NEUTRALISATION AND THE PERCEPTION OF CLOSE-MID AND OPEN-MID VOWELS: THE GRADIENT BETWEEN PHONOLOGICAL CATEGORIES

Megumi Im[1]; João Veloso[1]; Maria Conceição Costa[2]; Luis M. T. Jesus[3]

[1]University of Porto, Porto, Portugal; [2]Department of Mathematics (DMat) and Centre of Research and Development in Mathematics and Applications (CIDMA), University of Aveiro, Portugal [3]School of Health Sciences (ESSUA), Institute of Electronics and Informatics Engineering of Aveiro (IEETA), Intelligent Systems Associate Laboratory (LASI), University of Aveiro, Aveiro, Portugal
megumiim1004@gmail.com; jveloso@letras.up.pt; lopescosta@ua.pt; lmtj@ua.pt

ABSTRACT

A prototype theoretical framework was the basis of a study on the impact of neutralisation on the perception of Portuguese close-mid and open-mid vowels and the gradient between phonological categories.

Fifteen Portuguese listeners from Lisbon participated in identification and goodness rating tasks of 20 random repetitions of 31 stimuli from a /i/-/e/-/ɛ/-/a/ continuum. Boundaries between categories were explored using logistic regression curves and analysis of variance.

Results from the identification task revealed four vowel categories marked by three statistically distinct boundaries. The prototype of a category, i.e., the stimulus with the highest goodness rating score, was approximately in the centroid and the goodness score decreased as the stimulus moved away from this centroid. The boundary between /e/ and /ɛ/, the underlying opposition that neutralises at surface level, was less steep than between /ɛ/ and /a/, the opposition that never neutralises, a possible influence of the phonological system on speech perception.

Keywords: Speech perception, Phonological system, Mid front vowels, The Prototype Theory.

1. INTRODUCTION

The way we decode speech signals into phonemic categories is still unclear, but some studies [4], [11], [15] show that prototypes play an important role in structuring the categories of speech sounds.

Prototypical phonemes stored in long-term memory may serve as cognitive reference points against which real items are judged, that is, listeners compare the *input* acoustic signals with the phonemes idealised in their linguistic knowledge [15]. When an *input* is similar to the prototype (the best representative of a given phonemic category), it is recognised by the listener as a member of that category [4], [15]. This model implies a hypothetical mechanism in which phonemic categories are represented in terms of prototypes as sole representatives of all members of the same category, and that it is the distance to these prototypes, that defines phonemic categories [11]. Thus, the members of a phonemic category do not have the same representative relevance.

Phonemes have a distinctive function in a given language and are abstract entities, stored as part of the speakers' knowledge of the language, and activated in the production and perception of speech sounds [23]. Listeners are less accurate in the perception of partial contrasts which neutralise in certain contexts [10] and are less sensitive to the difference between allophones than to the contrasting sounds of their own language [2].

The difference in the perception of the neutralizable opposition of the mid vowels and the constant opposition of other vowels has been previously observed in Brazilian Portuguese (BP) studies [21], [22]. As in European Portuguese (EP), the phonological system of BP has oppositions of mid vowels in the tonic syllable. Regarding the process of neutralising the mid vowels, BP presents different realisations in relation to those of EP within the unstressed, non-final context: [e] for front vowels and [o] for back vowels [1], [25].

Silva and Neves [21], [22] carried out two experiments to study the potential differences in the way in which the contrasts between mid-back vowels [o] and [ɔ] [21] and mid front vowels [e] and [ɛ] [22] are represented in the perceptual system of BP speakers. Results showed that the boundary between /e/ and /ɛ/ was less well defined and required a more elaborate processing, thus justifying the complex phonological descriptions [8], [9], [19] of these vowels.

This study was based on two theoretical assumptions: The perception of speech sounds is influenced by the relationships they have within their language system [2], [10], [23]; the categories of speech sounds, especially vowels, have an internal structure [13], [14]. Therefore, this study's main goal was to explore whether the neutralisation of vocalic

phonemes had an influence on the perception of EP, more specifically, ascertaining the gradient of the boundary between the categories of close-mid and open-mid front vowels as defined by Portuguese listeners, comparing it with the gradient of boundaries between other front vowel categories.

2. HYPOTHESIS

Considering the complexity observed in the Portuguese vowel system [3], [17], [18], [24], [26], [25], the proposals of the studies mentioned above, and the results obtained for BP [21], [22], we formulated the following hypotheses: H1. There are four phonemic categories with a gradient structure (although there may be dispersion intra and inter-listeners, as previously observed in speech production studies); H2. The gradient of the boundary between the mid vowels /e/ and /ɛ/ is lower than other boundaries.

