VI-V	0.830 (0.037)	0.924 (0.049)	0.929 (0.052)	0.934(0.059)
VI-VC	0.684 (0.101)	0.834 (0.157)	0.859 (0.120)	0.746(0.116)
VI-S	0.516 (0.120)	0.606 (0.136)	0.615 (0.160)	0.538(0.126)
YARD-VC	0.717 (0.150)	0.695 (0.133)	0.869 (0.113)	0.848(0.123)

Table 3: Absolute means for the statistically significant metrics and standard deviation (between parentheses) for BP, English-L1, English-L2 (1), English-L2(1) and English-L2(4).
* S stands for the phonetic syllable, which is the vowel-vowel (VV) unit.

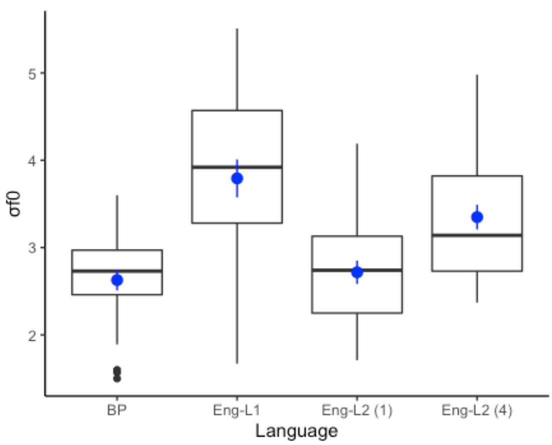
Acoustic Parameters

Five out of the twelve employed acoustic parameters reached statistical significance: f0peak, σf0, σΔ1- f0, spectral emphasis (emph) and speech rate (SR).

As for the standard deviation of f0 (Figure 3.1), English-L1 presented the highest standard deviation among the corpora analyzed ($\sigma f0_{Engl\text{-}L1} = 3.79$), followed by English-L2(4) ($\sigma f0_{Engl\text{-}L2(4)} = 3.34$), English-L2(1) ($\sigma f0_{Engl\text{-}L2(4)} = 2.71$) and BP ($\sigma f0_{BP} = 2.62$). The results for this parameter suggest a gradual prosodic development of the learners towards the f0 variation patterns of the target language. The standard deviation of f0 first derivative ($\sigma\Delta 1\text{-}f0$) (Figure 3.2) was also successful in the separation of the L1s and captured a similar course of development to that found by $\sigma f0$. The highest mean was scored by English-L1($\sigma\Delta 1\text{-}f0_{Eng\text{-}L1} = 5.51$), the lowest mean was scored by BP ($\sigma\Delta 1\text{-}f0_{PB} = 3.61$). The interlanguages registered intermediate values, but the mean English-L2(4) was much closer to English-L1 ($\sigma\Delta 1\text{-}f0_{Eng\text{-}L2(1)} = 3.73 < \sigma\Delta 1\text{-}f0_{Eng\text{-}L2(4)} = 4.61$).



(a)



(b)

Figure 3: Boxplots of the means $\sigma f0$ (Figure 3.1) and $\sigma\Delta 1\text{-} f0$ (Figure 3.2 for English-L1, English-L2(1), English-L2(4) and BP. The blue dots and lines represent the means and standard errors respectively.

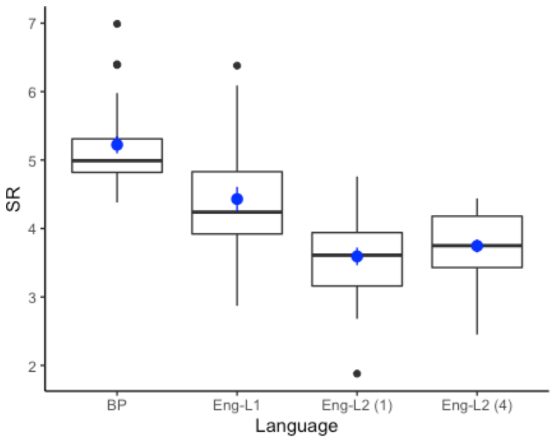
The results for the f0 dimension must be interpreted with caution, since there was an unbalance between male and female participants in both groups (control group: 1 male, 4 female; experimental group: 4 males, 1 female). In fact, the correlation between f0 and sex is evident when individual results are taken into consideration. For instance, in the experimental group, it was observed that participant N, the only female, is the one that had the highest f0 peak (97.16), as well as the widest scopes of f0 (f0 peak minus f0 min) for BP (17.18) and English-L2(4) (19.06). There was also a smaller variation between the male learners f0 scope of English-L2(1) (A = 12.28; F = 15.68; K = 15.5; L 13.37) and English-L2(4) (A = 12.81; F = 15.09; K = 14.43; L = 15.1), in comparison to the variation of the female participant, who went from 15.59 to 19.06 in the last recording.

In the dimension of intensity, as visually demonstrated in Figure 4.1, spectral emphasis was able to separate the L1s, with the highest mean for English-L1 among the analyzed corpora (emphEng-L1 = 4.34), which was higher than PB (emphPB = 2.73). If we consider works that show the correlation between spectral emphasis and phrasal stress [?], this result suggests that native English speakers make more effort as an acoustic clue in stress marking than Portuguese speakers. Regarding the interlanguages, English-L2 (1) obtained the lowest mean of spectral emphasis, very close to BP values, (emphEng-L2(1) = 2.56), and English-L2 (4) got much closer to English-L1(emphEngl-L2 (4) = 3.23). This indicates L1 transfer at the intensity dimension, and a tendency towards the prosodic patterns of English-L1 in the last recording.

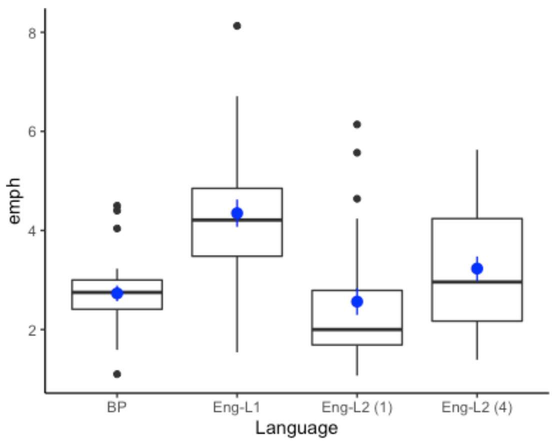
As expected, the L1s presented higher speech rates, with BP registering a higher mean compared to English-L1 (SRPB = 5.22 > SREngl-L1 = 4.43). In addition, English-L2 (1) presented the lowest speech rate among the corpora analyzed (SREngl-L2(1) = 3.59) and English-L2(4) registered a slightly higher mean, closer to English-L1 (SREng-L2(4) = 3.74). The increase in the speech rate of the interlanguages between the first and last recording may be related to the effects of explicit instruction.

Discussion

The metrics and parameters positioned BP, English-L1, English-L2 (1) and English-L2 (4) as rhythmically different systems. There were differences between the English-L2 of the speakers in the two different stages of development and different developmental paths were captured as function of the dimension of prominence. This developmental path



(a)



(b)

Figure 4: Boxplots of the means spectral emphasis (Figure 4 (a)) and speech rate (Figure 4 (b)) for English-L1, English-L2(1), English-L2(4) and BP. The blue dots and lines represent the means and standard errors respectively.

is consistent with the definition of interlanguage that presents itself as a relatively independent system of L1 and L2 [Li and Post, 2014], and with the non-linearity of the L2 development process [Lima Jr and Alves, 2019].

At the durational level, the dominant distribution pattern ([+ stress-timed] English-L2 (1) > English-L2 (4) > English-L1 > BP [+ syllable-timed]), with English-L2(1) assuming the highest means among the four corpora, and English-L2(4) getting closer values to English-L1. One possible explanation for such behavior for English-L2(1) is that learners may have mobilized a process of dissimilation of phonetic categories, displaying exaggerated durational values to maintain the distinction between L1 and L2, similarly to what is predicted by the Speech Learning Model for the segmental level [?Flege et al., 2021].

At the f0 dimension, the dominant distribution pattern ([+ f0 variability] English-L1 > English-L2 (4) > English-L2 (1) > BP [- f0 variability]) placed English-L1 with the highest means among the 4 corpora and BP with the lowest means, which suggests English native speakers mobilize more complex and varied f0 contours in speech. L1 transfer was more salient at the f0 level and seems to be more persistent among men, which adds to Urbani [2012]. A tendency towards the f0 prosodic patterns of the target language was also identified at this level and could be an effect of explicit instruction. This effect may have also influenced the greater speech rate ([+ speech rate] BP > English-L1 > English-L2 (4) > English-L2 (1) [- speech rate]) and spectral emphasis ([+ spectral emph] English-L1 > English-L2 (4) > BP > English-L2 (1) [- spectral emph]) for English-L2 (4), suggesting more fluency of the learners, and an overall improvement in marking syllable stress, respectively.

Conclusion

The metrics and acoustic parameters confirmed the first hypothesis that North American English-L1, the Brazilian English-L2, and the BP-L1 are rhythmically different systems. We confirmed the hypothesis that the data of English-L2 (1) would be more dissimilar in relation to English-L1 compared to the data of English-L2 (4), but orthogonal patterns of rhythmic development seem to coexist as a function of the different dimensions of prominence. Nevertheless, the approximation between the means of English-L2(4) and English-L1 in all dimensions suggest positive effects of explicit pronunciation in the development of prosodic features by learners of non-native languages. As future work, we intend to expand the analyzed corpora including the recordings of the 2nd and 3rd semesters as well as the other paragraphs of the text, and to analyze the correlation between the metrics and acoustic parameters and perceived degrees of foreign accent, intelligibility, and comprehensibility.

Acknowledgment

This study integrates a project that is partially financed by CNPq, process 438823/2018-4.

Bibliography

David Abercrombie. Elements of general phonetics. Aldine Atherton, Chicago, 1971. URL https://archive.org/details/elementsofgenera0000aber_u2o1. OCLC: 1256469368.

Catherine T. Best and Michael D. Tyler. Nonnative and second-language speech perception: Commonalities and complementarities. In Ocke-Schwen Bohn and Murray J. Munro, editors, Language Learning & Language Teaching, volume 17, pages 13–34. John Benjamins Publishing Company, Amsterdam, 2007. ISBN 9789027219732 9789027292872. DOI: 10.1075/lllt.17.07bes. URL https://benjamins.com/catalog/lllt.17.07bes.

Paul Boersma and David Weenink. Praat: doing phonetics by computer. 6.0.20, 2019. URL http://www.praat.org.

Ruth Elizabeth Cumming. Speech rhythm: the language-specific integration of pitch and duration. PhD thesis, University of Cambridge, Cambridge, 11 2010.

James Emil Flege, Katsura Aoyama, and Ocke-Schwen Bohn. The revised speech learning model (SLM-r) applied. In Ratree Wayland, editor, Second Language Speech Learning, pages 84–118. Cambridge University Press, 1 edition, February 2021. ISBN 9781108886901 9781108840637 9781108814614. DOI: 10.1017/9781108886901.003. URL https://www.cambridge.org/core/product/identifier/9781108886901%23CN-bp-2/type/book_part.

Robert Fuchs. Speech rhythm in varieties of English: evidence from educated Indian English and British English. PhD thesis, 2016. OCLC: 1084748564.

Peter Ladefoged. Course in Phonetics. Houghton Mifflin Harcourt P, New York, 1975. ISBN 9780155151802.

Aike Li and Brechtje Post. L2 acquisition of prosodic properties of speech rhythm: evidence from L1 mandarin and german learners of english. Studies in Second Language Acquisition, 36 (2):223–255, June 2014. ISSN 0272-2631, 1470-1545. DOI: 10.1017/S0272263113000752. URL https://www.cambridge.org/core/product/identifier/S0272263113000752/type/journal_article.

Ronaldo Mangueira Lima Jr and Ubiratã Kickhöfel Alves. A dynamic perspective on L2 pronunciation development: bridging research and communicative teaching practice. Revista do GEL, 16(2): 27–56, December 2019. ISSN 1984-591X, 1806-4906. DOI: 10.21165/gel.v16i2.2417. URL https://revistas.gel.org.br/rg/article/view/2417.

Kenneth L. Pike. The intonation of American English. University of Michigan publications Linguistics. University of Michigan Publications, Ann Arbor, 1945.

Franck Ramus, Marina Nespor, and Jacques Mehler. Correlates of linguistic rhythm in the speech signal. Cognition, 73(3):265–292, December 1999. ISSN 00100277. DOI: 10.1016/S0010-0277(99)00058-X. URL https://linkinghub.elsevier.com/retrieve/pii/S001002779900058X.

Leônidas José da Silva Junior and Plínio Almeida Barbosa. Speech rhythm of english as L2: an investigation of prosodic variables on the production of Brazilian Portuguese speakers. Journal of Speech Sciences, 8(2):37–57, August 2020. ISSN 2236-9740. DOI: 10.20396/joss.v8i2.14996. URL https://econtents.bc.unicamp.br/inpec/index.php/joss/article/view/14996.

R. I. Thomson and T. M. Derwing. The effectiveness of L2 pronunciation instruction: a narrative review. Applied Linguistics, 36(3):326–344, July 2015. ISSN 0142-6001, 1477-450X. DOI: 10.1093/applin/amu076. URL https://academic.oup.com/applij/article-lookup/doi/10.1093/applin/amu076.

Martina Urbani. Pitch range in L1/L2 English. An analysis of F0 using LTD and linguistic measures. Methodological Perspectives on Second Language Prosody, pages 79–83, 2012.

Change-Point Analysis in language development: a study of voice onset time production in a multilingual system

Laura Castilhos Schereschewsky[1,2], Ubiratã Kickhöfel Alves[1,3]

1 . Universidade Federal do Rio Grande do Sul

2 . CAPES

3 . CNPq

Introduction

According to Complex Dynamic Systems Theory (CDST) [1], when it comes to multilingual development, we need to think about the interconnectedness of the system components. Departing from this assumption, we follow Kupske's concept of language attrition[2], which characterizes this phenomenon as the force resulting from the contact of two bodies, in this case, two languages, that are in constant movement [Kupske, 2016, p. 39–40]. This concept embraces the CDST premise that change is inherent to development. Thus, if the system is in constant movement, we may find it in continuous change in a given state, and language variability is expected to be found. Sometimes, the system may go through significant changes that exceed its current state [van Dijk and van Geert, 2007]. If these particular changes lead to the reorganization of the system as a whole (as in the emergence of a new attractor state), we call them 'phase transitions' or 'phase shifts'. According to Hepford [2020], this new attractor state is not necessarily something new to learners, as it "could be a language form that they

[1] ; ; ; and

[2] In this paper, we do not differentiate 'language attrition' from 'language transfer' or 'language drift'. We will use the three terms interchangeably.

are exposed to regularly but have not had the cognitive ability to adapt to, or an event that pushes a learner to adapt and self-organize resulting in using a new form" [Hepford, 2020, p. 162–163].

Based on the aforementioned assumptions, this study aims to investigate the phenomena that occur in the development of the additional languages of trilingual speakers, native speakers of Brazilian Portuguese (BP-L1) and non-native speakers of English (L2) and French (L3). Specifically, this longitudinal study analyzes, over a period of three months (with 12 weekly datapoints), the development of the production of Voice Onset Time (VOT), observing possible phase shifts through change-point analyses [Taylor, 2000, cf.] provided by the Change-point analyzer v.2.3 software [TAYLOR ENTERPRISES]. The study included a period of pedagogical intervention to accelerate the development of the positive VOT pattern with the characteristic aspiration of English. This teaching intervention took place over six explicit pronunciation instruction sessions, conducted in the weeks of datapoints 4 to 9. We aimed to discuss to what extent the accelerated development of an L2 with a typologically different VOT pattern causes changes in the development of the L3 and L1 subsystems, as well as show the inter-relation of the two additional languages over time.

According to the literature [Schereschewsky, 2021, cf.], from the study of VOT, we can observe the multidirectionality of transfer and the adaptability and the self-organization of language subsystems. Therefore, this study intends to provide empirical and theoretical input into a larger understanding of these aspects. This may shed light on the development of additional languages in the light of CDST. As this is essentially a theory about change, we aim to raise issues such as language development and its ongoing "process" in time [Lowie and Verspoor, 2015, 2019, cf.], the interconnectivity of typologically different subsystems, data variability, and the emergence of new attractor states and phase shifts.

Method

As addressed in the previous section, the main goal of this study was to inferentially verify possible phase shifts in VOT patterns, in each of the language subsystems, especially after the beginning of explicit pronunciation instruction in English. For that, we carried out change-point analyses [Taylor, 2000, cf.].

In order to achieve our goal, we proposed a methodology in which changes and interactions among the subsystems of multilingual speakers could be investigated through accelerating L2 VOT development. The experiment was built with a longitudinal design in an A-B-A format [Hiver and Al-Hoorie, 2020, cf.], with 12 datapoints, which were intersected in the midpoints with 6 sessions of explicit pronunciation instruction in English. This intervention took place between the weeks referring to datapoints 4 and 9, and all instructional sessions were conducted with a communicative approach.

In this study, we replicated different process-oriented analyses to encompass and address variability [de Bot et al., 2013], conducting the same experiment with five participants from different backgrounds, different ages, with different proficiency levels in their additional languages, and different routines. Due to space restrictions, in this paper we will focus on the results from one particular participant[3], who was 24 years old at the time, a graduate student who worked as a French teacher. This participant took a self-evaluation test [?]and graded herself a 6 in English and a 10 in French[4].

The participant was presented with three different reading tasks. In each data collection session, she received 23 carrier sentences (repeated three times each) with 18 target words with /p/, /t/ and /k/ in word-initial position and 5 distractor words. The BP and English instruments were the same as in Kupske (2016), for both matters of consistency and comparisons of results with previous studies. We also used the same methodological control as Kupske in the development of the French instrument [Schereschewsky, 2021, cf.]. Because she received the same target words, the order of the carrier sentences was randomized and the distractor words were changed each week. This study was conducted during the pandemic of COVID-19, so the participant accomplished the experiment in an individual setting and was asked to complete each task taking time intervals between them.

All audio recordings from the reading tasks were analyzed acoustically in the Praat v.6.1.16 software [?]. Due to time and space restrictions, only the absolute values of VOT production were considered. As for VOT measurements, similar criteria to previous works were used: selecting the voiceless interval between the burst of the stop consonant and the first regular pulse of the following vowel.

As for the statistical analyses, the participants' developmental trajectories were plotted, considering minimum, maximum, and mean values from the tokens of each stop consonant in each datapoint. Following that, change-point analyses [Taylor, 2000, Steenbeek et al., 2012, Baba and Nitta, 2014, Han and Hiver, 2018, Englhardt et al., 2020, Henry et al., 2021, cf.] were conducted. Change-point analysis is an infer-

[3] This participant is referred to as Participant 5 in previous works from this project. For more information on the other participants, see Schereschewsky [2021].

[4] It is interesting to note that her L3 was more active than her L2, because she worked as a French teacher, even though she had started studying English before she even started learning French.

ential method that uses resampling and cumulative sums to identify a pattern shift, or the point of change, in a set of longitudinal data.

The Change-Point Analyzer software is able to detect several longitudinal changes. By running a fast analysis of cumulative sums and bootstrapping, for each change in pattern, the software provides practical information, including the confidence level, which indicates a probability that a change has actually occurred, and the confidence interval, which indicates when that change has occurred. Figure 1 shows the first output tab from the software, with the change-point analysis visual plot.

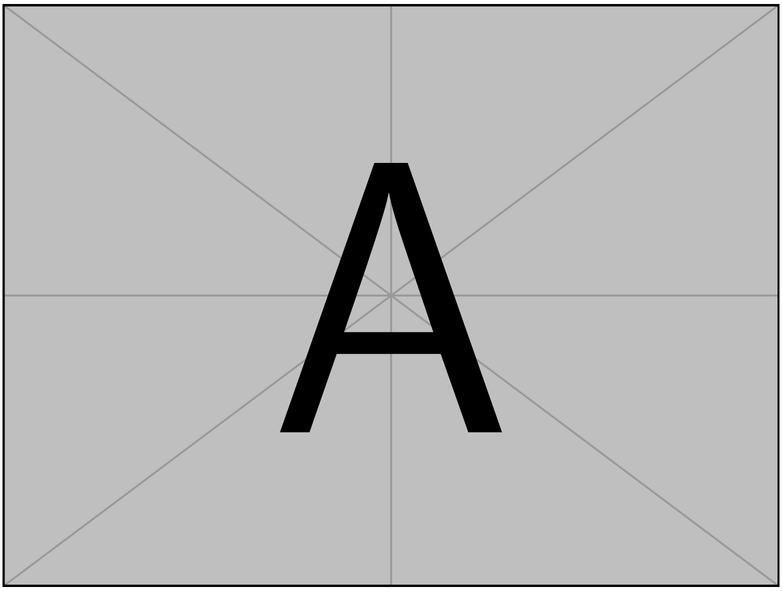


Figure 5: Outputs - change-point analysis visual plot.

In Figure 5, the bold black line in the graph represents the raw data of the mean VOT values of [p] in Participant 5's L1 over the 12 collection points. The dark blue lines represent the amplitude range of the control limits, that is, the maximum range of variation in which the values can fluctuate, assuming that no change has occurred (if the black line exceeds the control limits, we will have a first indicative that a change has taken place, which may simply be an outlier or an indication of an actual phase shift). The lighter blue background represents the area that should contain all values varying within the control limits. The displacement of this area in light blue at the bottom of the graph actually indicates a phase shift, as the average values within the first segment show a sudden change, starting to vary in a different range, represented by the second segment of the area in lighter blue. Figure 6 shows the second output tab from the software, with the significant changes of the set of data.

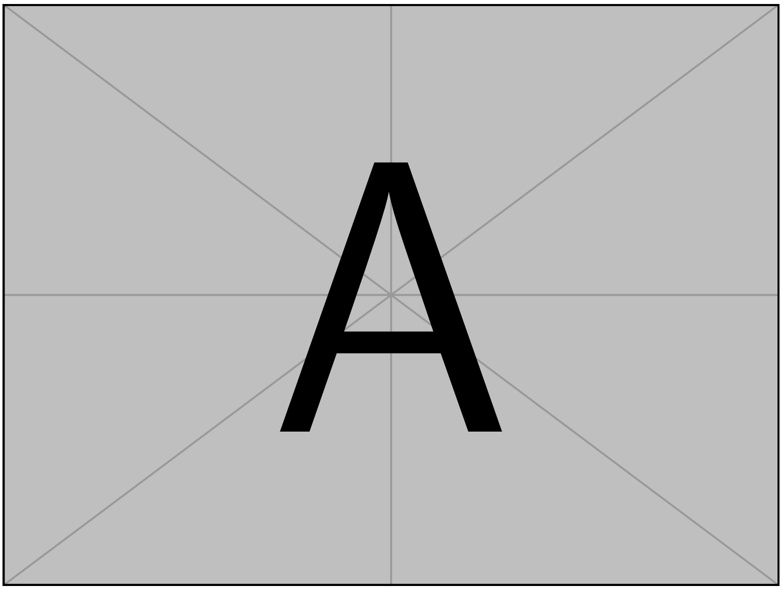


Figure 6: Outputs - table of significant changes.