To test the first hypothesis (H1) the following questions were posed: Q1. Can four phonemic categories be observed? Q2. Do these categories have a gradient internal structure?

To test the second hypothesis (H2), we analysed four front vowel categories and their boundaries, the presence, in these categories, of an internal structure with prototypes, and the gradient of the boundaries. Postulating an internal structure of the phonemic category and, consequently, distinct boundaries between categories, the gradient of these boundaries was studied using goodness rating tasks.

3. METHOD

The stimuli used in the perception experiments were synthesised with a morphing technique [12] that is based on natural speech samples. Having observed [5] a convergence in the production of mid front vowels by young speakers from Lisbon and considering the role perception units play in speech production [6], [7], [20], stimuli were based on a male speaker from Lisbon, maintaining the characteristics of natural speech during the synthesis process.

3.1. Selection of the speaker

We selected a speaker with the same characteristics as those recruited for the study of Escudero et al. [5], that is, young males with permanent residence in Lisbon, in order to generate stimuli from natural speech. Five speakers (aged between 21 and 35) were recruited who met the criteria and participated in the recording, having previously signed an informed consent.

The speech samples were recorded individually in Lisbon and Porto, and participants were asked to speak spontaneously for 2 minutes on a common topic. The recordings were classified by 8 experts, on whether the recording had the typical characteristics of the vowels of the Lisbon dialect and whether the quality of the voice on the recording was pleasant. We selected the speaker with the highest average score and the lowest standard deviation for the two criteria.

3.2. Recording of the target-vowels

New recordings were made in Lisbon with the selected speaker (21 years of age; university student of the 3rd year of an Undergraduate Degree). We used an *AKG Perception 120 USB* condenser microphone connected to a laptop computer. The recordings were made with *Praat version 6.0.37*, at a sampling frequency of 48000 Hz, with 16 bits per sample, and the data stored in mono .wav format (Windows PCM) with no compression. The speaker produced the target vowels [i, e, ɛ, a] in isolation three times. We chose to record samples with a drop in tone at the end, a procedure similar to that used in the study by Silva and Neves [22], thus seeking to generate the stimuli in the most natural fashion possible. We used a duration of 400 ms, similar to that of the study by Masapollo et al. [16].

3.3. Creation of stimuli

We selected one repetition of each of the four vowels, with the following criteria: A drop in the f_o of 20-25 Hz from beginning to end; values of the frequency of F_1 and F_2, as close as possible to those of Escudero et al. [5].

These four samples were used as anchors to generate 31 stimuli with the *morphing* procedure in the *TANDEM-STRAIGHT monolithic Package 014* (function *Morphing Menu Last Modified by GUIDE v2.5 19-Jul-2016 01:42:59*), while dividing each trajectory into 11 stages: The stimulus stimulus001, stimulus011, stimulus021 and stimulus031 correspond to the anchors of the vowels /i/, /e/, /ɛ/ and /a/, respectively. As the value of F_1 increased, the values of F_2 and F_3 decreased, thus reflecting the characteristics of the stimulus filter.

3.4. Perception pre-test

The pre-test was used to verify the experimental procedures and if there was any issue with the stimuli generated. We recruited a homogeneous group, all of them from the North of Portugal, namely, Braga and Porto. All pre-test experiments were carried out in a quiet room at the University of Minho.

Six (females = 3) EP speakers, from the North of Portugal, university students, aged between 20 and 24

(average = 22.3 years of age), participated voluntarily in the pre-test.

3.5. Perception test

Fifteen (females = 9) EP speakers from Lisbon, with academic qualifications above an undergraduate degree or attending an undergraduate degree, aged between 19 and 34 (average = 23.7 years of age), voluntarily participated in the main test.

The participants listened to each stimulus 20 times, randomly ordered (620 stimuli in total), with an optional pause after every 155 stimuli. There was no possibility of repetition. The experiment included two tasks: 1) to choose one of the words presented on the screen containing the tonic vowel they heard, with an option X in case of no match (identification task); 2) to classify the quality of the sound heard on a scale of 1 to 5 according to the vowel chosen in the task, 1 being a bad example of this vowel and 5 a good example (goodness rating task).

The main test was carried out in the city of Lisbon, and the stimuli presented via *Praat version 6.0.37*, installed in a *TOSHIBA dynabook PT65DGP-RJA* laptop computer, using *Sennheiser* HD 380 Pro headphones connected to its internal sound card. During the presentation of the stimuli, the two-syllable words with the target-vowel present in the tonic syllable, <pico>, <medo>, <teto>, <pato> and <X> were displayed on the computer screen. In addition to the words mentioned above, a scale of 1 to 5 was also displayed at the bottom of the screen.

The answers were given using the computer mouse, by selecting a word and a number for each sound heard. Data from the identification task (task 1) and from the goodness rating task (task 2) were collected with *Praat version 6.0.37*.

3.6. Data processing

The data were statistically analysed with *R studio software Version 1.1.453* and *Excel 2016*, which were also used in the creation of graphs. Since it was intended to statistically analyse the boundaries between categories in the identification task (task 1), and the internal structures of each category and the gradient of each boundary in the goodness rating task (task 2), the number of stimuli were considered as independent variables, each being a distinct identity defined by the values of the acoustic parameters, and the responses of the participants in each task as dependent variables.

4. RESULTS

This section presents the analysis and discussion of the results.

4.1. Identification task

The first research hypothesis (H1) has to do with the existence of speech perception boundaries between the phonemic categories /i/-/e/, /e/-/ɛ/ and /ɛ/-/a/. In order to determine the boundaries between two categories, data from the identification task was initially explored using logistic regression curves that allowed us to estimate the boundary between two categories, which corresponds to 0.5 probability of response. Three boundaries, shown in Fig. 1, between categories were estimated for each participant.

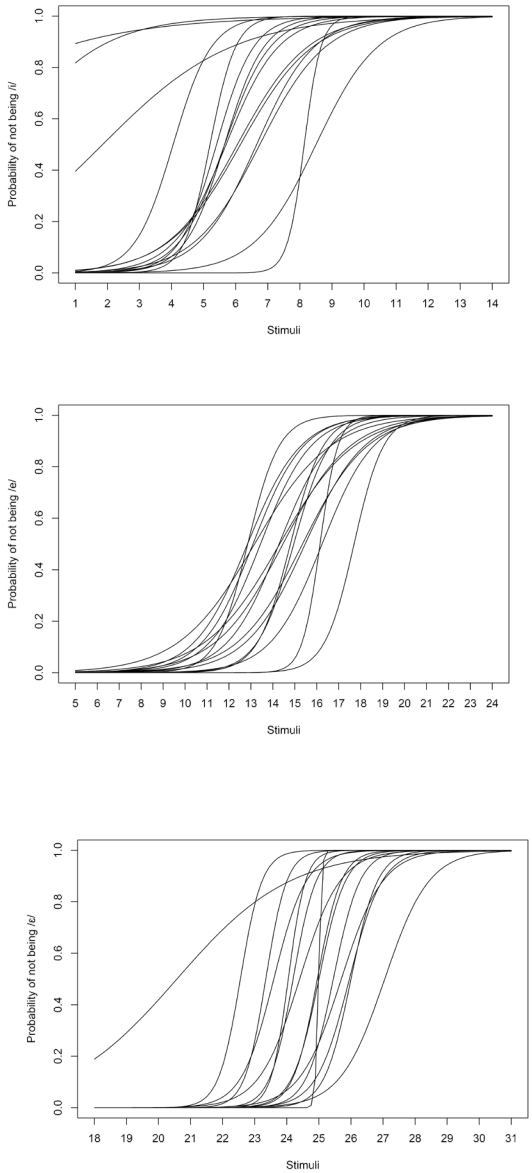


Figure 1: The logistic regression curves of all participants for the /i/-/e/ (top), /e/-/ɛ/ (middle) and /ɛ/-/a/ (bottom) transitions.

Estimates (median / mean ± standard deviation) for the three boundaries were obtained from the logistic regression curves adjusted for all participants:

/i/-/e/ (boundary 1) – 5.65 / 5.28 ± 2.40; /e/-/ɛ/ (boundary 2) – 14.46 / 14.61 ± 1.42; /ɛ/-/a/ (boundary 3) – 24.95 / 24.52 ± 1.59.

Analysis of Variance (ANOVA) was carried out to assess whether there was a significant distance between the samples of the three boundaries for each participant. The null hypothesis of equality of medians between groups was rejected (F=403.7, df=2, *p*-value $<2 \times 10^{-16}$) and multiple comparisons were performed using the Tukey's HSD test. None of the 95% confidence intervals (CI) for the mean of the differences between groups contained zero: Comparison between boundaries 1 and 2 – CI (-10.97; -7.68); boundaries 1 and 3 – CI (-20.89; -17.60); boundaries 2 and 3 – CI (-11.56; -8.27).