The table in Figure 6 indicates the estimated point of change to another phase, in this case, in Datapoint #5, with a confidence level of 97%[5], indicated by the confidence interval (which, in this case, points exactly to session 5). Next, the table indicates the values before and after the change, that is, the average values of variation in the first phase (considering the average of all inputs within this first phase), which go from 34.25ms to 46.29ms in the second phase. Finally, the level of change indicates its importance. In this particular example, the Level 1 change indicates that this was the first significant change identified by the software in the first analysis run of the data. Other change levels may appear, depending on how many phase changes are identified and whether these are significant. Finally, Figure 7 shows the third output tab from the software, with the visual chart showing the cumulative sums.

The chart in Figure 7 represents the cumulative sum analyses (CUSUM). According to Taylor [2000, p. 6], ”they are the cumulative sums of differences between the values and the average”. These dif-

[5] The Change-point Analyzer only presents, in the outputs, intervals that have at least 95% confidence. The more spaced the confidence interval, the lower the confidence level for a change to have occurred at the point identified by the software.

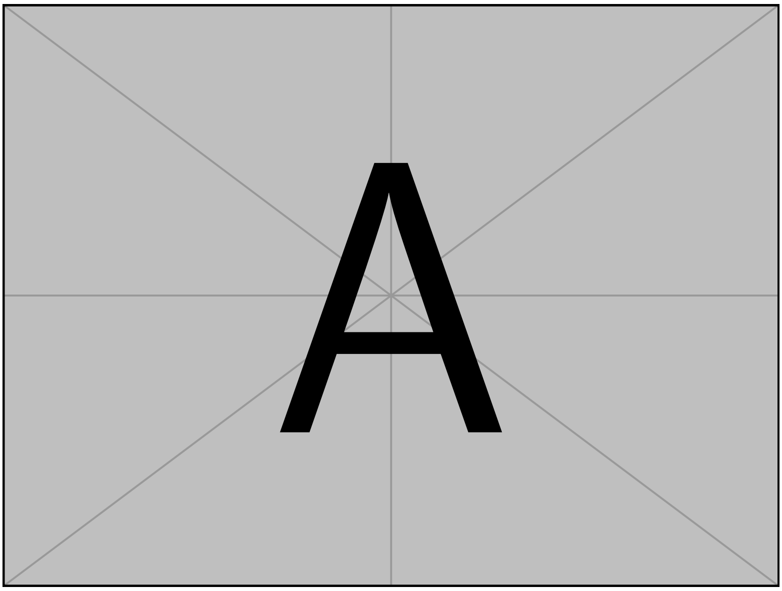


Figure 7: Outputs - table of significant changes.

ferences sum to zero so the cumulative sum always ends at zero. Thus, in the CUSUM graph from this data, a downward sloping line can be seen, indicating that the values in that period have a tendency to be below the general average, until there is a change in the direction of the line, starting to have an upward slope, indicating that values from that portion of the graph tend to be above the overall mean. We should also look at the shaded background of the chart, which indicates whether and where there has been a significant change in the slope of the line, referring to the table with confidence intervals. Another important information brought by the CUSUM chart is that the straighter the line, regardless of its direction (up or down), the greater the certainty that no change occurred in that period. On the other hand, the more curved the line is, as is the case between points 4 and 9, the greater the possibility that other changes (from other levels) have taken place.

Results

As previously mentioned, change-point analyses are used to identify the points at which a pattern change occurs in a longitudinal dataset. Thus, change-point analyses help identify the developmental stages of VOT production, checking attractor states in phases of relative stability in each language. In this section, only a summary of the significant changes and the most relevant charts for the discussion will be presented. A table will presented for each language, with the significant results split by consonant (/p, t, k/), of the data collection session the change took place, the confidence interval, the confidence level (in percentages), the mean values before and after the change (which is related to the averages of variation of values within the control limits of each phase) and the level of change (degree of importance in the analysis by the software). Table 1 shows the results of Brazilian Portuguese-L1.

Stop	Measure	Session	Conf.level	From	To	Shift
[p]	Mean	5	97%	34,252	46,289	↗
[p]	Max	5	95%	69,097	84,571	↗
[t]	Mean	11	100%	40,181	32,775	↘
[t]	Max	11	99%	76,253	48,585	↘
[k]	Mean	4	96%	59,073	75,531	↗

Table 4: Change-point analysis of BP-L1.

First, we emphasize the interconnectedness of the language subsystems, which makes it possible for a native language to change, even if it is typologically different from a language that underwent an intervention, as we found significant phase changes in the production of VOT in Portuguese-L1 in the three stops.

For [p], we found a Level 1 phase shift in the means in Datapoint 5, when the averages change from 34.25ms to 46.29ms, and a Level 3 change in maximums around Datapoint 5, when the averages increase from 69.1ms to 84.57ms.

For [t], in which we also found significant phase changes in the averages and maximum instances, the data are somewhat more interesting. In both measures, Level 2 phase changes are found, in which the VOT decreases in duration. For the means, phases shifted from an average of 40.18ms to 32.78ms. For the maximum instances, they shifted from 76.25ms to 48.59ms. However, as both changes are Level 2 and this new phase with shorter VOT measures only starts by the end of the analyzed period, around Datapoint 11, it is also necessary to visually analyze these data, since there is also the possibility that another (non-significant) change occurred in previous datapoints. Figure 8 shows the change-point analysis plots for the two measures.

The graphs of the means and the maximum instances of [t] show that the two measures presented a very similar behavior during the analyzed period. Comparing the initial and final points of the mea-

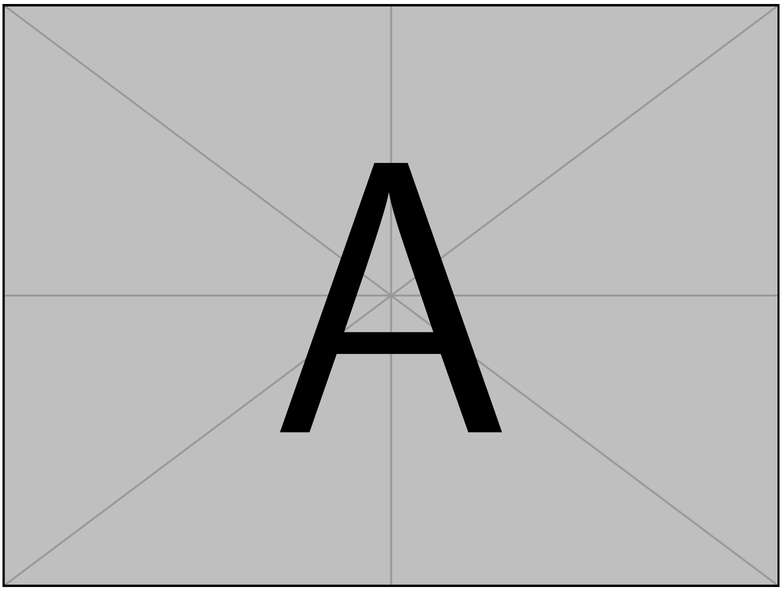


Figure 8: Change-point analysis of means and maximum instances of [t] in Portuguese-L1.

surements, there is a clear trend towards a decrease in the descriptive values of VOT, which is in accordance with the phase shift found with a decrease of averages. However, there is also a very clear indication that the data may have undergone another phase shift, around Datapoint 3, where the VOT appears to have increased in duration. The CUSUMs graph shows the possibility of another phase shift due to the sudden change in the direction of the cumulative sums line on the third datapoint. However, the software did not verify this change as significant to include it in the outputs. What was included in the outputs was a significant Level 1 phase shift in the means of [k], where there was an increase in the averages from 59.07ms to 75.53ms in Datapoint 4, thus after the beginning of the intervention, once again showing that even the subsystem of a typologically different language is subject to change as a result of another one changing.

The English-L2 results also bring valuable data to the discussion. For [p], there is a significant change in Datapoint 4, the first after the start of the intervention. When it comes to the means, the Level 2 phase shift occurs when the average changes from 54.55ms to 95.42ms. For the maximums, the Level 1 change occurs with an increase of the averages from a phase of 104.37ms to 142.81, with very high values of VOT production for a bilabial stop.

Stop	Measure	Session	Conf.level	From	To	Shift
[p]	Mean	4	97%	54,553	95,416	↗
[p]	Max	4	94%	104,37	142,81	↗
[t]	Min	12	91%	40,409	26,24	↘
[t]	Mean	4	94%	62,407	105,89	↗
[t]	Mean	11	93%	105,89	80,07	↘
[t]	Max	4	100%	110,1	147,2	↗
[k]	Mean	4	99%	82,73	112,96	↗
[k]	Max	5	91%	130,65	157,73	↗

Table 5: Change-point analysis of English-L2.

For [t], we found significant phase changes for the three analyzed measures, but each measure presented a different result. For minimums, for instance, a Level 3 phase shift occurs around Datapoint 12, with a decrease in averages from 40.41ms to 26.24ms. With such a large confidence interval, which covers the entire intervention until the end of the study, in addition to the fact that it is a Level 3 change, there remains a possibility of another, less significant phase shift, in some other datapoint in that interval. For the means, two significant phase changes were identified, one of Level 1, in Datapoint 4, with an increase in the averages from 62.40ms to 105.89ms, and one of Level 2 , at the end of the study, around Datapoint 11, this time with a decrease in averages (much like what happens in her L1) from 105.89ms to 80.07ms, a higher average than in the initial phase. The maximums of [t] present a third pattern of behavior, with a Level 3 phase shift around Datapoint 4, with an increase from 110.1ms to 147.2ms in the later phase.

Figure 9 shows the plots of the change-point analyses of the three analyzed measures of [t] in English-L2. Although the three measures show completely different behaviors in the outputs, as evidenced by the first graph of each one, the CUSUM graphs of the three are very similar, indicating sudden changes in the slope of the cumulatives sums line at least twice on each[6]. This pattern would be indicative of at least three distinct phases during the study, in which the first phase change would represent an increase in the average VOT values, and the second a slight decrease, as we verified in the means of [t], almost always involving the same datapoints. For minimums and maximums, however, a possible second phase change was not significant, leaving the observation only for a qualitative discussion.

[6] A downward-sloping CUSUM line indicates values below the overall average, while an upward-sloping line indicates values above the overall average. A change in the slope of the line represents a change in trend and, being a significant change, corresponds to a phase shift.

Finally, for [k], we found significant phase shifts in the means and in the maximums, both indicating an increase in VOT values. For the

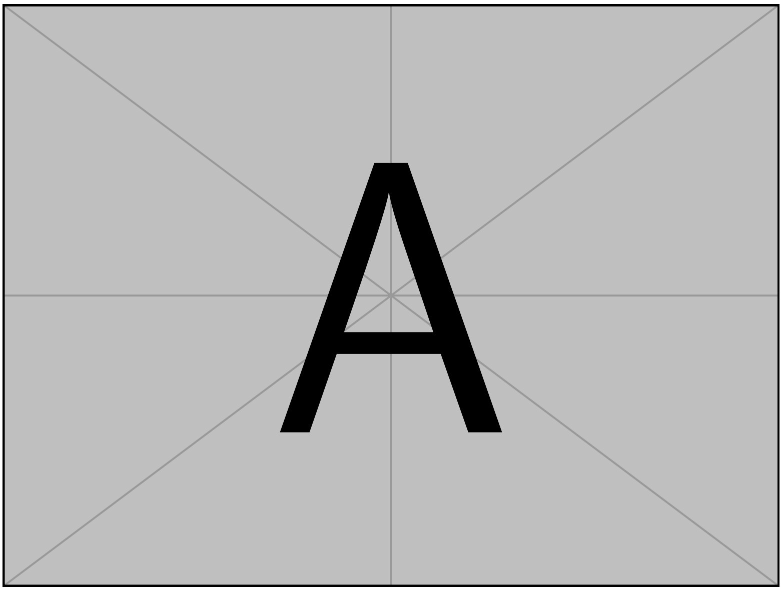


Figure 9: Change-point analyses of the minimums, means and maximums of [t] in English-L2.

means, the change occurred in Datapoint 4, where a new phase went from an average of 82.73ms to 112.96ms. For the maximums, the change occurred around Datapoint 5, when the results show an increase of the averages from 130.65ms to 157.73ms. Overall, all these English-L2 data are extremely valuable in showing the influence of explicit instruction in the development of new attractor states, that is, new phases developing a non-native positive VOT pattern with long-lag aspiration. Furthermore, these data highlight change as an inherent characteristic of a developing system, showing that the language remains in motion even after the end of an intervention.

Stop	Measure	Session	Conf.level	From	To	Shift
[p]	Mean	4	97%	30,987	40,782	↗
[p]	Max	4	96%	60,697	78,477	↗
[t]	Mean	4	99%	33,437	40,15	↗
[t]	Max	4	95%	58,91	70,327	↗
[k]	Mean	6	92%	59,452	66,256	↗

Table 6: Change-point analysis of French-L3.

Once again, we can observe significant phase shifts in the three consonants of a language subsystem that is typologically different from the language that received explicit instruction during the intervention, showing the interconnectivity of the system as a whole. Interestingly, all the identified changes present new phases with an increase in the VOT values in the French language. For [p], the means and maximums undergo phase shifts in Datapoint 4. For the means, the Level 2 shift showed a change in the averages from 30.99ms to 40.78ms. For the maximums, the Level 1 shift showed changed averages from 60.68ms to 78.48ms. For visualization purposes, the graphs referring to the phase shifts in the maximums of [p] are in the Figure 10.

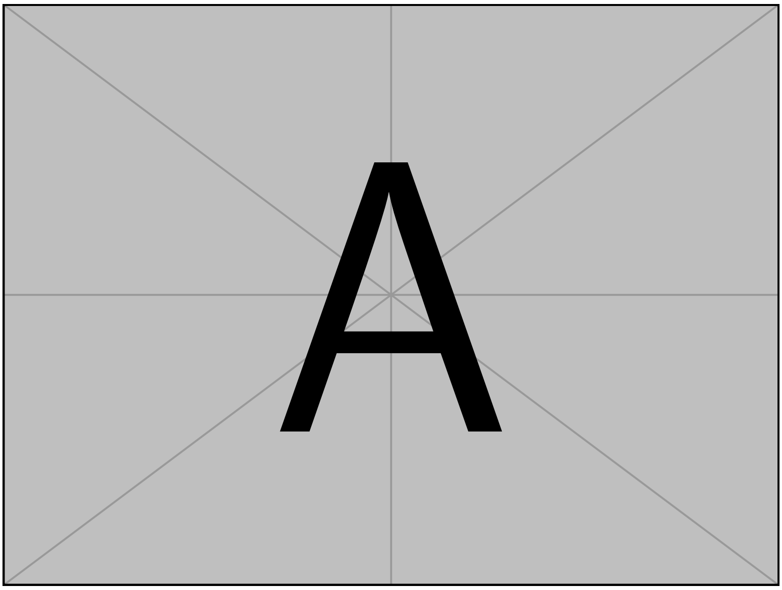


Figure 10: Change-point analyses of the maximums of [p] in French-L3.

For [t], Level 1 phase shifts were also identified in the means and maximums, occurring around Datapoint 4. For the means, the phase shift indicates an increase in average from 33.44ms to 40.15ms, and in the maximums, from 58.91ms to 70.33ms.

On the other hand, finally, we only found a significant phase shift in the means of [k]. The Level 1 change was identified around Datapoint 6, indicating an increase in average from 59.45ms to 66.26ms. Again, we reiterate the subsystem's ability to change under the influence of changes in other subsystems, especially the L2, since the new phases were always identified from the beginning of the instruction period in that language.

Final Considerations

We highlight the relevance of change-point analyses in verifying the emergence of new developmental phases and we emphasize that change-point analyses allow us to identify more than one change in each language subsystem, as shown in the L2 data. Considering that changes in a multilingual system are constant and that even attractor states are not permanent, an analysis of this sort provides valuable information on the process of multilingual development. As shown in our results, languages are entangled and interconnected in a multilingual system, and they influence one another. Finally, we hope to contribute to the area of language development in the light of Complex Dynamic Systems Theory. Discussing methods of analysis that verify developmental changes is always necessary. Specifically, change-point analyses help to identify the emergence of new stages of development. As a non-linear process, we acknowledge the fluctuations of the VOT values in the new developmental phase, as it probably refers to a less strong attractor state, and even the emergence of a third phase, different from both the initial phase and the phase under the influence of the pedagogical intervention. These results show that language and learning are constantly changing, demonstrating the relevance of an approach via CDST, given that this, after all, constitutes a theory essentially about change.

Bibliography

Kyoko Baba and Ryo Nitta. Phase Transitions in Development of Writing Fluency From a Complex Dynamic Systems Perspective: Phase Transition in Development of Writing Fluency. Language Learning, 64(1):1–35, March 2014. ISSN 00238333. DOI: 10.1111/lang.12033. URL `https://onlinelibrary.wiley.com/doi/10.1111/lang.12033`.

Kees de Bot, Wander Lowie, Steven L. Thorne, and Marjolijn Verspoor. Dynamic systems theory as a comprehensive theory of second language development. In María del Pilar García Mayo, María Junkal Gutierrez Mangado, and María Martínez Adrián, editors, AILA Applied Linguistics Series, volume 9, pages 199–220. John Benjamins Publishing Company, Amsterdam, 2013. ISBN 9789027205254 9789027205285 9789027272225. DOI: 10.1075/aals.9.13ch10. URL `https://benjamins.com/catalog/aals.9.13ch10`.

Adrian Englhardt, Jens Willkomm, Martin Schäler, and Klemens Böhm. Improving semantic change analysis by combining word embeddings and word frequencies. International Journal on Digital Libraries, 21(3):247–264, September 2020. ISSN 1432-5012, 1432-1300. DOI: 10.1007/s00799-019-00271-6. URL http://link.springer.com/10.1007/s00799-019-00271-6.

Jiwon Han and Phil Hiver. Genre-based L2 writing instruction and writing-specific psychological factors: The dynamics of change. Journal of Second Language Writing, 40:44–59, June 2018. ISSN 10603743. DOI: 10.1016/j.jslw.2018.03.001. URL https://linkinghub.elsevier.com/retrieve/pii/S1060374317304642.

Alastair Henry, Cecilia Thorsen, and Peter D. MacIntyre. Willingness to communicate in a multilingual context: part one, a time-serial study of developmental dynamics. Journal of Multilingual and Multicultural Development, pages 1–20, June 2021. ISSN 0143-4632, 1747-7557. DOI: 10.1080/01434632.2021.1931248. URL https://www.tandfonline.com/doi/full/10.1080/01434632.2021.1931248.

Elizabeth Hepford. Chapter 7. The elusive phase shift: Capturing changes in L2 writing development and interaction between the cognitive and social ecosystems. In Gary G. Fogal and Marjolijn H. Verspoor, editors, Language Learning & Language Teaching, volume 54, pages 161–182. John Benjamins Publishing Company, Amsterdam, June 2020. ISBN 9789027205575 9789027205582 9789027261144. DOI: 10.1075/lllt.54.07hep. URL https://benjamins.com/catalog/lllt.54.07hep.

Phil Hiver and Ali H. Al-Hoorie. Research methods for complexity theory in applied linguistics. Number 137 in Second language acquisition. Multilingual Matters, Bristol ; Blue Ridge Summit, 2020. ISBN 9781788925730 9781788925747.

Felipe Flores Kupske. Imigração, atrito e complexidade: a produção das oclusivas surdas iniciais do inglês e do português por sul-brasileiros residentes em Londres. PhD thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2016.

Diane Larsen-Freeman and Lynne Cameron. Complex systems and applied linguistics. Oxford applied linguistics. Oxford university press, Oxford, 2008. ISBN 9780194422444.

Wander Lowie and Marjolijn Verspoor. Variability and variation in second language acquisition orders: A dynamic reevaluation: Variability in acquisition orders: Dst. Language Learning, 65(1): 63–88, March 2015. ISSN 00238333. DOI: 10.1111/lang.12093. URL `https://onlinelibrary.wiley.com/doi/10.1111/lang.12093`.

Wander M. Lowie and Marjolijn H. Verspoor. Individual Differences and the Ergodicity Problem: Individual Differences and Ergodicity. Language Learning, 69:184–206, March 2019. ISSN 00238333. DOI: 10.1111/lang.12324. URL `https://onlinelibrary.wiley.com/doi/10.1111/lang.12324`.

Laura Castilhos Schereschewsky. Desenvolvimento de voice onset time em sistemas multilíngues (português - l1, inglês - l2 e francês - l3): discussões dinâmicas a partir de diferentes metodologias de análise de processo. Master's thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2021.

Henderien Steenbeek, Louise Jansen, and Paul van Geert. Scaffolding dynamics and the emergence of problematic learning trajectories. Learning and Individual Differences, 22(1):64–75, February 2012. ISSN 10416080. DOI: 10.1016/j.lindif.2011.11.014. URL `https://linkinghub.elsevier.com/retrieve/pii/S1041608011001646`.

Wayne Taylor. Change-Point Analysis: A Powerful New Tool For Detecting Changes, April 2000. URL `https://variation.com/change-point-analysis-a-powerful-new-tool-for-detecting-changes/`.

TAYLOR ENTERPRISES. Change-Point Analyzer. URL `https://variation.com/product/change-point-analyzer/`.

Marijn van Dijk and Paul van Geert. Wobbles, humps and sudden jumps: a case study of continuity, discontinuity and variability in early language development. Infant and Child Development, 16(1):7–33, February 2007. ISSN 15227227, 15227219. DOI: 10.1002/icd.506. URL `https://onlinelibrary.wiley.com/doi/10.1002/icd.506`.

Production of English [Cs] clusters by Brazilian speakers: effects of orthography, phonological environment and task type

Wellington Araujo Mendes Junior[1]
1 . Federal University of Minas Gerais

This study examines the effects of orthography, phonological environment and task type in the production of English [Cs] clusters by Brazilian Portuguese (BP) speakers. Two orthographic patterns were examined for English nouns whose plural is pronounced as a (stop + sibilant) cluster. One of the patterns presents two consonants word-finally - cups, cats, ducks - whereas the other one presents a silent vowel <e> between two consonants: grapes, plates, cakes. The goal was to assess whether these different orthographic patterns would trigger the production of an epenthetic vowel. Additionally, it was assessed whether different phonological environments would influence the voicing property of the final sibilant. As it is known, word-final English sibilants are prone to progressive assimilation (e.g. cups [kʌps], bags [bægz]), rather than regressive assimilation – as it occurs in BP (mês [mes], mês anterior [mez ə̃.te.ɾi.ˈoɾ]). An experiment was designed to test the production of [Cs] clusters in English nouns and in BP forms undergoing sound change. Harmonics-to-noise ratio (HNR) was used to measure sibilant voicing, whereas the presence of epenthetic vowels was assessed categorically. Results showed that English learners are more likely to pronounce a vowel when the orthographic pattern is <Ces> rather than <Cs>, and this occurs regardless of the visual presentation of the words. Moreover, HNR rates showed that fully voiced sibilants tend to occur in L2 English when the consonant is both preceded and followed by a vowel. These findings are discussed in light of the Exemplar Model in L2 Phonology (EML2P) [Cristófaro-Silva and Guimaraes, 2021, Mendes Jr. and Cristófaro-Silva, 2022 in press]. The analysis based on the EML2P showed that robust patterns from the L1 are adopted in L2, including fine phonetic detail that reflects subphonemic properties.

Introduction

Traditional phonological models assume that English plural suffixes and third person singular present forms are subject to a phonological rule.The underlying representation for regular plural and 3rd person singular present is assumed to be /z/ [Hayes, 2011]. A progressive assimilation rule predicts that if a vowel or a voiced consonant precedes /z/, the output is [z], as indogs [dɒgz], trees [triːz] and pies [paɪz]. If a voiceless consonant precedes /z/, it surfaces as[s], as in cups [kʌps], cats [kæts] and ducks [dʌks]. Finally, if an alveolar fricative or anaffricate precede the sibilant, the outcome is [ɪz], as inbuses [bʌsɪz], quizzes [kwɪzɪz] and watches[wɒtʃɪz]. However, when we consider the orthography of English plural forms, two possible spellings are associated with theaforementioned sound patterns. Nouns can either end in a consonant followed by the letter <s>, as in books andjobs or the by letters <es>, as in cakes and cubes. Brazilian Portuguese, on the other hand, presents mainly the <Ces> pattern, as it occurs incheques and clubes, whereas only some few nouns present the <Cs> pattern: biceps, forceps,volts.

As a consequence of an ongoing sound change, word-final [Cs] clusters[7] are currently very productive some Brazilian Portuguese plural forms: crepes, potes, cheques [Soares, 2016]. The alternation between [Cs] ~ [Cis] word-finally in BP follows from the reduction and eventual loss ofunstressed high front vowels when flanked between a consonant and a word-final sibilant. It seems that such alternation also applies toplural forms produced by Brazilian speakers of L2 English, as incakes [keɪks] ~ [’keɪ.kis].

[7] For the purpose of the present discussion, we refer to [Cs] as any (consonant + sibilant) sequence. However, as it will be discussed later, the sibilant may be either voiced or voiceless.

This paper intends to investigate [Cs] clusters in English regular plural forms (e.g. cups [kʌps], grapes [greɪps])produced by Brazilian speakers of L2 English in an attempt to address the question of whether an ongoing sound change from the L1 plays a rolein L2 learning. Additionally, we aim to assess how orthographic and phonological representations are related. Studies on the relationshipbetween orthography and phonology have increased in recent years [Rafat, 2015, Hamann and Colombo, 2017, Zhou, 2021]. The main research questions in this topic aim to explain how L2 learners mediate the relationship between the already knownphonological and orthographical knowledge from the L1 in order to buildan L2. Thus, an important question we pose is whether differentorthographic patterns trigger different pronunciations of [Cs] clusters in L2 English.

This paper is organized as follows. The next section reviews studies on the production of English [Cs] clusters by Brazilian speakers. Thethird section describes the methodology adopted in this study. The fourth section discusses our findings and is followed by theconclusions.

Production of English [Cs] clusters by Brazilian speakers

Several works have addressed the relationship between orthography and the pronunciation of L2 English forms by Brazilian speakers. One of themain concerns have been to assess whether the presence of an epenthetic vowel in L2 English is influenced by a letter corresponding to a vowel. Delatorre [2006] investigated the production of English [Cs] clusters that occur in past andparticiple forms by Brazilian speakers of L2 English (e.g. moved and robbed). Two epenthetic vowels were attested: one epentheticvowel breaks up the word-internal consonant cluster and the other one prevents word-final consonants, as in asked['as.ke.dʒi] and saved['seɪ.ve.dʒi]. Delatorre [2006] claimed that theorthographic input, which was present in a reading task, favored higher rates of an epenthetic vowel, as opposed to a free speech task, whichdid not present any orthographic stimulus. Therefore, she argued that the orthographic input favored the presence of epenthetic vowels in the pronunciation of L2 English by Brazilian speakers.

Although she did not focus on the production of [Cs] clusters, Silveira [2007] also investigated the production of word-final epenthesis in Brazilian speakers of L2 English. She compared words whose final letter was a consonant (e.g. mad[mæd]) to words whose final letter was a silent ⟨e⟩ (e.g. made [meɪd]). Her results showed that words ending in asilent ⟨e⟩ presented higher rates of epenthesis than words that ended in a consonantal letter. Akin to Delatorre [2006], the results of Silveira [2007] showed that a reading task favored higher rates of epenthetic vowels than a free speech task, indicating that orthographic input (and the task type) contributed to the production of an epenthetic vowel.

Another case of epenthetic vowels reported in the literature involves word-final consonant and sibilant sequences, [Cs], which typically appear in regular plural and 3rd person singular present forms in English. It is known that Brazilian speakers of L2English tended to insert an epenthetic vowel between two word-final consonants, as it occurs, for example, in cakes [keɪks]~ ['keɪ.kis] [?]. Interestingly, works that considered 3rd person singular present and regular plural forms in English spoken by Brazilian speakers did not account for an epenthetic vowel. They were rather concerned with voice agreement.

Zanfra [2013] studied sibilant voicingin L2 English by Brazilian speakers. Although her focus was not specifically on plural forms, her results shed some light on the current discussion. The author tested whether the BP voicing assimilation rule involving adjacent segments in word boundaries would apply in L2 English learners' productions. Her results showed that sibilants tended to be voiced when followed by a voiced consonant (e.g. Thehouse backyard is huge) or by a vowel (e.g. The mouse I saw is white).Conversely, a sibilant was voiceless when the following context was a pause (e.g. I won't go if he goes.) or a voiceless consonant (e.g. These pancakes are great). Zanfra [2013] suggested that Brazilian speakers of L2 English speakers transfer the BP regressive assimilation rule into their L2 English.

Fragozo [2017] investigated the voicing of sibilants in English regular plural forms and 3rd person singular presented by Brazilian speakers of L2 English. She assessed the extents to which a sibilant would be voiced after a voiced consonant, as in dogs or clubs, which would reflect the acquisition of a progressive assimilation rule from English. Fragozo [2017] also examined words in context to verify if the regressive assimilation rule, which applies to BP, would be transferred to L2 English. She found that voiced sibilants tended to follow the regressive assimilation rule from BP, whereas the English progressive assimilation rule had a very low rate in her data (0.6%). She argues that the low rates of voiced sibilants [z] in L2 English by Brazilian speakers follows from the fact that these consonants are only partially voiced in English. Data from her control group of native speakers presented 44% of expected voiced sibilants. Thus, as sibilants are partially voiced in English, they would not be accessible in L2 English.

Zanfra [2013] and Fragozo [2017] both investigated voicing agreement within a rule-based approach where there would be a competition between a regressive assimilation rule from BP and a progressive assimilation rule from English. A question that arises from this assumption is whether a rule that is transferred from the L1 to the L2 could change as time goes by. Another issue which is polemic lies on the role played by orthography, as in hou<se> orbu<s> [Zanfra, 2013]. Orthography cannot bemodelled within a rule-based approach as it is not part of Grammar. Furthermore, the rule-based approach adopted by Zanfra [2013] and Fragozo [2017] neglected the role played by an epenthetic vowel that may intervene between the two word-final consonants, as in cakes[keɪks] ~ [’keɪ.kis]. Additionally, they did not account for the gradience of sibilant voicing.

Unlike previous works which adopt rule-based approaches, this paper models L2 phonology within an Exemplar Model by considering representation robustness and the role of fine phonetic detail in shaping mental representations. Within this proposal, orthography is modelled as part of the linguistic knowledge of literate speakers and sound patterns display a great range of variability and gradience.

Methodology

A set of 36 plural nouns ending in a sequence of (stop + sibilant) were considered in BP. These words present a single orthographic pattern: <Ces>, as in cheques [ʃɛks] ~ [’ʃɛ.kis] ‘cheques’. For the L2 English case study, a set of 36 words were selected, where 15 words display the orthographic pattern <Ces>, as in grapes [greɪps], and the other 21 words display the orthographic pattern <Cs>, as in maps [mæps].

The experiment comprised two tasks. The first one consisted of a picture-counting task in which participants were asked to count and name the items shown in the pictures. Short carrier sentences that did not include orthographic stimuli of the target words were given. The second trial consisted of a reading task. Initially, participants were asked to read 72 BP sentences aloud. Alike the picture-counting task, BP nouns in the reading task were followed by either a vowel or a voiceless consonant. On the other hand, L2 English nouns were followed by either a vowel or a pause. The overall number of syllables was controlled for both languages: 4 in English and 12 in BP, considering the deletion of the [i] vowel. Sentence-level intonation and the morphological class of each word were also controlled.

A group of six Brazilians studying at the Federal Center for Technological Education of Minas Gerais, in the city of Araxá, participated in this study[8]. All participants were high school students who had been taking English classes as part of the school’s curriculum for about one year. The group consisted of 3 males and 3 females and their ages ranged from 15 to 17. All participants displayed either B1 or B2 proficiency levels (intermediate learners) of the Common European Framework of Reference for Languages.

Due to the recent COVID-19 pandemic, all interactions were performed remotely. Experiments were recorded with the Open Broadcaster Software Studio at 48 kHz sampling rate. The obtained recordings were converted into WAVEform audio format by the soft-

[8] This research has been approved by the ethics committee from the Universidade Federal de Minas Gerais, reference number: CAAE: 15116119.9.0000.5149.

ware Adobe Premiere 2020, which was able to maintain the same sampling rate as the original files. The average time to complete the experiment was 45 minutes. A total of 648 tokens were collected for the L2 English study. For the BP study, 432 tokens were collected. Samples were edited and manually annotated using Praat TextGrids [?].

Besides assessing the presence or absence of a vowel between [Cs] clusters, this research also considered the voice quality of word-final sibilants. In BP, only voiceless sibilants occur word-finally, unless a vowel follows it, to which a voiced sibilant occurs. In English, voiced and voiceless sibilants occur word-finally. When a vowel follows the sibilant, the voice quality remains as it formerly was (rather that changing as it occurs in BP). We posited that word-final voiceless sibilants would be favored in L2 English, as it is the more robust pattern in L1. We also posited that a voiced sibilant occurs at higher rates in an intervocalic position: [Cis] followed by a word-initial vowel.

Voicing was measured under Harmonics-to-noise ratio. Each token was extracted to a separate sound object and a harmonicity object was created, from which the mean harmonicity was calculated, hereafter the HNR. The details of its calculation can be found in Boersma [1993]. Harmonicity would seem to be a good measurement of voicing since vocal cord vibration produces “a complex periodic wave” [Johnson, 1997, p. 63]. Based on the discussion from Praat’s manual, higher values of HNR should correspond with higher voicing rates.

Results

Consider Figure 11, which shows the rates of [Cs] in regular plural forms in BP and L2 English.

The leftmost column shows that regular plural forms in BP, whose orthography is <Ces>, presented 62% of a consonant followed by a sibilant: [Cs]. That means that when a letter <e> appears in the orthography of BP plural forms, a vowel is manifested in 38% of the cases. The two rightmost columns report data from English spoken by Brazilian speakers. When the orthography in the plural form is <Ces>, a consonant followed by a sibilant [Cs] occurred in 83% of the cases, whereas in the cases where the orthography was <Cs>, a consonant followed by a sibilant occurred in 96% of the productions. This result shows that the pronunciation of [Cs] is more recurrent when the orthography is <Cs> than when the orthography is <Ces> in regular plural forms in English. In other words, a vowel will appear at higher rates when the orthographic pattern is <Ces> than when it is <Cs>. Thus, it is more likely that a plural form as tapes will

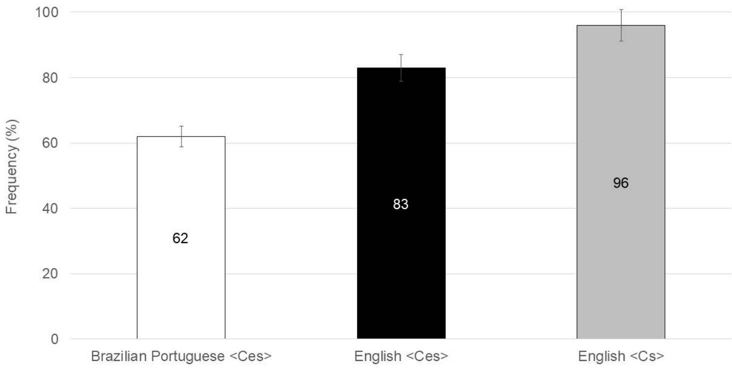


Figure 11: [Cs] rates by orthographic patterns.

have a vowel pronounced between the last two consonants than a plural form as maps. The difference between the data presented in the two rightmost columns is statistically significant for the orthographic patterns ($\chi 2 = 36.113$, df $= 1$, $p < 0.01$). The explanation for such difference lies in the different orthographic patterns.

We also considered whether different tasks could favor the production of an epenthetic vowel. According to Delatorre [2006] and Silveira [2007], visual input favors such non-target productions. In our experiment, the picture-counting task had no orthographic visual input, whereas orthography was available in the reading task. If Delatorre [2006] and Silveira [2007] are correct, then we expect that vowels would occur at higher rates in the reading task than in the picture-counting task in our experiment. However, no statistically significant differences were found between the picture-counting task and the reading task ($\chi 2 = 0.66$, df $= 1$, p-value $= 0.41$). This shows that it is the orthographic pattern rather the type of task that favors a vowel to occur in L2 English. Our claim is that once speakers are literate, orthography is part of their grammar, i.e., it has a permanent impact on mental representations. The EML2P model adopted in the current paper differs from Delatorre [2006], Silveira [2007] and Zanfra [2013] rule-based approach mainly by assuming that orthography is part of linguistic knowledge and not external to it.

Another research question we posited regarded the voice quality of the word-final sibilant in [Cs] and [Cis]. This was the main issue considered by Zanfra [2013], Fragozo [2017] within a rule-based approach. Their analysis claimed that voicing in L2 English did not achieve the target rates due to constraints of BP distribution of sibilants and regressive assimilation. BP only presents voiceless sibilants word-finally. However, across word-boundaries, BP sibilants are voiced when followed by a voiced consonant or a vowel: mês [mes] 'month', mês bonito [mez ˈbo.ni.tu] 'beautiful month', mês anterior [mez ə̃.te.ɾi.ˈoɾ] 'previous month'. In this paper, we offer an alternative view to the preceding rule-based approaches. Within the scope of the EML2P, it is suggested that generalizations from an ongoing sound change in BP phonology are transferred into L2 English, where phonetic detail plays an important role in shaping mental representations. Consider Figure 12.

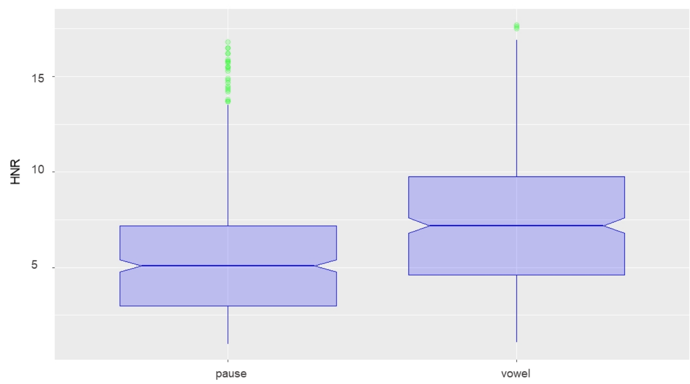


Figure 12: HNR per following phonetic environment in L2 English.

The boxplots in Figure 12 show harmonics-to-noise ratio per following phonetic environment in L2 English. We can see that when the sibilant is followed by a pause, it tends to be unvoiced, with HNR rates at around 5 decibels. Conversely, when the sibilant is followed by a vowel, voicing rates are higher. T-test results show that there is a significant difference in HNR between both following phonetic environments ($t = -8.8153$, $df = 821.37$, p-value $< 0{,}01$). However, even though such environments seem to influence voicing rates of the final sibilant, these rates are still lower when compared to English target forms. To put it another way, nouns that should be pronounced with a word-final voiced sibilant present more unexpected voiceless sibilants than voiced ones. This can be accounted by the fact that only voiceless sibilants occur word-finally in BP. Learners are likely unaware of the fact the [z] should be voiced in accordance with the voice property of the preceding segment. Now consider Figure 13.

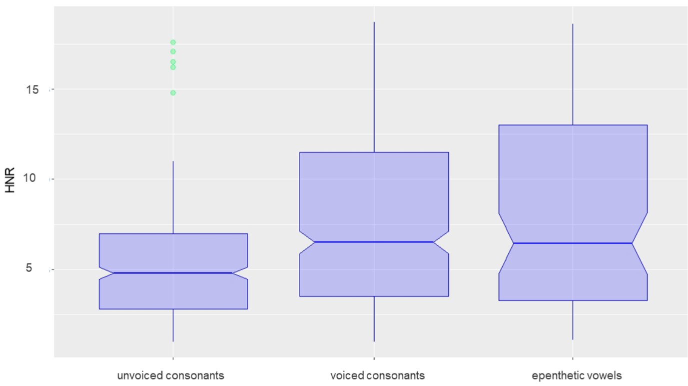


Figure 13: HNR per preceding phonetic environment in L2 English.

The boxplots in Figure 13 show harmonics-to-noise ratio per preceding phonetic environment in L2 English. Data is comprised of sibilants preceded either by an unvoiced consonant, a voiced consonant or an epenthetic vowel. At first sight, we can see the HNR rates are somewhat lower when the sibilant is preceded by an unvoiced consonant, and higher rates occur when the sibilant is followed by voiced consonants and epenthetic vowels. An analysis of variance (ANOVA) on these scores yielded significant variation among conditions: $F(2,858) = 59.99$, $p < 0.001$. A post-hoc Tukey test showed that the group comprised of unvoiced consonants differed significantly at $p < 0.05$; the voiced consonants group was not significantly different from the epenthetic vowels group. This result suggests that epenthetic vowels contribute to higher rates of voicing as much as other voiced segments in L2 English. Finally, an interaction between both preceding and following phonetic environments was attested [$F(2,858) = 7.797$, $p < 0.001$].

Our results throw some light on the line of research carried out by Zanfra [2013] and Fragozo [2017], who investigated the sibilant voicing followed by a vowel within rule-based approaches. We account for the fact that low HNR (which reflect voiceless sibilants) is recurrent in regular plural forms in L2 English, as [s] is the most robust exemplar in word-final position in BP. We also account for the fact that the pattern [Cis] favors a voiced sibilant in L2 English, as voiced sibilants are favored in similar contexts in BP (i.e., intervocalically). This indicates that L1 exemplar patterns, which reflect subphonemic information, are adopted in the L2. Finally, our analysis explains why [z] presents a low rate of production in L2 English spoken by Brazilian speakers: it is an emerging pattern in the L2, since it has no exemplars from the L1, at least not in word-final position. It will be through experience that these exemplars will become robust and more recurrent.

Conclusions

The aim of this paper was to investigate [Cs] clusters in English by Brazilian speakers. Its main contribution was to assess the role of orthography, phonological environment and task type not only on the production of epenthetic vowels, but also on the voicing property of the final sibilant. It also considered the role played by the [Cs] ~ [Cis] ongoing sound change from BP into L2 English. Results showed that the orthographic pattern <Ces> favors the production of an epenthetic vowel at higher rates that the <Cs> pattern. As for the task type, it was shown that it was not the visual access to orthographic forms that triggered a vowel to occur, but rather the orthographic patterns.

It was also shown that HNR is strongly influenced by phonetic/phonological environments, including preceding epenthetic vowels, which had not been accounted for in previous studies. We can assume that [z] poses a challenge to Brazilian speakers of L2 English due to the fact that it still has no exemplars from L1 in word-final position.

Concerning the role played by the BP ongoing sound change involving the [Cs] ~ [Cis] alternation, it was shown that robust patterns from the L1 are adopted in L2, including fine phonetic detail that reflects subphonemic properties. This sheds light to the fact that learners not only transfer sounds to the L2, but also phonological behaviors in which such sounds are subject to (as seen with how L2 voicing/HNR is influenced by L1 phonological patterns). We can assume, thus, that better generalizations are posited when the production of [Cs] clusters is assessed globally, rather than accounting for epenthesis and voicing agreement as separate, unrelated phenomena.

Bibliography

Paul Boersma. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In IFA Proceedings 17, pages 97–110, 1993.

Thais Cristófaro-Silva and Daniela Guimaraes. untitled. Paper presented at Seminário de Ciências da Fala, 9 2021.

Fernanda Delatorre. Brazilian efl learners production of vowel epenthesis in words ending in-ed. Master's thesis, Florianópolis, SC, 10 2006.

Carina Silva Fragozo. Aquisição de regras fonológicas do inglês por falantes de português brasileiro. PhD thesis, São Paulo, 12 2017. URL http://www.teses.usp.br/teses/disponiveis/8/8139/tde-21122017-124449/.

Silke Hamann and Ilaria E. Colombo. A formal account of the interaction of orthography and perception: English intervocalic consonants borrowed into Italian. Natural Language & Linguistic Theory, 35(3):683–714, August 2017. ISSN 0167-806X, 1573-0859. DOI: 10.1007/s11049-017-9362-3. URL http://link.springer.com/10.1007/s11049-017-9362-3.

Bruce Hayes. Introductory Phonology. 2011. ISBN 9781118315958 9781444360134. URL https://nbn-resolving.org/urn:nbn:de:101:1-201410212801. OCLC: 894709783.