4.2. Goodness rating task

The sum of all points (1 to 5) from the goodness rating task of each stimulus were calculated for each participant and represented in a bubble chart. The median value was used to represent the group of all participants (see Fig. 2).

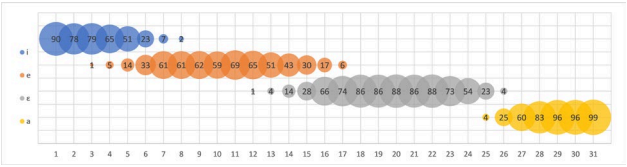


Figure 2: Median goodness rating of all participants, for each of the 31 stimuli (horizontal axis). Scores are represented by the sizes of the circles. The number inside each circle is the median score for each stimulus, considering all participants, which may range from 1 (worst) to 100 (best).

Fig. 2 shows that as the stimulus moves away from the centroid, the goodness score decreases. Most listeners used four categories with three boundaries, exhibiting an internal structure that included their prototype.

Considering the three boundaries revealed by the identification results and the goodness score, a central question of our study was finally addressed, that is, if the boundary between /e/ and /ɛ/ was less steep than other boundaries in the perception of EP vowels.

Median/mean comparison tests for paired samples (Wilcoxon tests) revealed that there were no significant differences between the median values (*p-value* = 0.5760) at boundary 1, but there were significant differences for both boundary 2 (*p-value* = 0.0499) and boundary 3 (*p-value* = 0.0170).

The gradient of the boundaries was also analysed empirically, by calculating the ratio between the median goodness scores at the boundaries between categories, while considering boundary 1 = stimulus 6, boundary 2 = stimulus 15 and boundary 3 = stimulus 25. Our assumption was: The closer to 1 the ratio between goodness scores on a boundary is, the more confusing the categorisation decision will be. When comparing the three boundaries, it was observed that boundary 2 was the one with the ratio closest to 1 (28 / 30 = 0.93), which suggests a greater degree of confusion. Second, there was boundary 1, with a ratio of 23 / 33 = 0.70, and finally, boundary 3 with a ratio of 4 / 23 = 0.17, which suggests a lower degree of confusion than in the previous two.

The gradient of boundary 3 could be considered the highest, because it corresponded to the lowest value of the ratio, and it was also the only boundary for which significant differences were found between the medians of the goodness scores. According to the statistical analysis, there was a considerable degree of confusion in the first two boundaries, since there were no significant differences between the mean/median values of their goodness scores. The empirical analysis of the ratio revealed boundary 2 as the least steep of the three.

5. CONCLUSIONS

The identification task revealed that the participants used four distinct vowel categories, which are in accordance with the phonological theory and with the what has been observed in Portuguese speech production studies. From the answers obtained in the goodness rating task, we were able to observe the internal structure of each category, which has its prototype in the centroid.

The boundaries between /i/ and /e/, and /e/ and /ɛ/ were less steep than the boundary between /ɛ/ and /a/, the latter being non-neutralizable. The gradient of the /e/-/ɛ/ boundary could be due to the complex phonological relations that these sounds establish in the vowel system of EP. There is also a possibility of dialectal specificity but this question remains open for future work.

Stimuli beyond /i/ and /a/ would have covered the entire categories of these two vowels. Variance comparison tests could be carried out, which would allow us to analyse the amplitude and variability of the categories. The same study could be extended to the BP variant and to other EP dialects, providing new evidence for the cognitive and/or phonological theoretical model.

6. ACKNOWLEDGEMENTS

This work was supported by National Funds through the FCT - Foundation for Science and Technology, in the context of the projects UIDB/00022/2020 (Centro de Linguística da Universidade do Porto – CLUP),

UIDB/00127/2020 (Institute of Electronics and Informatics Engineering of Aveiro – IEETA), UIDB/04106/2020 and UIDP/04106/2020 (Center for R&D in Mathematics and Applications – CIDMA).

7. REFERENCES

[1] L. Bisol, *Introdução a estudos de fonologia do português brasileiro*. Porto Alegre: Edipucls, 2001.

[2] A. Boomershine, K. C. Hall, E. Hume, and K. Johnson, "The impact of allophony versus contrast on speech perception," in *Phonological contrast: Perception and acquisition*, P. Avery, E. Dresher, and K. Rice, Eds. New York: Mouton de Gruyter, 2008, pp. 146–172.