Keith Johnson. Acoustic and auditory phonetics. Blackwell Publishers, Cambridge, Mass, 1997. ISBN 9780631200949 9780631200956.

Wellington Mendes Jr. and Thais Cristófaro-Silva. Plural formation in English: a Brazilian Portuguese case study. In Ubiratã Kickhöfel Alves and Jeniffer Imaregna Alcantara de Albuquerque, editors, Second Language Pronunciation: Different Approaches to Teaching and Training, volume 64 of Studies on Language Acquisition. De Gruyter Mouton, 2022 in press.

Yasaman Rafat. The interaction of acoustic and orthographic input in the acquisition of Spanish assibilated/fricative rhotics. Applied Psycholinguistics, 36(1):43–66, January 2015. ISSN 0142-7164, 1469-1817. DOI: 10.1017/S0142716414000423. URL https://www.cambridge.org/core/product/identifier/S0142716414000423/type/journal_article.

Rosane Silveira. O papel desempenhado pelo tipo de tarefa e pela ortografia na produção de consoantes em final de palavra. Revista de Estudos da Linguagem, 15(1), June 2007. DOI: 10.17851/2237-2083.15.1.147-180. URL https://doi.org/10.17851/2237-2083.15.1.147-180.

Victor Hugo Medina Soares. Encontros consonantais em final de palavra no português brasileiro. Master's thesis, Belo Horizonte, 2016.

Mayara Tsuchida Zanfra. Phonological context as a trigger of voicing change: a study on the production of english /s/ and /z/ in word-final position by brazilians. Master's thesis, Florianópolis, 12 2013.

Chao Zhou. L2 speech learning of European Portuguese /l/ and /flap/ by L1-Mandarin learners: Experimental evidence and theoretical modelling. Language Acquisition, pages 1–2, August 2021. ISSN 1048-9223, 1532-7817. DOI: 10.1080/10489223.2021.1952591. URL `https://www.tandfonline.com/doi/full/10.1080/10489223.2021.1952591`.

Radial Basis Function Artificial Neural Network for Automatic Identification of Interlanguage Transfer Phenomena

Atos Apollo Silva Borges[1], BrunoFerreira de Sousa[1], Aratuza Rodrigues Silva Rocha[2], Wilson Júnior de Araújo Carvalho[3], Fábio Rocha Barbosa[1], Ronaldo Mangueira Lima Júnior[4]

1 . Federal University of Piauí
2 . Faculdade Afonso Mafrense
3 . State University of Ceará
4 . Federal University of Ceará

Abstract

In the recent decades, especially for non-English speaking countries, the modern and more connected world has increased the urgency in learning a second language. Among the obstacles for beginners acquiring a new language are the grapho-phonic-phonological transfer phenomena between the two language systems, which may undermine their ability to communicate in the target language. The present work proposes a seed for an intelligent software designed to help language learners by providing automatic identification of transfer phenomena produced during their reading process.The algorithm is centered on a Radial Basis Function Artificial Neural Network (RBF-ANN) trained to automatically identify transfer processes between Brazilian Portuguese and English as ForeignLanguage. Five transfer processes already known in the literature were chosen to demonstrate the concept; however, as an initial approach, the audio samples used for training the algorithm weresynthetically generated by the Google Translate™ TTS system. To train the RBF-ANN algorithm we

used the f_0 mean and the mean of the first two Formant Frequencies as signal descriptors. The results presented a promising perspective for the development of a new computer-assisted pronunciation training software (CAPT) with accessible computational resources for Brazilian students and language institutes.

Introduction

During the learning of a new language, a process called interphonology is manifested. Interphonology is characterized as the creation of a linguistic system different from both the foreign language (L2) and native language (L1), but presenting characteristics from both languages simultaneously [Rocha, 2012]. The students in the processes of acquiring fluency on the second language transfer some of their knowledge of the L1 to the new language due to the already established structure of the L1. This phenomenon, which may be manifested during speech or oral reading, it is called grapho-phonic-phonological knowledge transfer [Zimmer and Alves, 2006]. The term grapho-phonic-phonological contemplates not only the transference of phonetic-phonological knowledge but also the transference of the grapheme-phoneme relationship of one language to the other, in the case of this work, Brazilian Portuguese (BP) as L1 to the English as Foreign Language (EFL), the L2. When the learner finds an unknown structure in the foreign language, it uses strategies to adapt L2 to a structure already known in L1.

Transfer phenomena between Portuguese as L1 and English L2 produced by Brazilian learners are well documented in the literature [Silveira et al., 2021]. However, the identification and classification of these processes are made mainly through transcriptions, a slow and laborious process done by specialized linguists. However, there is a shortage of works aimed at recognizing these processes in an automated way. Most studies carry this task as a general mispronunciation identification, comparing the input speech with a pre-recorded dataset of pronunciations, not taking advantage of the nature of each phenomenon. Most studies treat mispronunciation as a random process, not having any patter or regularity to be explored. Only two works were found proposing forms of automatic identification that take the nature of the phenomena as an important part of the recognition. The first was a categorization of BP speakers by a Self-Organizing Map (SOM) regarding the transfer of stress patterns between BP-L1 and English-L2 [Silva et al., 2011]. The second also aimed to identify transfer processes from BP to English-L2 of Brazilian students using a Multi-Layer Perceptron (MLP) neural network [Rocha, 2017]. The rapid identification of these phenomena would be of great value for software doing proficiency placement testsand could be used in language schools, distance education, computer-assisted pronunciation training (CAPT), researchers, and inclusion of neurodivergent people [Grund et al., 2020].

Therefore, this work proposes a seed for an intelligent software designed to help language learners by providing automatic identification of transfer phenomena produced during their reading process. The algorithm is centered on a Radial Basis Function Artificial Neural Network (RBF-ANN) trained to automatically identify transfer processes known in the literature of BP transfer to English-L2. The details of the algorithm are described on the RBF Neural Network section. Five transfer processes were chosen to demonstrate the concept and are described on the Acoustic data generation section; however, as an initial approach, the audio samples used for training the algorithm were synthetically generated by the Google Translate™ Text-To-Speech system. We assumed the hypothesis that even simple architectures of Artificial Neural Network, such as Radial Basis Function ANNs, are able to correctly identify the chosen transfer phenomena between Brazilian Portuguese-L1 to English-L2.

Acoustic data generation

The corpus of this study was constructed using the Corpus of Contemporary American English (COCA), an online and open-access corpus of English with a large variety of written and spoken words. Non-words were also incorporated to the study, all generated by the authors modifying existing words but still obeying English phonological patterns. As the pronunciations in this work should besynthetically generated, there were only two recordings for each word, one with the effects of the transfer phenomenon, as if pronounced by a Brazilian learner, and the other without it, as if pronounced by an English native speaker. A varied quantity of words must be used to be able to reach statistical significance. For this reason, a total of 508 words were used, generating a totalof 1016 recordings.

Five widely known transfer phenomena were chosen to be collected in the Google Translate™ TTS system. These phenomena are well documented and commonly found in the pronunciation of Brazilian beginning learners of English.

The first phenomenon investigated was the deletion of initial [h] in words beginning with <h> (henceforth, H-deletion), which corresponds to the deletion of the glottal fricative [h] at thebeginning of a word. As initial <h> has no corresponding sound in Portuguese, a

Brazilian learner might produce [i] and [u] in the beginning of ‘hilarious’ and ‘humorist’, respectively.

The second phenomenon was the deletion of initial [h] with a change of [aj] to [i] in words beginning with <hy> (henceforth, HY-i). As in the previous process, the deletion of [h] occurs due to the absence of a sound corresponding to the grapheme <h> in initial position in Portuguese, especially in cognate words such as ‘hyper’, ‘hydrant’ and ‘hydrogen’.

The third process chosen was only changing [aj] to [i] while keeping the pronunciation of initial [h] in words beginning with <hy> (henceforth, HY-hi). The HY-hi process goes in the opposite direction of the previous ones concerning the pronunciation of <h>. In H-deletion and HY-i processes, there is the deletion of initial [h], but in HY-hi the [h] is pronounced, with only a replacement of [aj] by [i], as described above.

The fourth process investigated is the pronunciation of silent <k> with the insertion an epenthetic [i] in words beginning with <kn> (henceforth, KN-kin). This transfer process is characterized by the pronunciation of [k] when <k> should be silent in words like ‘knife’ or ‘knickers’.

The last process investigated was the voicing of /s/ when <s> occurs between two vowels (henceforth, S-z). It is the pronunciation of voiced [z] when it should be voiceless [s]. The voicingoccurs in words like ‘basic’, ‘case’ or ‘fantasy’.

After the corpus selection, the speech collection took place on the Google Translate™ online platform. The Text-to-Speech (TTS) system embedded on the platform has the goal of generating a naturally-sounding speech waveform given a text to be synthesized. This process of mapping a sequence of discrete symbols (text) to a real-valued time series (waveform) is design to mimic the human speech production, emulating the periodic and aperiodic components present in human voice.

Recent researchers at Google have proposed the use of neural networks to perform the mapping between linguistic features and acoustic features [Tokuda and Zen, 2016, Zen et al., 2016]. In 2017, about 1/3 of all languages in Google’s TTS options already used Recurrent Neural Networks (RNN) as acoustic models and almost all options of languages in Android mobile devices already used RNN-based TTS systems [Zen, 2017]. The mapping of linguistic features to acoustic features using a parallel-distributed system is remarkably similar to the human reading process in the brain.

To collect the samples produced by Google Translate™, we used the open-source audio software Audacity (version 2.1.2). All the data in this research were collected in August of 2018. The productions were recorded at 44.1 kHz (standard) in Wave 32-bit float PCM. However, raw speech contains thousands of samples, which are often polluted with noise and unnecessary information. The solution is to convert the audio signal into a format with higher information density. To obtain these dense representations, we opted to use the PRAAT software (version 6.0.21).

To test different types of representation, we chose two descriptors: the mean of Formant Frequency (FF) and the mean of the Fundamental Frequency (f_0). The PRAAT software presents the oscillogram and spectrogram of audio files. This way, it is possible to select, in each word, the exact region where each researched phenomenon occurred. This specific region was selected, cut, and saved in Wave format, resulting in a file referring to the exclusive region of incidence of transfer processes. The objective was to extract both f_0 mean and the mean of F1 and F2 from the selected region. Although two different methods are used to obtain these values, the same audio file was used for both extractions.

RBF Neural Network

An artificial neural network is a system composed of ordered neurons in layers interconnected through synaptic weights. These synaptic weights ponder the connection between two neurons, or between an input and a neuron, assuming a higher value according to the influence of that connection to the output of the network. ANN has input nodes that receive stimuli from the external medium and output neurons that provide the network response. Usually, a layer between the input and output neurons is used, known as the hidden layer. The use of the hidden layer structure enables ANN to solve non-linearly separable problems.

A Radial Basis Function Artificial Neural Network is a three-layer feed-forward network that consists of one input layer, responsible for receiving the inputs, one middle layer, fully connectedto the input layer, and one output layer, also fully connected by weighted synapses and responsible for outputting the neural network prediction. Each input neuron corresponds to a component ofon input vector (in this case F1, F2 and f_0). The middle layer consists of N neurons and one bias neuron. Each middle neuron layer neuron computes a kernel function which is usuallythe Gaussian function [Hwang and Bang, 1997].

In order to specify the middle layer of an RBF-ANN we have to decide the number of neurons in the kernel layer. The simplest method is creating one neuron for each category present in the data.However, this method is not a good practice for most applications and can be suboptimal, especially when there is a large number of training patterns. Therefore, we used a K-means algorithm tocluster the training patterns in a reasonable number of groups. K-means is a kind of unsupervised clustering algorithm that search internal groups in the dataset, clustering the samples that belong to the same group enabling the adjustment of the centers and radius of the Gaussian kernels in the hidden layer [Chang et al., 2010].

Next, we use the standard statistical approach to calculate the weights between the middle layer and the output layer. The Least Mean Square Error (LMSE) procedure was used to determine the weights for each synaptic connection between the layers. This method finds the parameters of a linear function by the principle of Least Squares: minimizing the sum of the square of the differencesbetween the observed dependent variable and those predicted by the linear function of the independent variable. In our case, the independent variable is the vector of desired outputs, and the dependent variable is the vector of outputs of the hidden layer. The algorithm finds the linear relationship between these variables (hidden weights) using a nonlinear transformation.

As the neural networks are supervised algorithms, we manually classified the datasets and divided into a training subset (or memory subset) and a testing subset. The training subset is used as reference to the algorithm, presenting enough information about the behavior of the samples to allow for learning and generalization. With the training process completed, the neural network was tested with the testing subset. The samples of the training subset were never presented during the testing process or added in the reference data. This way we could test the accuracy and generalization levels of the model for new samples.

Results and Discussion

In summary, the results correspond to the average accuracy for each of the 50 iterations using randomized holdout for training, cross-validation and testing subsets. The average accuracy $\pm$ standard deviation obtained by the algorithm in each phenomenon is distributed in Table 7, presenting the performance in the test sets using both mean f_0 and the mean of the first two FF. We also displayed the optimal number of hidden neurons found by the K-means algorithm.

Results	Processes			
	H-deletion	HY-i/HY-hi	KN-kin	S-z
Accuracy	0.9203 ± 0.0234	0.9445 ± 0.0958	0.9441 ± 0.0368	0.9308 ± 0.0223
Kernels	6	2	2	2

Table 7: Accuracy obtained by the RBF-ANN in each phenomenon studied.

The results presented by the RBF-ANN algorithm were in general satisfactory for the identification goal. The algorithm obtained high levels of accuracy for all the phenomena with small variability on the results for the 50 iterations, providing a promising perspective for the development of a new computer-assisted pronunciation training software.

The number of kernels on the hidden layer found by K-means were expected for all process except for H-deletion. One explanation for the higher number of hidden units might be the existences of small clusters on the dataset. These clusters do not only separate the native-like and phenomenon samples, but also reveal internal structures inside the classes. Although these internal clusters are not directly being used to separate the classes globally, they can be used to enhance the accuracy of the decision boundary at the regions of superposition between the classes.

Conclusions

After the evidence presented by the results, a series of conclusions about the initial hypotheses could be drawn. The results indicated that RBF-ANN can identify the transfer processes producedby the TTS algorithm using the audio descriptor with high levels of accuracy and precision, providing ways to automatically identify the five processes with confidence. The algorithm can be trained with relatively small datasets, and it does not require huge computational power to be trained. These results provided a new perspective on the development of CAPT systems, demonstrating the advantages of using the already developed literature about the transfer phenomena to make the identification process more focused on the transfer patterns. This more efficient approach can be implemented on devices with limited processing power, such as mobile devices and online applications.

For future works, it is still necessary to expand the investigation with more phenomena and to acquire a greater number of samples for each process investigated. Expanding the number samples and testing new phenomena will provide new information for the development of a simple and efficient identification software. Further investigation can provide significant new information and ideas not only for software development but also about the phenomena themselves.

Bibliography

Gary W. Chang, Cheng-I Chen, and Yu-Feng Teng. Radial-basis-function-based neural network for harmonic detection. IEEE Transactions on Industrial Electronics, 57(6):2171–2179, 2010. DOI: 10.1109/TIE.2009.2034681.

Jonas Grund, Moritz Umfahrer, Lea Buchweitz, James Gay, Arthur Theil, and Oliver Korn. A gamified and adaptive learning system for neurodivergent workers in electronic assembling tasks. In Proceedings of the Conference on Mensch Und Computer, MuC '20, page 491–494, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450375405. DOI: 10.1145/3404983.3410420. URL `https://doi.org/10.1145/3404983.3410420`.

Young-Sup Hwang and Sung-Yang Bang. An efficient method to construct a radial basis function neural network classifier. Neural Networks, 10(8):1495–1503, 1997. ISSN 0893-6080. DOI: https://doi.org/10.1016/S0893-6080(97)00002-6. URL `https://www.sciencedirect.com/science/article/pii/S0893608097000026`.

Aratuza R. S. Rocha. Os efeitos da instrução explícita em fonologia na produção e percepção de consoantes da língua inglesa. Dissertation - Masters, Programa de Pós-Graduação em Linguística Aplicada, Universidade Estadual do Ceará, Fortaleza, Brazil, 2012. URL `http://www.uece.br/posla/dmdocuments/AratuzaRodriguesSilvaRocha.pdf`.

Aratuza Rodrigues Silva Rocha. Identificação de processos de transferência do português do Brasil para o Inglês (L2) por meio de rede neural artificial MLP. PhD, Programa de Pós-Graduação em Linguística Aplicada, Universidade Estadual do Ceará, Fortaleza, Brazil, 2017.

Ana Cristina C. Silva, Ana Cristina P. Macedo, and Guilherme A. Barreto. A SOM-based analysis of early prosodic acquisition of english by brazilian learners: Preliminary results. In Advances in

Self-Organizing Maps, pages 267–276. Springer Berlin Heidelberg, 2011.

R. Silveira, A. R. Gonçalves, F. Kupske, Ubiratã Kickhöfel Alves, and R. M. Lima Jr. Efeito da ortografia. In Investigando os sons de línguas não nativas: uma introdução. Editora da Abralin, 2021.

K. Tokuda and H. Zen. Directly modeling voiced and unvoiced components in speech waveforms by neural networks. In 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 5640–5644, March 2016. DOI: 10.1109/ICASSP.2016.7472757.

Heiga Zen. Generative Model-Based Text-to-Speech Synthesis, 2017. URL https://research.google.com/pubs/pub45882.html.

Heiga Zen, Yannis Agiomyrgiannakis, Niels Egberts, Fergus Henderson, and Przemysław Szczepaniak. Fast, Compact, and High Quality LSTM-RNN Based Statistical Parametric Speech Synthesizers for Mobile Devices. San Francisco, CA, USA, June 2016. URL https://research.google.com/pubs/pub45379.html.

Márcia Cristina Zimmer and Ubiratã Kickhöfel Alves. A produção de aspectos fonético-fonológicos da segunda língua: instrução explícita e conexionismo. Revista Linguagem & Ensino, 9(2):101–143, 2006. ISSN 1983-2400. URL http://www.rle.ucpel.tche.br/index.php/rle/article/view/168.

Part III: Methods for speech data collection, processing and analysis

> Fairy tales are more than true: not because they tell us that dragons exist, but because they tell us dragons can be beaten.
>
> C.K. Chesterton

A labeled dataset for analysing deviant orthographic forms in texts written by children

Adelaide H.P. Silva[1], Fabiano Silva[1], Wanderlan Carvalho[1], Lourenço Chacon[1]
1 . UFPR/DELLIN/LIAMF/CNPq

Introduction

According to Chacon and Pezarini [2018], "writing is a way of enunciating language, that is, a way of putting language to use in a given discursive situation, so that one can learn to write." Consequently, individuals have to understand the nature of the language's writing system, that is, its notational system, and also the functioning of the discursive aspects of writing, i.e., its social use. This is the same prediction made by official documents that offer guidelines to the process of teaching Portuguese Language, such as National Common Curriculum Base (BNCC).

We assume, like Chacon and Pezarini [2018], that orthography, taken as a notational system of a language, must have its operation understood by users so that they can learn to write and thus "using the language in a given discursive situation". In other words, learning the spelling of a language is like the first step for users of that language towards learning writing. Hence, the notational system of a language is addressed in the initial years of individuals' formal instruction. Despite being addressed in the early school years, the teaching of spelling is treated mainly, and according to Morais [2000], as a theme of systematic teaching verification, which results in a lack of planning about the spelling competence expected for students in each school grade. The lack of planning, in turn, is reflected in the absence of guidelines for teaching/learning spelling in the text of the Common National Curriculum Base (BNCC). This scenario causes 33.95% of students from the 5th grade in Elementary School to show an insufficient level in terms of spelling, according to data from the Ministry of Education, coming from the last National Literacy Assessment, in 2016.

For the above reasons, we started the elaboration of a system that learns the patterns of "errors" produced by children in Elementary School, in order to: 1) identify the most recurrent types of "errors" in written productions at each level of Elementary School; 2) help teachers to assess the child's performance in the spelling learning process, understanding the hypotheses that children build about the functioning of the language's notation system and identifying aspects to be worked on with the children to improve their performance; 3) help build guidelines for teaching spelling, so that a plan can be drawn up on which aspects to teach in the different grades of Elementary School. The last two points obviously depend on the first one, and are aspects to be developed in secondary or long term. In this work, we will address the identification of patterns of "errors" and the grouping of types of "errors" [Chacon and Pezarini, 2018] in the written productions of children from the third and fifth grades of Elementary School. The next section exposes the typology of "errors" by Chacon and Pezarini [2018] and the following sections deal with the functioning of the system.

A typology of spelling "errors"

In designing our system that analyzes non-standard spellings in texts written by children, the first step was to identify the "errors" and establish patterns for them. So, it was necessary to analyze an initial set of data in the light of parameters that would help us to recognize the "errors". To do so, we turn to Chacon and Pezarini [2018] who offer a typology of "errors" in children's written productions.

The authors propose that there is a gradiency in the relationship between the sounds and phonemes of Brazilian Portuguese and the orthographic system. This is a new approach in the literature, which differs from previous ones that predicted categories of "errors" (e.g., [Lemle, 1982, Cagliari, 1990, Morais, 2000]). Chacon and Pezarini [2018] propose that there are processes of omissions, transpositions and substitutions. In the first case, we would have, e.g., "pene" for "pente" (comb); in the second, "porfessor" for "professor" (teacher) and, in the third, "sebola" for "cebola" (onion). The authors explain that gradiency is revealed in the fact that, in omissions, there is no orthographic record of a sound, which puts them in a different situation from that of transpositions and substitutions, in which there is an orthographic record of the intended unit to be represented, although in transpositions the grapheme that represents a certain sound is not in the expected position and, in substitutions, the grapheme occupies the expected position, but is itself expected in the representation of a certain sound.

Still with regard to transpositions and substitutions, the authors argue that there is a gradiency within each fact. Thus, transpositions can involve permutations, when there is an exchange of position of two graphemes within a word (eg, “senera” for “serena”, meaning serene), or when there is an exchange of intersyllabic graphemes (eg, “drento” for “dentro” (inside), in which the change of position of a grapheme affects two syllables) or, even when there is a change of intrasyllabic graphemes and a grapheme moves from one position to another one in the same syllable (eg, “pregunta” for “pergunta” (question), in which the change position of a grapheme affects a single syllable). With regard to substitutions, Chacon and Pezarini [2018] note that they can be: hybrid substitutions, in cases where a grapheme is replaced by another one that can represent the same phoneme, in another context, such as “licid” for “líquido”, meaning liquid; non-phonological substitutions, in cases where a grapheme is used to represent a sound in a context where it was not anticipated, such as “rrato” for “rato”, i.e, mouse; phonological orthographic substitutions, in cases where the alteration of a grapheme changes what the authors call phonological value, as in “galo” (rooster) by “calo” (callus), in which the sonority of the consonant represented by the initial grapheme of the word is altered. It should be noted that the authors add that phonological orthographic substitutions behave differently from other substitutions and may involve sounds from the same class, as well as sounds from a different class. Thus, e.g., a grapheme substitution representing sound of a distinct class from the target grapheme can be found in “molacha” for “bolacha” (cookie). The substitutions of graphemes that represent sounds within the same class can be exemplified by the spelling “gola” (collar) by “cola” (glue).

Table 1 illustrates the types of “errors” predicted by Chacon and Pezarini [2018], as well as the subsets of “errors” within each type, with examples found in the texts of the dataset which we based our analysis on.

It is worth considering that the proposal by Chacon and Pezarini [2018] is a work in progress, so it does not include “errors” involving

type	subtype	example	reference
omission		redodo	redondo (round)
transposition	permutation	desromonou	desmoronou (fell apart)
	intersyllabic	despedirça	desperdiça (he wastes)
	intrasyllabic	quator	quatro (four)
substitution	hybrid	cachoro	cachorro (dog)
	non-phonological	fumasa	fumaça (smoke)
	phonological	futebou	futebol (soccer)

Table 1: Typology of "errors" from [Chacon and Pezarini, 2018].

vowels or "errors" related to prosodic domains, such as the phonological word. For this reason, when analyzing the dataset that we take as a starting point for our system, we established additional categories of "errors" that aim to deal with facts such as vowel harmony, e.g. "repitiu", for "repetiu" (he repeated) or the spelling of a phonological word as a single spelling word, e.g. "tabom", for "está bom" (ok, then), as well as the reverse, i.e. the spelling of a single phonological word as more than one spelling word, e.g. "com ver sar", for "conversar" (to talk). Therefore, we established the type "substitution", with subsets involving front and posterior vowels, as well as the subset "random" that comprises substitutions that involve vowels with no apparent motivation. We also established the type "others", that comprise phenomena related to word segmentation, motivated mainly, and apparently, by the notion of phonological word. Besides these types of "errors", we also established a type "insertion", that deals with the incorporation of an unexpected grapheme into a word, and a type "combined facts", that gathers errors from different types within the same word.

Table 2 shows the additional types of "errors" we refer to and illustrates them with examples found in the texts of the dataset.

type	subtype	example	reference
vowel substitution	front	eli	ele (he)
	porterior	pulicial	policial (policeman)
	random	uruba	urubu (vulture)
insertion		seisi	seis (six)
others		eamãe	e a mãe (and the mother)
combined facts		pasarros	pássaros (birds)

Table 2: Additional proposed types of "errors".

Dataset

The initial set of children's writing data for our system is the written language database assembled by the "Language Studies Research

Group" (GPEL). The database [Chacon and Pezarini, 2018] brings together texts produced especially for the constitution of the corpus by children from the first to the fifth year of elementary school in a public school in the city of Marília (SP).

The texts result from a task of retelling a narrative that was read in the classroom by the teacher to the students. This procedure was adopted to all the five school grades that were the target of the database. The initial objective of the UNESP colleagues was to collect four texts from each child in each grade, but this was not possible for all the children, because some of them missed class on the day of one or more collections. In addition, only the texts of children whose parents signed the Informed Consent Term were included in the database. In this initial stage of elaborating the system, we selected the texts of the third and fifth grades, which constitute the end, respectively, the first and second cycles of Elementary School and, therefore, involve the evaluation of children's productions. Due to this methodological decision, our dataset consists of 65 texts from 3rd grade students and 63 texts from 5th grade students, making a total of 128 texts, with 45561 words and 356 words per text, on average.

In addition to selecting a subset of texts for the constitution of our dataset, another methodological decision we made was to adopt orthographic words, and not texts, as the units of analysis. Orthographic words are strings delimited by blank spaces or punctuation marks. Our methodological decision is due to the fact that, as our focus is spelling, taking the word as a unit of analysis allows us to reach the syllable an the segment, two other linguistic units in which "errors" manifest themselves (still according to Chacon and Pezarini [2018]).

Once the units of analysis were defined, we made a manual classification of the "errors" in each text, following the types presented in section 2. With the typology of the "errors" of the dataset, we made a profiling of the data, using Pandas, an open Python library specifically for data analysis.

Profilling results and discussion

The profiling of the "errors" in texts of 3rd. grade (Figure 1) reveals that there is a prevalence of the combined facts type in the data. This type of "error" refers to occurrences of facts of a different nature in the same word. Thus, e.g., "enveconhados", exhibits 'r' omission in syllable coda and a phonological substitution that consists in replacing two segments within the same class, that is, 'g' by 'c'. Omissions are also quite frequent. The “others” type, as seen in the histogram, is one of the least frequent.

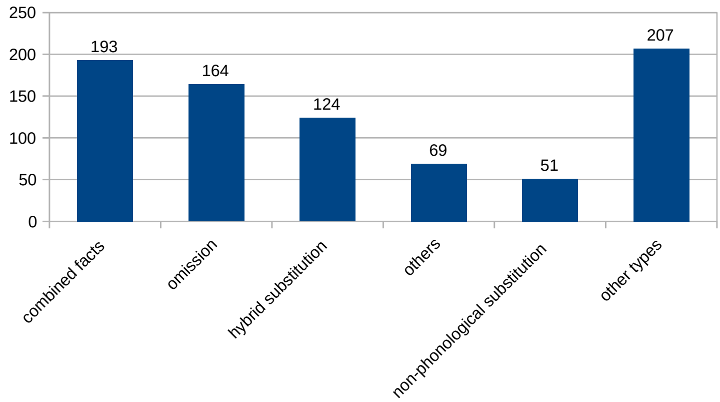


Figure 1: Number of occurrences by type for 3^{rd} grade data.

Checking now the distribution of “errors” in the texts of 5^{th} grade children, we notice a different distribution of data.

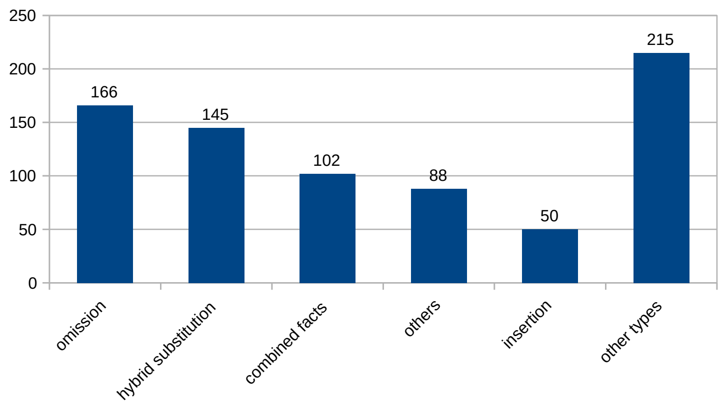


Figure 2: Number of occurrences by type for 5^{th} grade data.

Figure 2 shows a decrease in the frequency of occurrence of combined facts in the 5^{th} grade data, compared to the 3^{rd} grade. In the 3^{rd} grade, cases o combined facts account for 24% of the types of “errors”; in 5^{th} grade data, by 13% of them. On the other hand, the type omission remains very close in both series. An examination of data from the 5^{th} grade reveals that omissions occur mainly in the reduction of diphthongs of the 3^{rd} person, singular, in forms in past perfect tense (e.g., “assopro” instead of “assoprou”, he blew); in the deletion of the ‘r’ of verbal infinitive forms (e.g., “espirra”, instead of “espirrar”, to sneeze) or in forms of the verb to be, in which the entire first syllable is deleted, as in “tava”, instead of “estava” (he was). Faced with these cases, the cases of omission in the 3^{rd} grade texts follow another pattern: in them, there is deletion of the grapheme ‘n’, e.g., as a mark of the nasality of the vowel (as in “ligua”, instead of “língua”, tongue), or of the ‘r’ in syllable coda (as in “Macelo”, instead of “Marcelo”), of the ‘s’ in the same position (as in “contruir”, instead of “construir”, to build), or even more than one grapheme, as in “reportes”, instead of “repórteres”. These data suggest that, from the 3^{rd} to the 5^{th} grade, omissions are more localized and, to some extent, approximate the written forms to the spoken forms in a colloquial register. During their school time, children seem to learn that a letter is needed to mark the nasality of a vowel, for example. At the same time, they begin to register, in writing, syllabic structures that are more complex than CV. The others type, in turn, has a different path: it increases from the 3^{rd} to the 5^{th} grade. Interestingly, this kind of “error” is related to prosody – it gathers facts like “denovo”, instead of “de novo” (again) for example. Apparently, children start to build hypotheses about the segmentation of the speech chain more recurrently.

Bibliography

L. C. Cagliari. Alfabetização e lingüística. Ed. Scipione, São Paulo, 1990.

L. Chacon and I. O. Pezarini. Gradiência na correspondência fonema/grafema: uma proposta de caracterização do desempenho ortográfico infantil. In A. B. P. César, M. P. Seno, and Capellinim S. A., editors, Tópicos em Transtornos de Aprendizagem: Parte VI. Ed. Pulso, Ribeirão Preto, 2018.

M. Lemle. A tarefa da alfabetização-etapas e problemas no português. Letras de hoje, 17(4), 1982.

A. G. Morais. Ortografia: ensinar e aprender. Ed. Ática, São Paulo, 2000.

Behavioral and Neurophysiological Representations of Speech Phonemic Units

ADRIELLE C. SANTANA[1], ADRIANO V. BARBOSA[2], HANI C. YEHIA[3], RAFAEL LABOISSIÈRE[4]

1 . Universidade Federal de Ouro Preto

2 . Universidade Federal de Minas Gerais

3 . Université Grenoble Alpes

Introduction

Many studies showed that we perceive the world around us by categorizing the sensory input. This was studied, for example, for the perception of emotions [McCullough and Emmorey, 2009] and speech [Liberman et al., 1957].

For centuries, researchers around the world try to understand how our brain process speech and they try to propose a neurobiological model that explains the mechanisms that underlie the production-perception relationship. The dual-stream model is one example [Hickok and Poeppel, 2007]. But how such models relate or explain the categorical perception of speech? Many works tried to address this issue.

In the works of Alho et al. [2016], Chevillet et al. [2013] and Möttönen et al. [2014] the authors worked with a continua based on formant variations. In general, in those works the authors concluded about a sensorimotor integration of auditory and motor areas for early categorization which will occur around $120 - 170$ ms after stimulus onset. In Bouton et al. [2018], for a similar continuum, the authors identified the encoding of the second formant frequency around $95 - 120$ ms and again at 175 ms. In Bidelman et al. [2013] the authors synthesized an

/u/-/a/ continuum and observed categorical perception around 175 ms after stimulus onset. With this same continuum Bidelman and Walker [2017] concluded that the phonemic categorization was dependent of attention. However, in Chang et al. [2010] the authors identified the phonemic categorization around 110 ms in a task without attention (passive).

Based on those works we propose to perform the investigation of the neural correlates of categorical perception of speech sounds taking into account the attention and the acoustic cue influence as well the brain region measured. We also observed that the works reviewed selected the continuum a priori and did not performed any kind of dissociation of the physical (ϕ) and psychophysical (ψ) characteristics of the stimuli, so we took into account these issues in our study as well.

In order to perform this dissociation we considered the hypothesis illustrated in the Figure 3. It shows four stimuli in a phonemic continuum, in this example between the syllables $/da/$ and $/ta/$, differing in the Voice Onset Time (VOT). The physical values of VOT is represented in the horizontal axis. The vertical axis shows the probability of $/ta/$ responses in an identification task. The first (blue) and last (red) stimuli would be unambiguously identified as either $/da/$ or $/ta/$. However, the central stimuli (yellow and green), which are close to each other in terms of physical characteristics, would be represented rather distantly from each other in the psychophysical domain. Then, we hypothesize that is possible to identify two separate axes in the neurophysiological space: one related to the physical characteristics of the stimuli and another related to its psychophysical categorical perception.

Methodology

We performed electroencephalogram (EEG) acquisitions in eleven participants, right-handed and measured five signals from the electrodes difference: Cz–Tp9, Cz–Tp10, Cz–Fz, Cz–F7 and Cz–F8. The experiments were randomized across participants and each one performed both tasks (passive or active) with both continua (based on VOT variations or formants variations). For the acquisitions we selected five stimuli based on the psychometric curve of each subject for each continuum as represented in Figure 4 or a given participant.

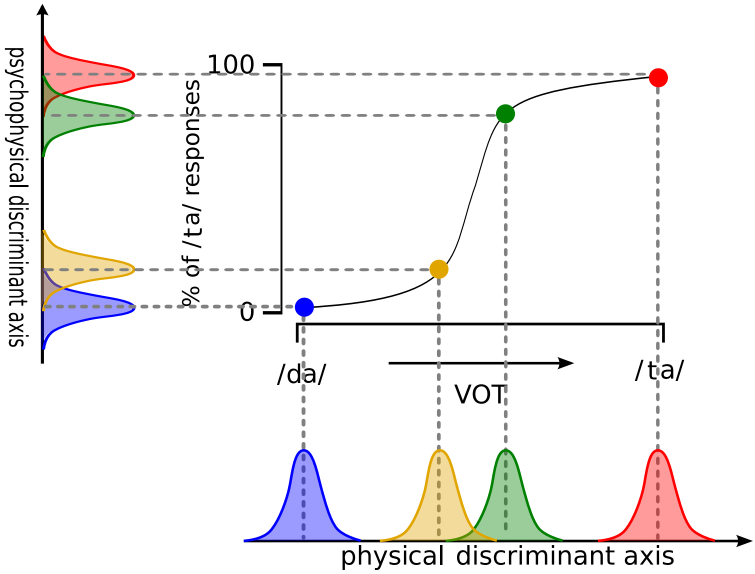


Figure 3: Physical and psychophysical (categorical) neurophysiological axes for the $/da/ - /ta/$ continuum. The physical values of VOT is represented in the horizontal axis. In the vertical axis, is represented the probability of $/ta/$ responses in an identification task.

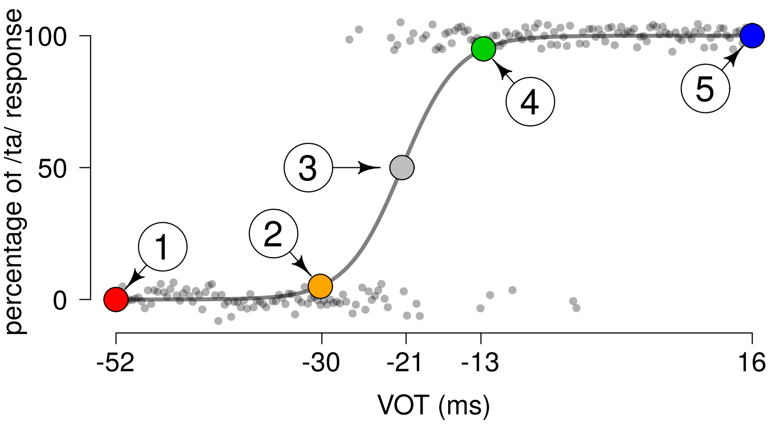


Figure 4: Psychometric curve for a given subject, for the VOT-continuum and the five stimuli selected

Time-domain processing

To evaluate how the brain oscillations are involved in the coding of the acoustic cues e to obtain the ϕ and ψ representations of the stimuli is interesting to work in the time-frequency domain.

For the processing we organized the data by electrode, participant, stimuli, continuum and task. The data was resampled for the execution of the Discrete Wavelet Transform (DWT) with decomposition in 9 levels. This way, the last levels presented a bands similar to those of the main brain oscillations ($\delta, \theta, \alpha, \beta$ and γ).

In general, we want to relate the behavior (observed in the psychometric curve) with the neural representation for each participant. However we arrived in a High Dimension Low Sample Size (HDLSS) problem with five observations and 800 wavelet coefficients (features). Then, we developed a regression technique to address this problem named Regression on Low-Dimension Spanned Input Space (RoLDSIS) [Santana et al., 2020]. Then we were able to obtain the ϕ and ψ neural discriminant axes which we evaluated in two ways: through the angle between them and through the Euclidean distance between them, which we called discrepancy.

With the angle we compared how it relate with the slope of the psychometric curve, which is reported as a measure of the categorical perception of the participant [Bidelman and Walker, 2017]. With the discrepancy analysis we worked with its value in different regions of the scalogram (graphical representation of the DWT), related to the N1 and P2 latencies and the main brain oscillations as illustrated in Figure 5. We obtained mixed-effects models for each region considering factors related to the continuum, task and electrodes, then we performed an ANOVA followed by a contrast analysis of the factors which presented significant effects.

Results

Time-domain results

Following it will be reported our main results for our time-domain analysis and some works that corroborate what we observed.

- We observed that there is a left hemisphere dominance for speech processing [Hickok and Poeppel, 2007, Boemio et al., 2005];
- A spectrotemporal analysis of the acoustic cue happens at the temporal region [Hickok and Poeppel, 2007];

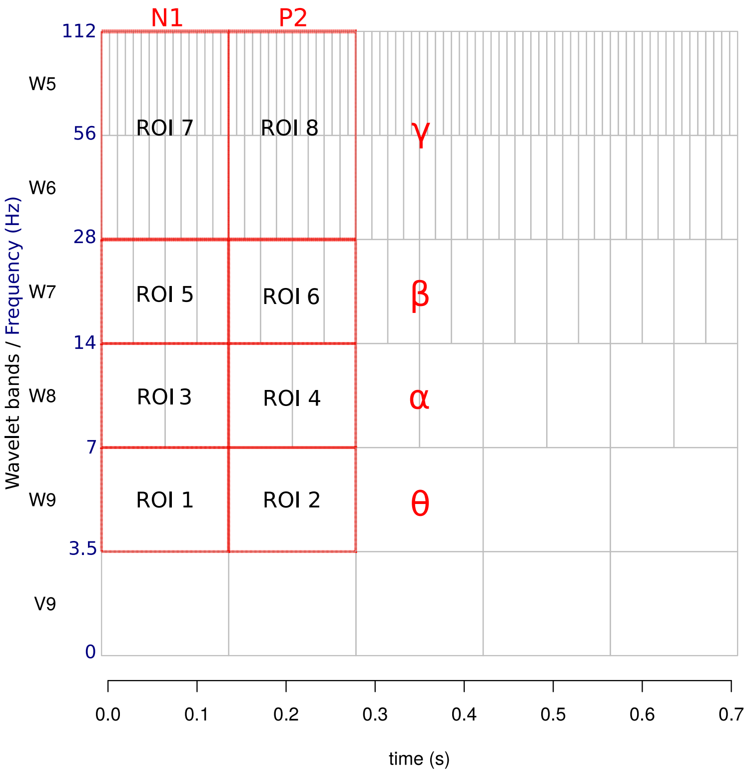


Figure 5: Definition of the ROIs as dependent variables for the models. Each ROI corresponds to specific a frequency band and specific a time interval.

- Generators of N1 and P2 are more laterally localized;
- Formants and VOT evoke different behaviors in N1 and P2 generators;
- N1 is sensitive to VOT variations [Steinschneider et al., 1995, Eggermont, 1995];
- Stimuli are processed differently when there is attention to the task and attention influences the speed of stimuli processing [Möttönen et al., 2014, Alho et al., 2016];
- Ambiguity is reflected into ERP amplitudes [Bidelman and Walker, 2017]
- P2 seems to code ambiguity or effort for speech perception [Rao et al., 2010];
- Attention influences the generators recruited to process stimuli [Hillyard et al., 1973];
- Attention influences more left hemisphere generators than right ones.

Time-frequency domain results

Following it will be reported our main results for our time-frequency domain analysis.

- Participants which categorize better have larger difference in internal ϕ and ψ neural representations of acoustic cues. The Figure 6 illustrates the positive and significant correlation ($r = 0.788$) between the axes' angles and the slopes of the psychometric curve for the formant continuum with the active task;
- It was observed a more lateral location of the speech structures;
- γ activity was observed in all scalp regions measured and this is probably related to integration/synchronization of these regions;
- Enhancement of α activity with attention;
- Larger discrepancies associated with the temporal region than the frontal or medial one;
- Brain oscillations involved in high level speech processing present strong activity as early as the N1 time frame.

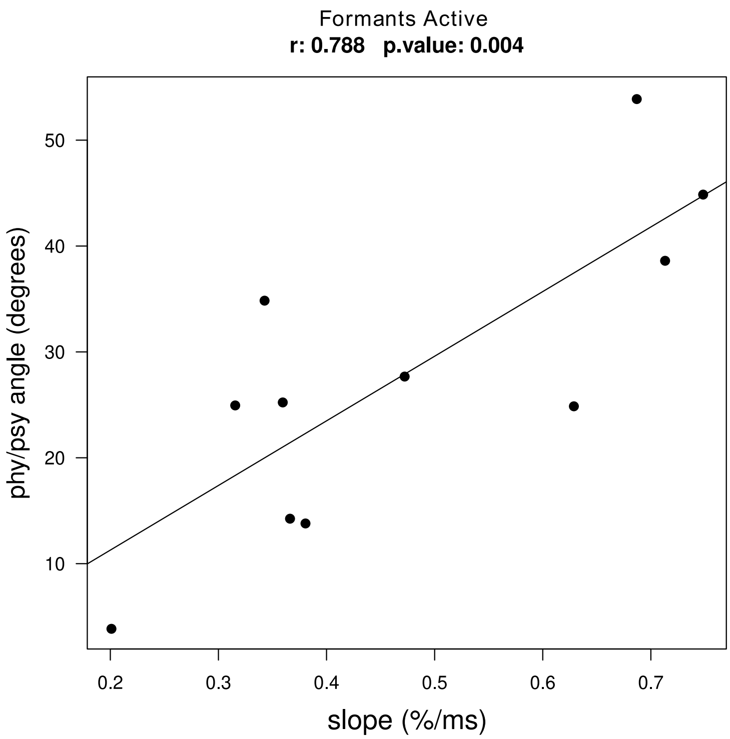


Figure 6: Relationship between the slope of the psychometric curve and the angle between the neurophysiological axes. In this population scatter plot, each point represents a participant. The horizontal and vertical axes represent, respectively, the slope of the fitted psychometric curve at 50% and the angle between the physical and the psychophysical directions obtained by the RoLDSIS procedure. The black line corresponds to the correlation line.

Conclusions

In this work we investigated the neural correlates of categorical perception of speech sounds, specifically of Brazilian Portuguese phonemes, evaluating ERPs in the scope of the stimulus acoustic characteristic (VOT and formant frequencies). We studied the brain cortical regions involved in speech perception (temporal and frontal), manipulating the degree of attention to the identification task and using data acquired with the use of a non-invasive method. In our analysis, we propose to identify the physical and psychophysical responses in the ERP, in order to show how the modulations in the time and frequency characteristics of the ERP can be related to the phonemic categorical perception (CP).

We saw that each frequency band and latency seems to code different aspects of the sound for the speech processing. It was observed that participants who presented behaviorally stronger CP had a larger difference between their physical and psychophysical neural representation of the stimuli. This difference was pronounced for the VOT acoustic cue than for the formants and for active tasks than for the passive ones. It was also shown that the CP occurs when there is no attention to the auditory task but only for the formant-based acoustic cue. Hemispheric differences were observed, with stronger activity at the left hemisphere. Differences were also observed between frontal and temporal cortical regions coded by low-frequency rhythms with more activity at the temporal region. In the gamma band we observed no significant difference between the activity at the frontal and temporal regions.

Our results also showed that temporal region structures may also perform some categorization besides the processing of physical acoustic characteristics of the sounds. We also show how the acoustic cue and task dynamically reconfigure the speech network which should be took into account by a neurobiological model for speech perception.

This study compared different factors related to categorical speech perception in Brazilian Portuguese using a reproducible protocol developed for the study and the evaluation of phonemic categorical perception, and confirmed many of the results found in the literature for other languages.

Data and Materials

The data that support the findings of this study, as well as the scripts four reproducing the results and the thesis, are available in the following repositories:

- `https://github.com/Adrielle-Santana/ThesisScripts`

- `https://github.com/RoLDSIS/code`
- `http://hdl.handle.net/1843/35151`

Bibliography

Jussi Alho, Brannon M Green, Patrick JC May, Mikko Sams, Hannu Tiitinen, Josef P Rauschecker, and Iiro P Jääskeläinen. Early-latency categorical speech sound representations in the left inferior frontal gyrus. Neuroimage, 129:214–223, 2016.

Gavin M Bidelman and Breya S Walker. Attentional modulation and domain-specificity underlying the neural organization of auditory categorical perception. European Journal of Neuroscience, 45(5): 690–699, 2017.

Gavin M. Bidelman, Sylvain Moreno, and Claude Alain. Tracing the emergence of categorical speech perception in the human auditory system. NeuroImage, 79:201–212, 2013. ISSN 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2013.04.093.

Anthony Boemio, Stephen Fromm, Allen Braun, and David Poeppel. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nature neuroscience, 8(3):389, 2005.

Sophie Bouton, Valérian Chambon, Rémi Tyrand, Adrian G. Guggisberg, Margitta Seeck, Sami Karkar, Dimitri van de Ville, and Anne-Lise Giraud. Focal versus distributed temporal cortex activity for speech sound category assignment. Proceedings of the National Academy of Sciences, 115(6):E1299–E1308, 2018. ISSN 0027-8424. DOI: 10.1073/pnas.1714279115. URL `http://www.pnas.org/content/115/6/E1299`.

Edward F Chang, Jochem W Rieger, Keith Johnson, Mitchel S Berger, Nicholas M Barbaro, and Robert T Knight. Categorical speech representation in human superior temporal gyrus. Nature neuroscience, 13 (11):1428, 2010. DOI: 10.1038/nn.2641.

Mark A Chevillet, Xiong Jiang, Josef P Rauschecker, and Maximilian Riesenhuber. Automatic phoneme category selectivity in the dorsal auditory stream. Journal of Neuroscience, 33(12):5208–5215, 2013.

Jos J Eggermont. Representation of a voice onset time continuum in primary auditory cortex of the cat. The Journal of the Acoustical Society of America, 98(2):911–920, 1995.

Gregory Hickok and David Poeppel. The cortical organization of speech processing. Nature reviews neuroscience, 8(5):393, 2007.

Steven A Hillyard, Robert F Hink, Vincent L Schwent, and Terence W Picton. Electrical signs of selective attention in the human brain. Science, 182(4108):177–180, 1973.

Alvin M Liberman, Katherine Safford Harris, Howard S Hoffman, and Belver C Griffith. The discrimination of speech sounds within and across phoneme boundaries. Journal of experimental psychology, 54 (5):358, 1957.

Stephen McCullough and Karen Emmorey. Categorical perception of affective and linguistic facial expressions. Cognition, 110(2): 208–221, 2009.

Riikka Möttönen, Gido M van de Ven, and Kate E Watkins. Attention fine-tunes auditory–motor processing of speech sounds. Journal of Neuroscience, 34(11):4064–4069, 2014.

Aparna Rao, Yang Zhang, and Sharon Miller. Selective listening of concurrent auditory stimuli: an event-related potential study. Hearing research, 268(1-2):123–132, 2010.

Adrielle C. Santana, Adriano V. Barbosa, Yehia Hani C, and Rafael Laboissière. A dimension reduction technique applied to regression on high dimension, low sample size neurophysiological data sets. BMC Neuroscience, 2020. (in press).

Mitchell Steinschneider, Charles E Schroeder, Joseph C Arezzo, and Herbert G Vaughan. Physiologic correlates of the voice onset time boundary in primary auditory cortex (a1) of the awake monkey: temporal response patterns. Brain and language, 48(3):326–340, 1995.

A generic representation for orthographic structure in texts written by children

ADELAIDE H.P. SILVA[1], FABIANO SILVA[1], WANDERLAN CARVALHO[1], LOURENÇO CHACON[1]
1 . UFPR/DELLIN/LIAMF/CNPq

Introduction

Even though spelling is taken as a parameter to evaluate whether an individual is literate or not, Brazilian students from 3rd and 5th grade of Elementary School achieve a low performance in tests that verify their knowledge of the Brazilian Portuguese spelling. The last edition of the National Literacy Exam, held in 2016, points that 34% of the children evaluated did not achieve the expected scores for literate students. So, we designed a software system, Scriba, that generates orthographic forms such as the ones found in texts written by 3rd and 5th grade children. We expect Scriba to offer subsidies for teachers to understand the hypotheses underlying the forms that deviate from standard orthographic rules. We also expect that Scriba can provide teachers resources to evaluate whether or not the "errors" fit a given grade.

For the system to learn how to produce deviant forms and simulate the "errors", the first step was the analysis of data to classify "errors" and to detect patterns that group them following criteria such as type of "error" and school grade. Concomitantly to grouping and classifying the "errors", we propose a generic representation for orthographic structure in texts written by children, so that Scriba can help understanding the hypotheses that lead to deviations.

Both tasks – "errors" classification and the elaboration of a representation for orthographic structure – are based on a corpus of 168 texts, containing 45561 words. We assumed the orthographic words as the unities of analysis because they are the domain in which the spelling deviations occur. Furthermore, and considering the findings of Chacon and Pezarini [2018], we focus on the syllable, within the orthographic word. This is because the syllable allows us to observe relationships between graphemes, as well as to make predictions about these relationships and the sounds they annotate. In the following sections we present and discuss the representation we elaborated to annotate the internal structure of orthographic words.

Criteria for making the labels

Predictions are at the heart of the algorithm for Scriba. In order to be able to predict the "errors", their nature, the graphemes involved in them and if they occur within the limits of a syllable, or extrapolate them, we proposed a set of labels, still as part of the initial process of formal representation of existing elements in a spelling word. The tags include: syllabic bondaries inside spelling words; internal structure of syllables; placement of primary stress in words; stress degree; consonant class. Table 3 presents the set of the labels.

Variable	Orthographic representation	Labels
Plosive consonants	p, b, t, d, c, qu, g, gu	O
Fricative consonants	f, v, s, ss, c, x, z, ch, j, g	F
Nasal consonants	m, n, nh	N
Liquid consonants	l, lh, r, rr	L
Vowels	i, e, a, o, u, ẽ, ã, õ	V
Onset	O, F, N, L	SA
Nucleus	V	SN
Coda	p, t, d, c, g, f, s, z, m, n, l, r	SC
First unit in complex onset	p, b, t, d, c, g, f, v	CA1
Second unit in complex onset	l, r, s, m, n	CA2
First unit in complex nucleus	i, e, a, o, u, ã, õ	CN1
Second unit in complex nucleus	i, u, e, o	CN2
First unit in complex coda	n, r	CC1
Second unit in complex coda	s	CC2
Stressed syllable		3
Pre and post-tonic syllable		1
Post-tonic final syllable		0

Table 3: Spelling word representation labels.

As can be seen in Table 3, each variable is assigned a label. Here, labels attached to variables are based on Brazilian Portuguese, and that's the reason why some of them may seem weird at first sight. The variables cover: the type of segments, such as vowels and consonants and, within the set of consonants, subsets were established following the parameter manner of articulation. The set of vowels gather oral and nasal sounds. Within each subset, or type of segment, consonants and vowels are associated to the possible graphemes that annotate them.

Variables also take into account the position of each unit within the syllable. Following the phonotactic constraints in Brazilian Portuguese, as presented, e.g., by Collischonn [1996], we assume that only vowels can ocuppy syllable nucleus. We also assume that the subset of consonants that occur in syllable codas is smaller than the subset of consonnats that occur in syllable onset. Notice that there is also the prediction on the units that can occur in the second position in complex syllabic constituents and we assign a numerical index for the consonant to indicate whether it is the first or the second unit in a complex constituent. By doing so, we offer an internal representation for the syllables in Brazilian Portuguese (BP).

The prosodic structure of the word is captured by the labels by assigning stress levels to the syllables, such as predicted by Camara Jr. [1970]. Thus, numerical index 3 is attributed to the syllable that carries the primary stress in the word. Numerical index 1 is attributed to pretonic and postonic syllables. In the case of the postonic syllables, index 1 applies only in cases where the syllable is not the last one in the word, i.e., in syllables that immediately follow the stressed one in proparoxytone words. Numerical index 0 is attributed to postonic syllables that are also the last ones in the word. Numerical index 2, in turn, is assigned to secondary stress, i.e,. in cases where the primary stress of the basis turns into the secondary one by means of morphological operations, such as derivations, e.g., "café" - "cafezinho" (coffee and its corresponding diminutive form). In "cafezinho", the suffix "-inho" is stressed, and carries the primary stress of the form, because the most prominence domain lays at the rightmost position in BP. As a consequence, the syllable "fe" loses intensity and turns out to be the second more prominent syllable in the word, thus carrying the secondary stress.

It is worth adding that the labels allow us to better capture and understand how the units relate to each other and predict possible sequences, as well as the sequences of units that violate constraints of well-formedness, such as sequences of graphemes that write sound sequences that do not obey the sonority scale, and also sequences of graphemes that annotate randomized sequences of consonants, with no intervenient vowels. Thus, e.g., the labels allow us to predict that

BP has a syllable such as “por”, in words like “porta” (door), but has no syllable with “rt”. This prediction has an obvious consequence for machine learning tasks: BP native speakers know the internal structure of the syllables in the language, so they perform quite well in establishing syllable boundaries. But the same is not true for a machine, that does not know the internal structure of BP syllables. So, considering that the syllable is the domain where “errors” apply, it is necessary to teach the machine the internal structure of syllables in the language. As the labels are machine-readable, the system can learn the internal structure of syllables.

It is worth mentioning that the labels do not take into account the number of the syllalbles within the word, because Chacon and Pezarini [2018] did not observe possible influences of this variable on children’s spelling “errors”. But the information can be added to the labels, if it’s necessary some time, since we plan to put aditional data into the corpus.

Application of the labels to words

Table 4 displays a set of examples of labeled data from the corpus of written texts by children. Data are organized according to number of syllables within the word and primary stress placement. In Table 4, segments boundaries are indicated by parentheses and syllables boundaries are indicated by brackets. This strategy for indicating boundaries was adopted in the light of the considerations presented in section 2 concerning machine learning and the need to teach the system which syllable structures are possible in BP.

Size	Example	Labels for different consonant, types and stress levels
1	mau	[(SAN)(CN1)(CN2)]3
1	sai	[(SAF)(CN1)(CN2)]3
2 ox	senhor	[(SAF)(SN)]1[(SAN)(SN)(SCL)]3
2 ox	inflei	[(SN)(SCN)]1[(CA1F)(CA2L)(CN1)(CN2)]3
2 par	porco	[(SAO)(SN)(SCL)]3[(SAO)(SN)]0
2 par	crânio	[(CA1O)(CA2L)(SN)]3[(SAN)(CN1)(CN2)]0
3 ox	arrombar	[(SN)]1[(SAL)(SN)(SCN)]1[(SAO)(SN)(SCL)]3
3 ox	derrubei	[(SAO)(SN)]1[(SAL)(SN)]1[(SAO)(CN1)(CN2)]3
3 par	bochecha	[(SAO)(SN)]1[(SAF)(SN)]3[(SAF)(SN)]0
3 par	açúcar	[(SN)]1[(SAF)(SN)]3[(SAO)(SN)(SCL)]0
3 prop	xícara	[(SAF)(SN)]3[(SAO)(SN)]1[(SAL)(SN)]0
3 prop	vítima	[(SAF)(SN)]3[(SAO)(SN)]1[(SAN)(SN)]0
4+ par	vovozinha	[(SAF)(SN)]1[(SAF)(SN)]2[(SAF)(SN)]3[(SAN)(SN)]0
4+ par	aniversário	[(SN)]1[(SAN)(SN)]1[(SAF)(SV)(SCL)]1[(SAF)(SN)]3[(SOL)(CN1)(CN2)]0

Table 4: Labeling examples for words in the dataset.

The numerical stress indices (0, 1, 2, 3) follow the proposal of Camara Jr. [1970]. ‘S’, at the beginning of the label sequence, within a syllable, marks Onset or Simple Nucleus, while ‘C’, preceding these tags, marks complex syllabic constituents. In the case of Complex Codas, we have a CC sequence. In the case of complex syllabic constituents, the proposed tag system makes it possible to identify whether a consonant is the first or second within the sequence. For diphthongs, we follow Collischonn [1996] proposal, which predicts the existence of a complex syllabic nucleus in BP. To clarify the labels interpretation, take the notation for the word “vovozinha”, a paroxytone with four syllables, in:

[(SAF)(SN)]1[(SAF)(SN)]2[(SAF)(SN)]3[(SAN)(SN)]0

from left to right: the first syllable has two units, being the first one a fricative consonant F occuring in a simple S onset A. It is followed by a simple S nucleus N and has stress level 1, i.e., it is pretonic. The second syllable has also two units. Like in the preceding syllable, the first unit is a fricative consonant, followed by a vowel. But the second syllable is assigned stress level 2, because it carries the former information on the primary stress placement in the word “vovó” before a derivational process applied. The third syllable has also two units, out of which the first one is a fricative consonant and the second one, a vowel. This is the stressed syllable in the word, and the primary stress is informed by the stress level 3. Finally, the last syllable has a nasal N consonant in simple S onset A, followed by a single vowel in the nucleus. The last syllable is assigned level 0, since it is the postonic one and, consequently, the less prominent syllable within the word.

Final remarks

The labels we propose here are machine readable and correspond to some abstraction departing from real data. As a consequence, they can provide teachers a way to understand the hypotheses children formulate when they make spelling “errors”. And the machine can learn to simulate such errors. Nevertheless, the labels do not specify which vowel occurs in syllable nucleus. According to Table 3, seven oral vowels and five nasal ones can be the nucleus of syllables in BP. There are “errors” that encompass vowel quality, such as “rechunchuda” (for “rechonchuda”, chubby) or “ispludiu” (for “explodiu”, it exploded). One possible way to solve the problem is to propose additional labels that deal with different opening degrees of aperture, because this is the parameter related to jaw opening that differentiates between /ɔ/, as in "vovó" (grandma) and /o/, as in "vovô" (grandpa). Besides that, vowels also differentiate according to the place of the vocal tract they are articulated in: the palate, the velum or an intermediate point.

To accommodate these information in our labels, we propose four degrees of aperture, expressed by numerical indices that follow the labels 'fr', i.e., frontal, for vowels articulated in the hard palate; 'ct', or central, for vowels articulated in an intermediate place between the hard palate and the velum, and 'pt', or posterior, for vowels articulated in the velum. Then, we can rewrite the "vovozinha" example as in:

[(SAF)(SNpt2)]1[(SAF)(SNpt3)]2[(SAF)(SNfr1)]3[(SAN)(SNct)]0

By doing so, we can properly differentiate between the vowels in the first and second syllables (from left to right) in the word. The next step, then, is make the system learn the labels and, departing from them, learn to produce non deviant orthographic forms for a set of words, as well as the deviant corresponding ones.

Bibliography

J. M. Camara Jr. Estrutura da língua portuguesa. Ed. Vozes, Petrópolis, 1970.

L. Chacon and I. O. Pezarini. Gradiência na correspondência fonema/grafema: uma proposta de caracterização do desempenho ortográfico infantil. In A. B. P. César, M. P. Seno, and Capellinim S. A., editors, Tópicos em Transtornos de Aprendizagem: Parte VI. Ed. Pulso, Ribeirão Preto, 2018.

G. Collischonn. A sílaba em português. In L. Bisol, editor, Introdução a Estudos de Fonologia do Português Brasileiro. EDIPUCRS, Porto Alegre, 1996.

Part IV:

Brazilian Portuguese Phonetics and Phonology

Fairy tales are more than true: not because they tell us that dragons exist, but because they tell us dragons can be beaten.

C.K. Chesterton

The implementation of phonic voicing contrast in children's speech: some explorations of clinical data

Fabiana Nogueira Gregio[1], Zuleica Camargo[1]
1 . Pontifícia Universidade Católica de São Paulo

Abstract

The objective was to investigate the implementation strategies of phonic voicing contrast in Brazilian Portuguese in a group of children with speech disorders in comparison to a control group. From the production of target words with unvoiced and voiced plosive sounds, in contexts of tonic and post-tonic syllables, a set of acoustic measures was extracted and analyzed, a perception experiment was applied, and the acoustic and auditory spheres were explored by means of statistical analysis. The investigation showed that the subjects performed intermediate productions towards the determinant characteristics of the voicing contrast. More than one acoustic cue was implemented for auditory judgment of the voicing contrast.

Introduction

Speech disorders trigger continual investigations into the refined mechanisms of speech. Among the complaints of speech disorders, the speech therapy clinical setting is faced with the demand to care for cases of "absence/exchange of voiced sounds" [Keske-Soares et al., 2004, Mota et al., 2012]. In traditional phonological views, such disorder is regarded as the absence or substitution of phonemes or distinctive features. However, clinical evaluations supported by acoustic-phonetic analyses have shown the presence of acoustic cues indicative of the realization of the sound considered absent.

Studies suggest that speakers mark a phonic distinction potentially perceived by them and revealed through intermediate rather than categorical productions, yet not always identified by the listener [Levy, 1993, Ficker, 2003, Gregio and Camargo, 2005, Rodrigues et al., 2008, Gregio et al., 2011, Pereira, 2012].

These productions can be investigated if enlightened by speech production and perception theoretical models that contemplate the dynamic and gradient character of speech, fundamentally relying on the use of speech analysis instruments for explanation [Silva et al., 2001, Gregio et al., 2011, Albano, 2007, Silva, 2010].

Regarding the plosive obstruent consonants of Brazilian Portuguese (BP), the unvoiced-voiced pairs [p]-[b], [t]-[d], and [k]-[g] constitute the repertoire of sounds, characterized by the respective places of articulation bilabial, dental, and velar [Silva et al., 2001, Camargo and Navas, 2008].

Physiologically, the voicing production in plosives involves fine motor coordination of glottic (vocal fold vibration) and supraglottic (vocal tract obstruction) movements [Sweeting and Baken, 1982, Shimizu, 1996, Gregio and Camargo, 2005]. Integrity of lung volume, aerodynamic conditions of the glottis, laryngeal muscles, phono-articulatory organs, and auditory system are required [Hoit et al., 1993, Shimizu, 1996, Hoole et al., 1999].

Acoustically, the production of unvoiced plosives involves the generation of a transient noise source, the acoustic result of complete constriction at some point in the vocal tract followed by its release. In voiced plosives, there are two sound sources: the transient noise coupled with the voice source resulting from the vibration of the vocal folds [Kent and Read, 1992, Johnson, 2003].

One of the acoustic measures used and studied in the investigation of voicing contrasts is voice-onset-time (VOT) [Lisker and Abramson, 1964, Behlau, 1986, Kent and Read, 1992, Levy, 1993, Shimizu, 1996, Cho and Ladefoged, 1999, Camargo et al., 2000, Rocca, 2003]. Other acoustic measures have been reported in voicing contrast studies, such as consonant duration, duration of vowels adjacent to the consonant, fundamental frequency (f0) at the beginning of the vowel following the consonant, frequency of the first formant (F1) at the beginning of the vowel following the consonant, and burst [Barton and Macken, 1980, Shimizu, 1996, Veloso, 1997, van Alphen and Smits, 2004, Benkí, 2005, Lousada et al., 2005, Barroco et al., 2007, Whalen et al., 2007, ?, Hanson, 2009, Tachibana et al., 2012].

In the BP literature, especially in the clinical context, VOT is still highlighted in the voicing contrast. Studies on this issue in various speech situations suggest that more than one acoustic cue seems to be involved [Behlau, 1986, Levy, 1993, Barbosa, 1996, Madureira et al., Ficker, 2003, Gregio and Camargo, 2005, Gurgueira, 2006, Barzaghi et al., 2007, Bonatto, 2007, de Oliveira e Britto, 2010, Gregio et al., 2011, Schliemann, 2011, Souza et al., 2010, Berti et al., 2012, Melo et al., 2012, Pereira, 2012].

Thus, the objective of this study was to investigate the implementation strategies of phonic voicing contrast in BP in a group of children with speech disorders in comparison to a control group.

Methods

Six subjects aged between 7 and 10 years old participated in this study, two females and four males, of whom three had a diagnosis of speech disorders related to voicing contrast (studied group) and three had no speech disorders (control group).

The selected subjects presented audiological evaluation with normal hearing thresholds, had a negative history of neurological, voice disorders and other speech disorders unrelated to voicing contrast, and were BP speakers with no reference to bilingualism.

The age group did not include the voice change phase, which could have affected the voice source as a result of physiological changes in the vocal tract [Behlau, 1995]. Within the established age limits, differences in laryngeal acoustic measures were not significant between male and female children [Andrade, 2009].

The subjects participated in the speech production data collection, carried out in a speech laboratory. The corpus consisted of recordings of the subjects reading, with five repetitions in random order, sentences following the syllabic structure C1V1C2V2 (consonant1-vowel1-consonant2-vowel2). The target words contained the unvoiced and voiced plosive sounds of BP (papa, baba, tata, dada, caca, gaga), inserted in the carrier-sentence “Diga ____ baixinho”.

The subjects presented different speech productions regarding the stress of the target word. Despite the proposed paroxytone stress pattern (for example, “PApa”), as it is considered the most common in BP, some children produced an oxytone stress pattern (“paPA”). Thus, the second syllable of the target word was performed as stressed and post-tonic. To guarantee the reliability of the data, with the subject maintaining the same stress of the word throughout its repetitions, the collected corpus was explored through statistical treatment in order to define distinct contexts that could interfere with the reading of the data.

Because the acoustic duration parameter is considered the main correlate of lexical stress in BP (Barbosa, 1996; Aquino, 1997; Gama-Rossi, 1999), duration measures of the V1C2V2 segments were extracted. As a result (figure 1), the acoustic measures showed 100% predictive value in segregating these speech samples into two groups and the most influential variables were duration of C2 and V2, equivalent to the duration of the syllable, finding support in the literature mentioned.

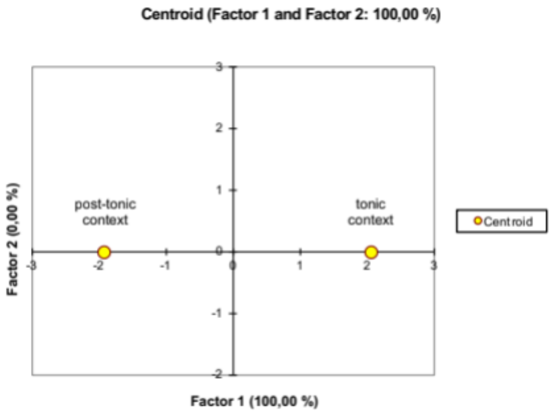


Figure 1: Centroid graph of the discriminant analysis for estimating the subjects' speech productions in the tonic and post-tonic contexts, based on the extracted acoustic measures.

The speech samples were then classified into studied group and control group and according to tonic and post-tonic contexts.

The collected data were analyzed acoustically using the PRAAT software and involved the acoustic inspection of the waveform and broadband spectrogram, and extraction of the measurements: f0 at the beginning of vowel following the consonant (V2); f0 at the stationary point of V2; F1 at the beginning of V2; F1 at the stationary point of V2; measures of duration of the plosive consonant (C2), duration of the previous vowel (V1) and of the following vowel (V2) to the consonant, and duration of the V1C2V2 excerpt of the target word; measures of duration of the VOT; and duration measures of the voicing bar. The voicing bar measures were extracted to contemplate gradient productions and included: duration of the voicing period (voicing bar period in the consonant stretch before the articulation release); duration of the voiceless period (length of the stretch in which there is no voicing bar before the articulation is released); duration of the voicing pre-plosion period (duration of the voicing bar when it is performed after a period of silence in an excerpt before the articulation is released); duration of the total pre-plosion period (total duration of the stretch prior to the release of the articulation, regardless of whether or not there is a voice bar); plosion duration (burst duration: period between the starting point of articulation release and the beginning of the vowel). The relative duration measures of all extracted duration measures described above were calculated to eliminate influences from the subject's speech rate. For sequence of analyses, measures of relative duration were used.

The speech production data were submitted to a battery of statistical tests to consider the existing variants in speech, as proposed in research developed by the partnership between researchers and professors from LIAAC-PUCSP and the Actuaries and Quantitative Methods Department-PUCSP.

The collection of speech perception data was carried out through an experiment elaborated using the PRAAT software. The sound stimuli consisted of the target words of the carrier sentences of the subjects' speech productions. Although the analysis and data reading refer to the V1C2V2 excerpt of the target word, for the experiment, the target word was edited in its entirety (C1V1C2V2) to avoid the effects of sample editing. The stimuli were presented in random order to the judges of the perception experiment.

Thirty-nine judges were selected of the same age and level of education, without hearing and/or speech complaints and with no connections to the fields of languages and speech-language pathology. The procedure was performed individually and with the use of headphones, and each judge orthographically transcribed the word as he or she heard it (word identification).

Next, a statistical analysis was performed to verify the reliability of the judges' answers, which resulted in the exclusion of four judges (figure 2). To analyze the data from the perception experiment, the responses of thirty-five judges were considered.

Based on Johnson [2003], the answers were tabulated in confusion matrices. Afterwards, the calculation of the auditory distances of the consonant pairs was performed, allowing visualization in the form of graphs.

After the speech production data and perception experiment data stage, a logistic regression analysis was performed to explore the acoustic and auditory spheres. Research was approved by the Ethics Committee, protocol number 119/09.

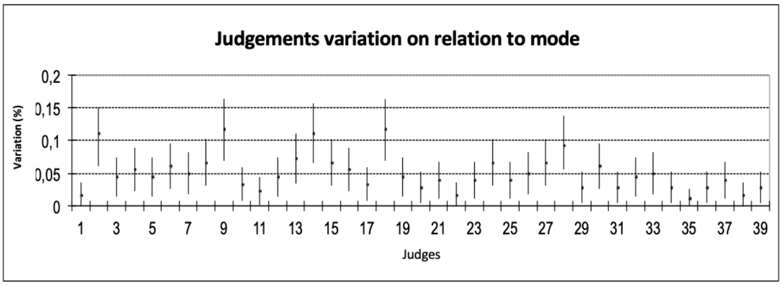


Figure 2: Representation of the auditory judgment variation of each judge in relation to mode value for verification of the judge's performance.

Results and discussion

Acoustic measures were compared between unvoiced and voiced plosive consonant pairs and, in general, revealed significant differences and allowed for classification of the control and studied groups for both stress contexts (figures 3 and 4).

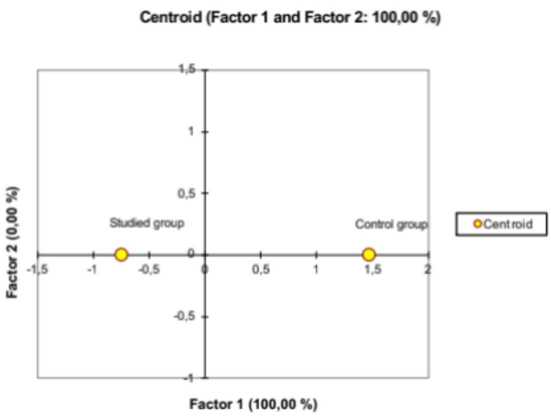


Figure 3: Centroid graph of the discriminant analysis of the estimation of the group of subjects (studied and control) in the tonic context from the acoustic measures.

The results of the calculation of the relative durations of the V1C2V2 excerpt (figures 5 to 8), as well as the voicing bar details (figures 9 to 12), showed differences between the control and studied groups for both stress contexts.

In terms of perception, auditory distances were smaller for samples from the studied group compared to the control group in the tonic context (figure 13). For the post-tonic context, the auditory distances were similar in both groups (figure 14).

The logistic regression analysis revealed that the most influential acoustic measures in the auditory judgments for the unvoiced plosive consonant were, in the tonic context, duration of the previous vowel

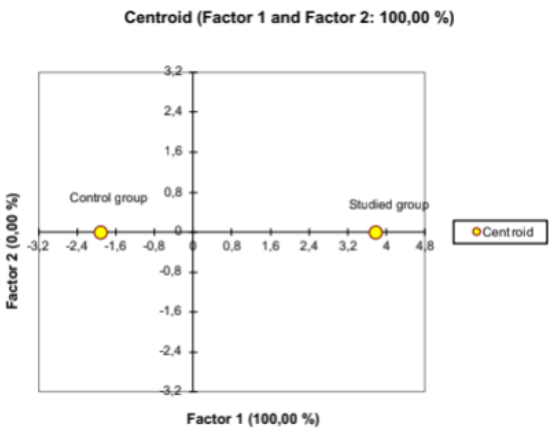


Figure 4: Centroid graph of the discriminant analysis of the estimation of the group of subjects (studied and control) in the post-tonic context from the acoustic measures.

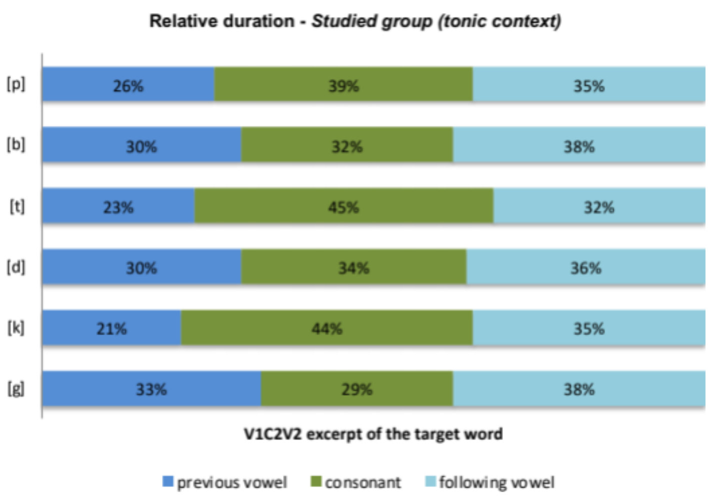


Figure 5: Schematic representation of the relative duration (%) of the vowel preceding the consonant, the plosive consonant and the vowel following the consonant, in relation to the V1C2V2 excerpt of the target word, of the speech samples in the tonic context of the studied group.

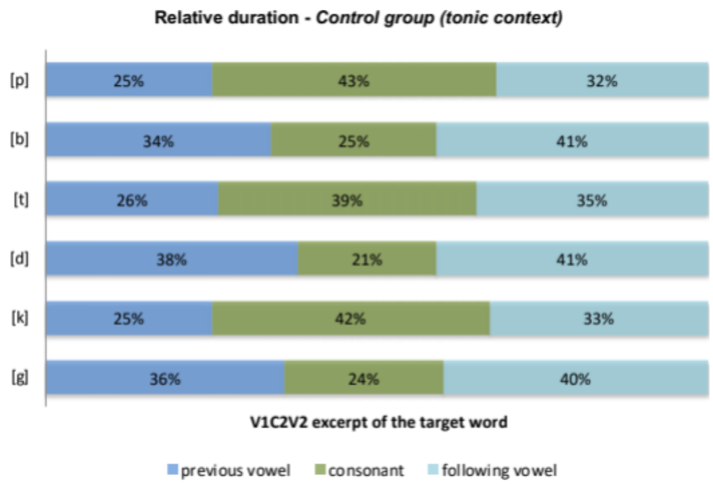


Figure 6: Schematic representation of the relative duration (%) of the vowel preceding the consonant, the plosive consonant, and the vowel following the consonant, in relation to the V1C2V2 excerpt of the target word, of the speech samples in the tonic context of the control group.

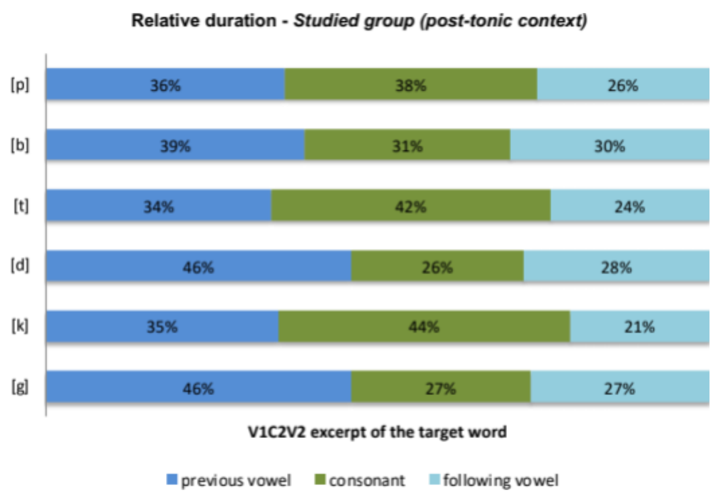


Figure 7: Schematic representation of the relative duration (%) of the vowel preceding the consonant, the plosive consonant, and the vowel following the consonant, in relation to the V1C2V2 excerpt of the target word, of the speech samples in the post-tonic context of the studied group.

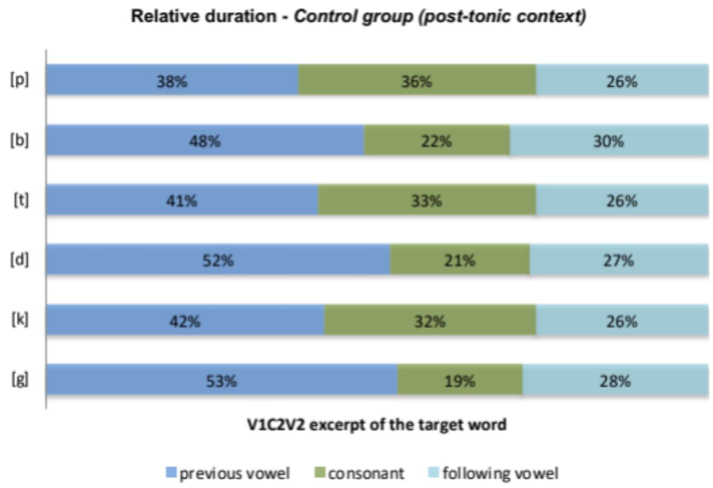


Figure 8: Schematic representation of the relative duration (%) of the vowel preceding the consonant, the plosive consonant, and the vowel following the consonant, in relation to the V1C2V2 excerpt of the target word, of the speech samples in the post-tonic context of the control group.

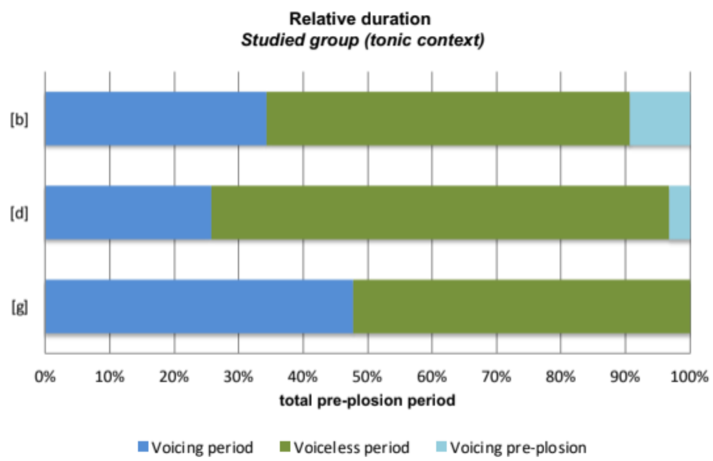


Figure 9: Schematic representation of the relative duration (%) of the voicing period, voiceless period, and the voicing pre-plosion period, in relation to the relative duration of the total pre-plosion period, of the speech samples in the tonic context of the studied group.

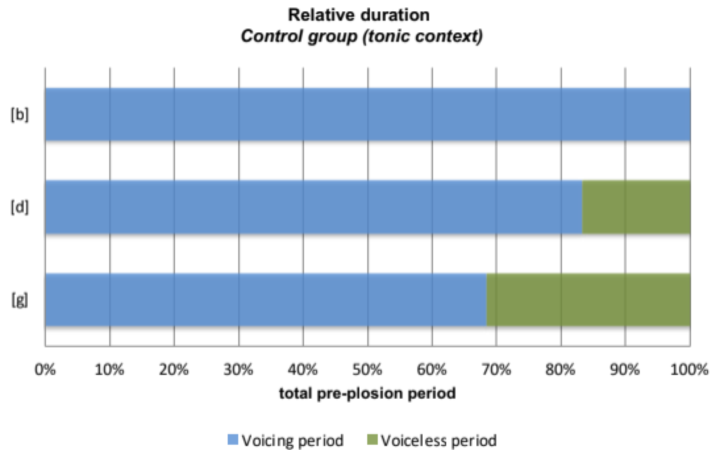


Figure 10: Schematic representation of the relative duration (%) of the voicing period and voiceless period, in relation to the relative duration of the total pre-plosion period, of the speech samples in the tonic context of the control group.

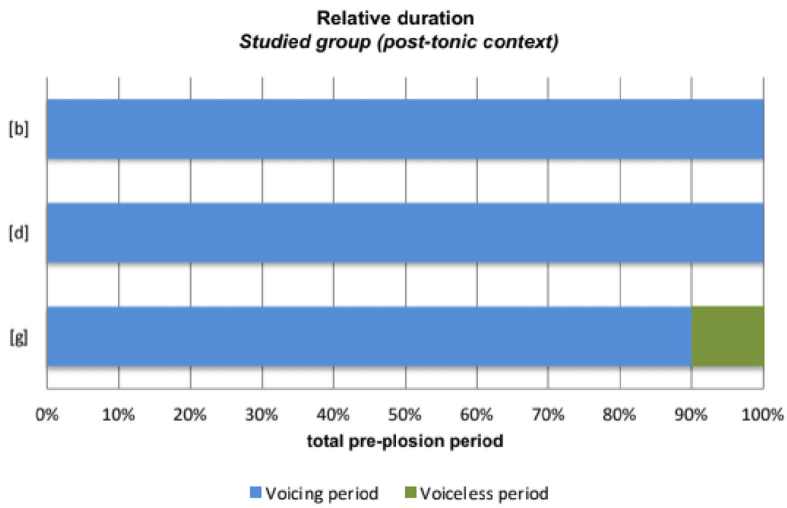


Figure 11: Schematic representation of the relative duration (%) of the voicing period and voiceless period, in relation to the relative duration of the total pre-plosion period, of the speech samples in the post-tonic context of the studied group.

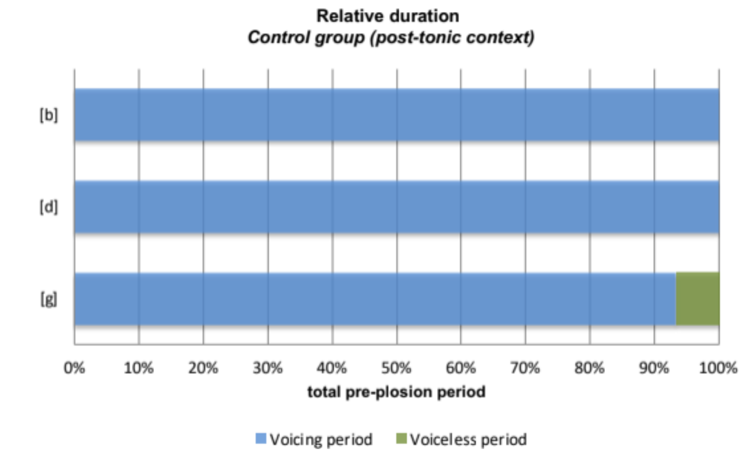


Figure 12: Schematic representation of the relative duration (%) of the voicing period and voiceless period, in relation to the relative duration of the total pre-plosion period, of the speech samples in the post-tonic context of the control group.

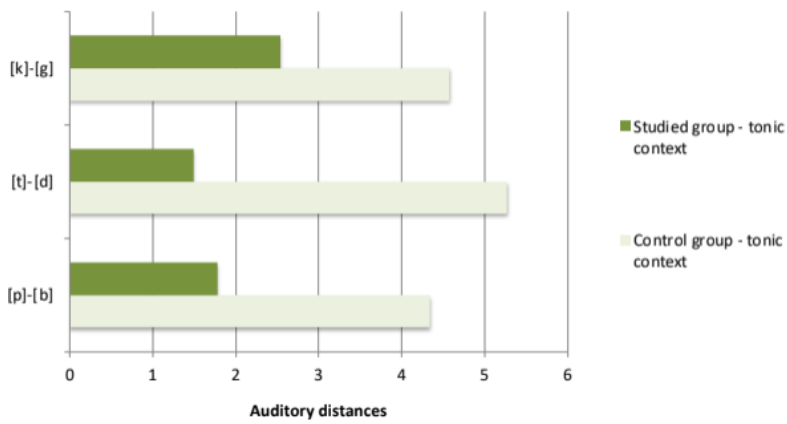


Figure 13: Auditory distances between voiceless and voiced pairs of speech productions in the tonic context of the studied and control groups as a function of the judges' responses in the perception experiment.

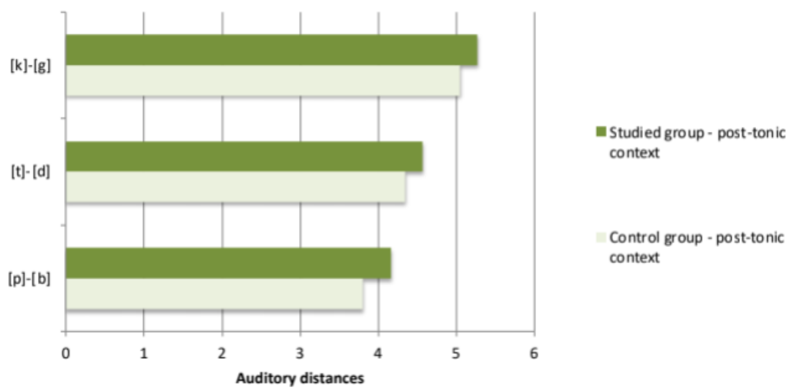


Figure 14: Auditory distances between voiceless and voiced pairs of speech productions in the post-tonic context of studied and control groups as a function of the judges' responses in the perception experiment.

(V1) and f0 at the beginning of the vowel (V2) (figure 15), and in the post-tonic context, duration of the plosive consonant (C2) and f0 at the beginning of the vowel (V2) (figure 16).

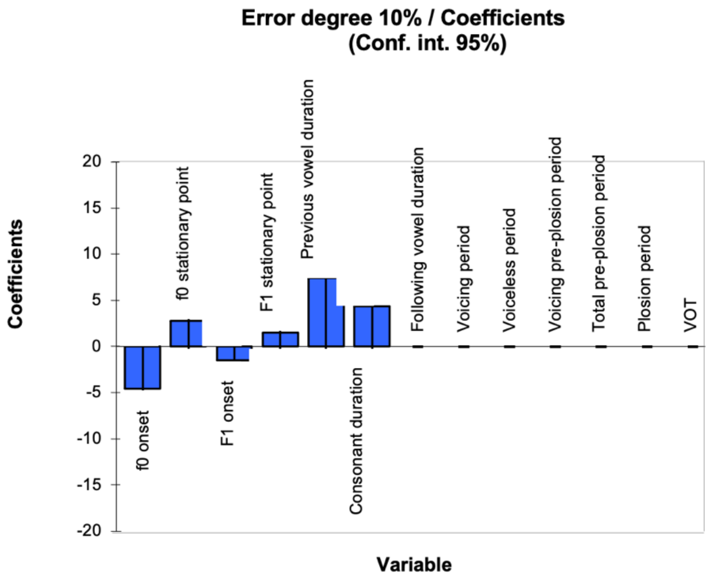


Figure 15: Logistic regression analysis graph for the estimation of auditory judgments from the acoustic measures of the production of unvoiced plosive consonants in the tonic context by the studied and control groups.

For the voiced plosive consonant, the most influential acoustic measures in the auditory judgments were, in the tonic context, the duration of the plosive consonant (C2), the duration of the total preplosion and the duration of the voiceless period (figure 17), and in the post-tonic context, the duration of the total pre-plosion period and duration of the voicing period (figure 18).

The VOT duration measure was not revealed as a predictive acoustic cue in the auditory judgment of voicing, in both stress contexts, while other duration measures involved in the voicing bar details were deemed complementary in explaining the implementations that speakers with disorders make and influencing the perception of altered speech.

The data are corroborated by the BP studies that considered the gradient productions from the voicing bar details in the productions of subjects with hearing impairment [Ficker, 2003, Barzaghi et al., 2007, Pereira, 2012] and without speech impairment [Gregio et al., 2011], suggesting the relevance of several acoustic cues for implementing voicing contrast.

In the voiced bilabial plosives of the studied group tonic context, a longer period of voicing pre-plosion period was observed, suggesting

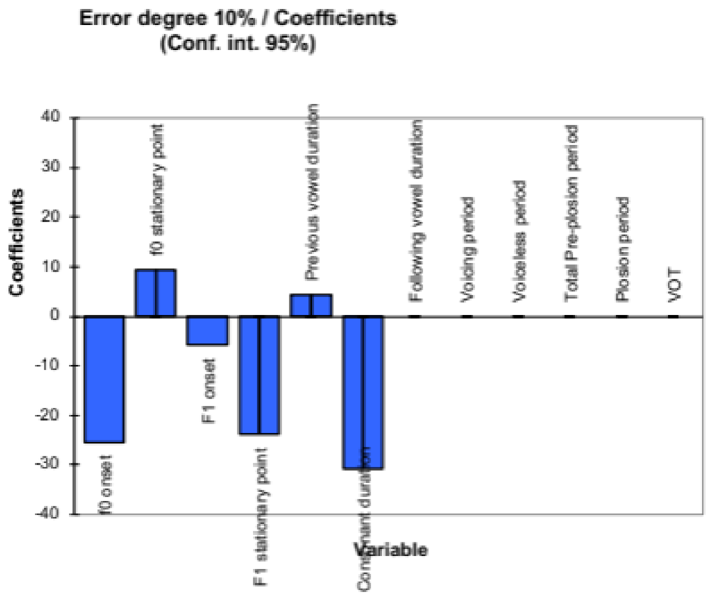


Figure 16: Logistic regression analysis graph for the estimation of auditory judgments from the acoustic measures of the production of unvoiced plosive consonants in the post-tonic context by the studied and control groups.

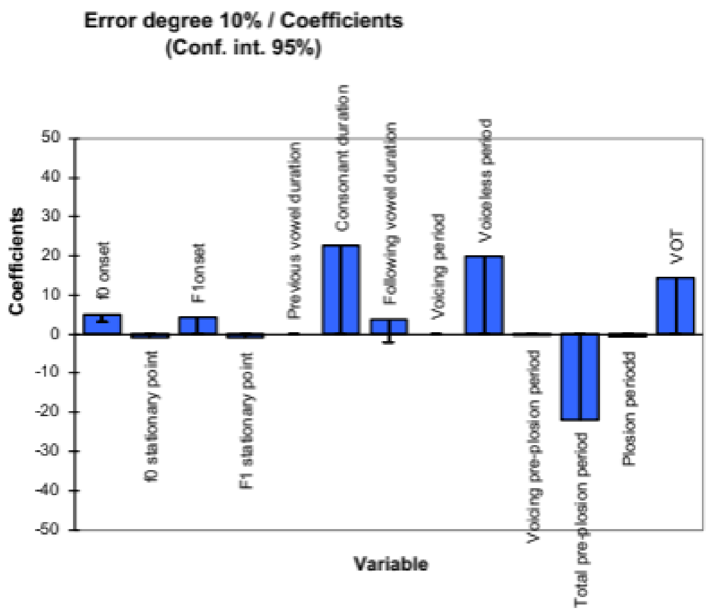


Figure 17: Logistic regression analysis graph for the estimation of auditory judgments from the acoustic measures of the productions of voiced plosive consonants in the tonic context by the studied and control groups.

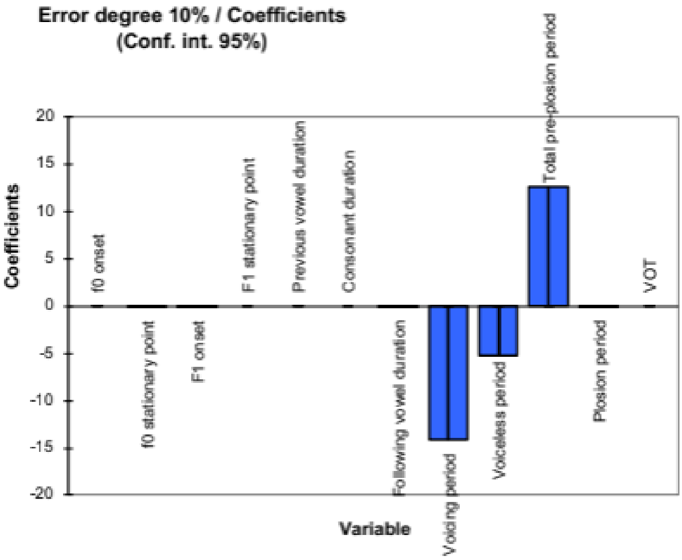


Figure 18: Logistic regression analysis graph for the estimation of auditory judgments from the acoustic measures of voiced plosive consonant productions in the post-tonic context by the studied and control groups.

an attempt to guarantee the necessary voicing period, since bilabials, according to the literature, have a longer voicing period. As for voiced plosive velar, no voicing pre-plosion period was observed, suggesting that the speaker perceives these differentiated cues for the articulatory points in trying to produce the voicing contrast. The literature indicates that velar plosives have a shorter duration of the voicing period [van Alphen and Smits, 2004, Lousada et al., 2005, Barzaghi et al., 2007, Pereira, 2012].

As for the influence of the f0 at the beginning of the vowel, the voicing contrast judgment revealed a predictive value with regard to unvoiced plosive in both stress contexts. As the f0 measurement results from vocal fold activity, which involves aerodynamic and physiological aspects, it tends to be higher at the beginning of the vowel following an unvoiced consonant [Shimizu, 1996]. The f0 measurement has been identified as an acoustic cue in voicing contrasts [Whalen et al., 1993, Hanson, 2009, Gregio et al., 2011]. F1 measures, in turn, did not reveal a predictive value for the auditory judgment of voicing in either stress context.

Regarding the duration measures of the V1C2V2 excerpt, the studied group differentiated the duration of voiced and unvoiced segments, as it kept the voiced plosive consonants' duration shorter than the duration of their respective unvoiced pairs, as expected based on the literature. However, the studied group made this differentiation in a smaller proportion compared to the control group, suggesting difficulty in synchronizing the glottic and supraglottic adjustments, given different timing of overlapping gestures. The BP literature points to higher duration values for vowels preceding and following consonants in the production of voiced plosive segments [Barbosa, 1996, Gurgueira, 2006, de Oliveira e Britto, 2010], justifying their influence on the auditory judgment of speech disorder.

As a final consideration, the acoustic measures that influenced the perception of the auditory judgment of the voicing contrast have been shown to be different for each stress context. The listener seems to attribute different relevance to the acoustic cues involved in the auditory judgment of sound. The perception of voicing contrasts showed that listeners integrate several acoustic cues to identify and categorize a sound. Thus, a clinical diagnosis based on only one acoustic measurement can be inaccurate.

Most of the speech samples from the studied group, who presented clinical demand for speech therapy, could not be categorized as "sound exchanges or absences" in view of the data exploration in this study. The subjects revealed knowledge about the language, as they performed intermediate productions towards the determinant characteristics of the voicing contrast. Such signs denote that the subjects perceive differences and seek to implement different actions to support the voicing contrast at different articulation points. The acoustic cues relevant to the construction of voicing information resided in parameters of duration, which suggest clues about the process of neuromotor maturation of speech movements, aspects that have been suggested in previous studies with children [Levy, 1993, ?, Albano, 2007].

The demand for temporal refinement surpassed the issues of implementation of the f0 acoustic cue. Such aspects relate to the issue of synchronization of glottic and supraglottic gestures, which is important for the construction of voicing contrast information.

Thus, the exploration of the acoustic signal of the speech samples of the studied group indicated an attempt to mark the voicing contrast, suggesting that speakers perceive and try to differentiate in their production one sound category from the other, yet these attempts are not always processed as relevant information by the listener's perception.

Subjects control and organize their articulatory gestures in terms of physical aspects and in terms of perceptual feedback (Gama-Rossi, 1999; Albano, 2001; Albano, 2007). It is up to the professional to guide the child in their attempt to achieve phonic contrast, producing articulatory targets that are audible to the listener.

The challenge of working with issues that lie at the interface between speech production and perception offers a rich field of reflection on the nature of speech disorders. Such a challenge may result, in the future, in therapeutic actions that consider the particularities of the manifestation in question, which means contemplating the difficulties and recognizing the implementations made by the speaker, which although not audible at first glance, can be unveiled through instrumental investigation.

Conclusion

The investigation showed evidence of more than one acoustic cue for the implementation of voicing contrast. The duration of the plosive consonant, voiceless period, and total pre-plosion period (tonic context) and the total pre-plosion period and voicing period (post-tonic context) revealed predictive power of the auditory judgment of the altered speech voicing contrast for voiced plosives. For unvoiced plosive consonants, the influential measures were f0 at the beginning of the vowel and duration of the previous vowel (tonic context) and f0 at the beginning of the vowel and duration of the plosive consonant (post-tonic context).

Bibliography

Eleonora Cavalcante Albano. Representações dinâmicas e distribuídas: indícios do português brasileiro adulto e infantil. Letras de Hoje, 42(1), August 2007. ISSN 1984-7726. URL https://revistaseletronicas.pucrs.br/index.php/fale/article/view/675.

Flávia Viegas de Andrade. Análise de parâmetros espectrais da voz em crianças saudáveis de 4 a 8 anos. Master's thesis, Universidade Veiga de Almeida, Rio de Janeiro, 2009.

Plínio Almeida Barbosa. At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration: emphasis on automatic segmental duration generation. Cadernos de Estudos Linguísticos, 31, 1996. ISSN 2447-0686. DOI: 10.20396/cel.v31i0.8636991. URL https://periodicos.sbu.unicamp.br/ojs/index.php/cel/article/view/8636991.

Mário André Lopes Barroco, Marta Teresa Pedrosa Domingues, Maria de Fátima Marques de Oliveira Pires, Marisa Lousada, and Luis

M. T. Jesus. Análise temporal das oclusivas orais do Português Europeu: um estudo de caso de normalidade e perturbação fonológica. Revista CEFAC, 9(2):154–163, June 2007. ISSN 1516-1846. DOI: 10.1590/S1516-18462007000200003.

David Barton and Marlys A. Macken. An instrumental analysis of the voicing contrast in word-initial stops in the speech of four-year-old english-speaking children. Language and Speech, 23 (2):159–169, April 1980. ISSN 0023-8309, 1756-6053. DOI: 10.1177/002383098002300203. URL `http://journals.sagepub.com/doi/10.1177/002383098002300203`.

Luisa Barzaghi, Kátia Barbosa, and Samar M. El Malt. Deficiência de audição e contraste de vozeamento em oclusivas do português brasileiro: análise acústica e perceptiva. Distúrbios da Comunicação, 19(3):343–355, 2007.

M. S. Behlau. Análise de tempo de início de sonorização na discriminação dos sons do português. PhD thesis, Escola Paulista de Medicina, São Paulo, 1986.

Mara Behlau. Avaliacao e tratamento das disfonias. Editora Lovise, São Paulo, 1 edition, January 1995. ISBN 9788585274269.

José R. Benkí. Perception of VOT and first formant onset by Spanish and English speakers. In In, pages 240–248. Cascadilla Press, 2005.

Larissa Cristina Berti, Ana Elisa Falavigna, Jéssica Blanca dos Santos, and Rita Aparecida de Oliveira. Desempenho perceptivo-auditivo de crianças na identificação de contrastes fonológicos entre as oclusivas. Jornal da Sociedade Brasileira de Fonoaudiologia, 24 (4):348–354, 2012. ISSN 2179-6491. DOI: 10.1590/S2179-64912012000400010.

M.T.R.L. Bonatto. Vozes Infantis: a caracterização do contraste do vozeamento das consoantes plosivas do português brasileiro na fala de crianças de 3 a 12 anos. PhD thesis, Pontifícia Universidade Católica de São Paulo, São Paulo, 2007.

Z. A. Camargo and A. L. G. P. Navas. Fonética e fonologia aplicadas à aprendizagem. In J. Zorzi and S. Capellini, editors, Dislexia e outros distúrbios da leitura-escrita: letras desafiando a aprendizagem. Pulso, São José dos Campos, 2008.

Z.A. Camargo, M. A. S. Fontes, and S. Madureira. Introdução ao estudo dos sons da fala. apostila da disciplina de Fonética e Fonologia do curso de Fonoaudiologia-PUCSP, 2000.

Taehong Cho and Peter Ladefoged. Variation and universals in VOT: evidence from 18 languages. Journal of Phonetics, 27(2):207–229, April 1999. ISSN 00954470. DOI: 10.1006/jpho.1999.0094. URL https://linkinghub.elsevier.com/retrieve/pii/S0095447099900943.

Ana Teresa Brandão de Oliveira e Britto. Estudo do contraste de vozeamento em sujeitos com e sem desvio fonológico. PhD thesis, Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, 2010.

L. Barzaghi Ficker. Estudo da produção e percepção das plosivas do português brasileiro por um sujeito com deficiência auditiva. PhD thesis, Pontifícia Universidade Católica de São Paulo, São Paulo, 2003.

Fabiana Nogueira Gregio and Zuleica Antonia de Camargo. Dados de tempo de início do vozeamento (VOT) na avaliação do sinal vocal de indivíduos com paralisia unilateral de prega vocal. Distúrbios da Comunicação, 17(3):289–97, 2005.

Fabiana Nogueira Gregio, Renata de Moraes Queiroz, Andrea Baldi de Freitas Sacco, and Zuleica Camargo. O uso da eletroglotografia na investigação do vozeamento em adultos sem queixa de fala. Intercâmbio. Revista do Programa de Estudos Pós-Graduados em Linguística Aplicada e Estudos da Linguagem, 23, 2011. ISSN 2237-759X. URL https://revistas.pucsp.br/index.php/intercambio/article/view/8890.

Adriana Limongeli Gurgueira. Estudo acústico do "voice-onset-time" e da duração da vogal na distinção da sonoridade dos sons plosivos em crianças com transtorno fonológico. PhD thesis, Universidade de São Paulo, São Paulo, 2006.

Helen M. Hanson. Effects of obstruent consonants on fundamental frequency at vowel onset in English. The Journal of the Acoustical Society of America, 125(1):425–441, January 2009. ISSN 0001-4966. DOI: 10.1121/1.3021306. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2677272/.

Jeannette D. Hoit, Nancy Pearl Solomon, and Thomas J. Hixon. Effect of lung volume on voice onset time (VOT). Journal of Speech, Language, and Hearing Research, 36(3):516–520, June 1993. ISSN 1092-4388, 1558-9102. DOI: 10.1044/jshr.3603.516. URL http://pubs.asha.org/doi/10.1044/jshr.3603.516.

P. Hoole, C. Gobl, and A. N. Chasaide. Laryngeal coarticulation. In W. J. Hardcastle and N. Hewlet, editors, Coarticulation – theory, data and techniques. Cambridge University Press, Cambridge, 1999.

Keith Johnson. Acoustic and Auditory Phonetics. Wiley-Blackwell, 3 edition, July 2003.

Raymond D. Kent and Charles Read. The acoustic analysis of speech. Singular Pub. Group, San Diego, Calif, 1992. ISBN 9781879105430.

Márcia Keske-Soares, Ana Paula Félix Blanco, and Helena Bolli Mota. O desvio fonológico caracterizado por índices de substituição e omissão. Revista da Sociedade Brasileira de Fonoaudiologia, 9(1): 10–18, 2004.

Ivone Panhoca Levy. Uma nova face da nau dos insensatos: a dificuldade de vozear obstruintes em crianças de idade escolar. PhD thesis, Universidade Estadual de Campinas, Campinas, 1993.

Leigh Lisker and Arthur S. Abramson. A cross-language study of voicing in initial stops: Acoustical measurements. WORD, 20(3):384–422, January 1964. ISSN 0043-7956, 2373-5112. DOI: 10.1080/00437956.1964.11659830. URL `http://www.tandfonline.com/doi/full/10.1080/00437956.1964.11659830`.

M. Lousada, P. Martins, and L. M. T. Jesus. Estudo do pré-vozeamento, frequência do burst e locus do f2 das oclusivas orais do português europeu. In Actas do XXI Encontro Nacional da APL, pages 485–494, Porto, Portugal, 2005.

Sandra Madureira, Luisa Barzaghi, and Beatriz Mendes. Voicing contrasts and the deaf: Production and perception issues. In Investigations in Clinical Phonetics and Linguistics.

Roberta Michelon Melo, Helena Bolli Mota, Carolina Lisbôa Mezzomo, Brunah de Castro Brasil, Liane Lovatto, and Leonardo Arzeno. Parâmetros acústicos do contraste de sonoridade das plosivas no desenvolvimento fonológico típico e no desviante. Revista da Sociedade Brasileira de Fonoaudiologia, 17(3):304–312, 2012. ISSN 1516-8034. DOI: 10.1590/S1516-80342012000300012.

Helena Bolli Mota, Aline Berticelli, Cintia da Conceição Costa, Fernanda Marafiga Wiethan, and Roberta Michelon Melo. Ocorrência de dessonorização no desvio fonológico: relação com fonemas mais acometidos, gravidade do desvio e idade. Revista da Sociedade

Brasileira de Fonoaudiologia, 17(4):430–434, December 2012. ISSN 1516-8034. DOI: 10.1590/S1516-80342012000400011.

Lílian Cristina Kuhn Pereira. As consoantes plosivas do PB: um estudo acústico e perceptivo sobre dados de falade sujeitos com deficiência auditiva. PhD thesis, Pontifícia Universidade Católica de São Paulo, São Paulo, 2012.

Paulina D. Artimonte Rocca. O desempenho de falantes bilíngües: evidências advindas da investigação do VOT de oclusivas surdas do inglês e do português. DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada, 19(2):303–328, 2003. ISSN 0102-4450. DOI: 10.1590/S0102-44502003000200004.

Luciana Lessa Rodrigues, Maria Cláudia Freitas, Eleonora Cavalcante Albano, and Larissa Cristina Berti. Acertos gradientes nos chamados erros de pronúncia. Letras, 0(36):85–112, December 2008. ISSN 2176-1485. DOI: 10.5902/2176148511968. URL `https://periodicos.ufsm.br/letras/article/view/11968`.

Lucila Rey Rocha Schliemann. Contraste de vozeamento por crianças entre 6-8 anos: uma abordagem dinâmica. Master's thesis, Universidade Estadual de Campinas, Campinas, 2011.

Katsumasa Shimizu. A cross language study of voicing contrasts of stop consonants in Asian languages. Seibido, Tokyo, 1996. ISBN 9784791966486.

Adelaide Silva, Vera Pacheco, and Leonardo Oliveira. Por uma abordagem dinâmica dos processos fônicos. Revista Letras, 55(0), 2001. ISSN 2236-0999. DOI: 10.5380/rel.v55i0.2821. URL `https://revistas.ufpr.br/letras/article/view/2821`.

Adelaide Hercília Pescatori Silva. O estatuto da análise acústica nos estudos fônicos. Cadernos de Letras UFF. Dossiê: Letras e cognição, 41(1):213–229, 2010.

Ana Paula Ramos de Souza, Lisiane Collares Scott, Carolina Lisbôa Mezzomo, Roberta Freitas Dias, and Vanessa Giacchini. Avaliações acústica e perceptiva de fala nos processos de dessonorização de obstruintes. Revista CEFAC, 13(6):1127–1132, May 2010. ISSN 1982-0216. DOI: 10.1590/S1516-18462010005000039.

Patricia M. Sweeting and Ronald J. Baken. Voice onset time in a normal-aged population. Journal of Speech, Language, and Hearing Research, 25(1):129–134, March 1982. ISSN 1092-4388, 1558-9102.

DOI: 10.1044/jshr.2501.129. URL http://pubs.asha.org/doi/10.1044/jshr.2501.129.

Ryosuke O. Tachibana, Tatsuya Kitamura, and Masako Fujimoto. Differences in articulatory movement between voiced and voiceless stop consonants. Acoustical Science and Technology, 33(6):391–393, 2012. ISSN 1346-3969, 1347-5177. DOI: 10.1250/ast.33.391. URL https://www.jstage.jst.go.jp/article/ast/33/6/33_E1255/_article.

Petra M. van Alphen and Roel Smits. Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: the role of prevoicing. Journal of Phonetics, 32(4):455–491, October 2004. ISSN 00954470. DOI: 10.1016/j.wocn.2004.05.001. URL https://linkinghub.elsevier.com/retrieve/pii/S0095447004000324.

João Veloso. Vozeamento, duração e tensão nas oposições de sonoridade das oclusivas orais do português. Revista da Faculdade de Letras : Línguas e Literaturas / Linguística, 14:59–80, 1997.

D. H. Whalen, Arthur S. Abramson, Leigh Lisker, and Maria Mody. F0 gives voicing information even with unambiguous voice onset times. The Journal of the Acoustical Society of America, 93(4):2152–2159, April 1993. ISSN 0001-4966. DOI: 10.1121/1.406678. URL http://asa.scitation.org/doi/10.1121/1.406678.

D. H. Whalen, Andrea G. Levitt, and Louis M. Goldstein. VOT in the babbling of French- and English-learning infants. Journal of phonetics, 35(3):341–352, July 2007. ISSN 0095-4470. DOI: 10.1016/j.wocn.2006.10.001. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2717044/.

PART V:

TEST

Fairy tales are more than true: not because they tell us that dragons exist, but because they tell us dragons can be beaten.

C.K. CHESTERTON

this is a test