[3] J. B. Carvalho, "Contrastive hierarchies, privative features, and portuguese vowels," *Linguística: Revista de Estudos Linguísticos da Universidade do Porto*, vol. 6, no. 1, pp. 51–66, 2011.

[4] O. Eerola, J. Savela, J.-P. Laaksonen, and O. Aaltonen, "The effect of duration on vowel categorization and perceptual prototypes in a quantity language," *J Phon*, vol. 40, no. 2, pp. 315–328, Mar. 2012.

[5] P. Escudero, P. Boersma, A. S. Rauber, and R. A. H. Bion, "A cross-dialect acoustic description of vowels: Brazilian and European Portuguese," *J Acoust Soc Am*, vol. 126, no. 3, pp. 1379–1393, Sep. 2009.

[6] B. G. Evans and P. Iverson, "Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences," *J Acoust Soc Am*, vol. 115, no. 1, pp. 352–361, Jan. 2004.

[7] B. G. Evans and P. Iverson, "Plasticity in vowel perception and production: A study of accent change in young adults," *J Acoust Soc Am*, vol. 121, no. 6, pp. 3814–3826, 2007.

[8] J. Harris, *English Sound Structure*. Cambridge: Blackwell, 1994.

[9] J. Harris and G. Lindsey, "Vowel patterns in mind and sound," in *Phonological knowledge: Conceptual and empirical issues*, N. B. Roberts, P. Carr, and G. Docherty, Eds. Oxford University Press, 2000.

[10] E. Hume and K. Johnson, "The impact of partial phonological contrast on speech perception," in *Proceedings of the 15th International Congress of Phonetic Sciences*, 2003, pp. 2385–2388.

[11] P. Iverson and P. K. Kuhl, "Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism?," *Percept Psychophys*, vol. 62, no. 4, pp. 874–886, Jan. 2000.

[12] H. Kawahara, T. Takahashi, M. Morise, and H. Banno, "Development of exploratory research tools based on TANDEM-STRAIGHT," in *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, 2009, pp. 111–120.

[13] P. K. Kuhl, "Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not," *Percept Psychophys*, vol. 50, no. 2, pp. 93–107, Mar. 1991.

[14] P. K. Kuhl, K. Williams, F. Lacerda, K. Stevens, and B. Lindblom, "Linguistic experience alters phonetic perception in infants by 6 months of age," *Science (1979)*, vol. 255, no. 5044, pp. 606–608, Jan. 1992.

[15] S. E. Lively and D. B. Pisoni, "On prototypes and phonetic categories: a critical assessment of the perceptual magnet effect in speech perception.," *J Exp Psychol Hum Percept Perform*, vol. 23, no. 6, pp. 1665–79, Dec. 1997.

[16] M. Masapollo, L. Polka, M. Molnar, and L. Ménard, "Directional asymmetries reveal a universal bias in adult vowel perception," *J Acoust Soc Am*, vol. 141, no. 4, pp. 2857–2869, Apr. 2017.

[17] M. H. M. Mateus, "Aspectos Fonológicos e Prosódicos da Gramática do Português," in *Gramática da Língua Portuguesa*, A. M. Brito, I. Duarte, and I. H. Faria, Eds. 2003, pp. 987–1053.

[18] M. H. M. Mateus and E. Andrade, *The Phonology of Portuguese*. Oxford: Oxford University Press, 2000.

[19] A. Nevins, "Vowel lenition and fortition in Brazilian Portuguese," *Let Hoje*, vol. 47, pp. 228–233, 2012.

[20] R. S. Newman, "Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report," *J Acoust Soc Am*, vol. 113, no. 5, pp. 2850–2860, May 2003.

[21] D. M. R. Silva and R. R. Neves, "Um estudo experimental sobre a percepção do contraste entre as vogais médias posteriores do português brasileiro," *DELTA*, vol. 25, no. 2, pp. 319–345, 2009.

[22] D. M. R. Silva and R. R. Neves, "Perception of height and categorization of Brazilian Portuguese front vowels," *DELTA*, vol. 32, no. 2, pp. 355–373, Aug. 2016.

[23] N. S. Troubetzkoy, *Grundzuge der Phonologie*. 1939.

[24] J. Veloso, "O sistema vocálico e a redução e neutralização das vogais átonas em português europeu contemporâneo," in *Manual de Linguística Portuguesa*, A. Martins and E. Carrilho, Eds. Berlin: De Gruyter, 2016, pp. 636–662.

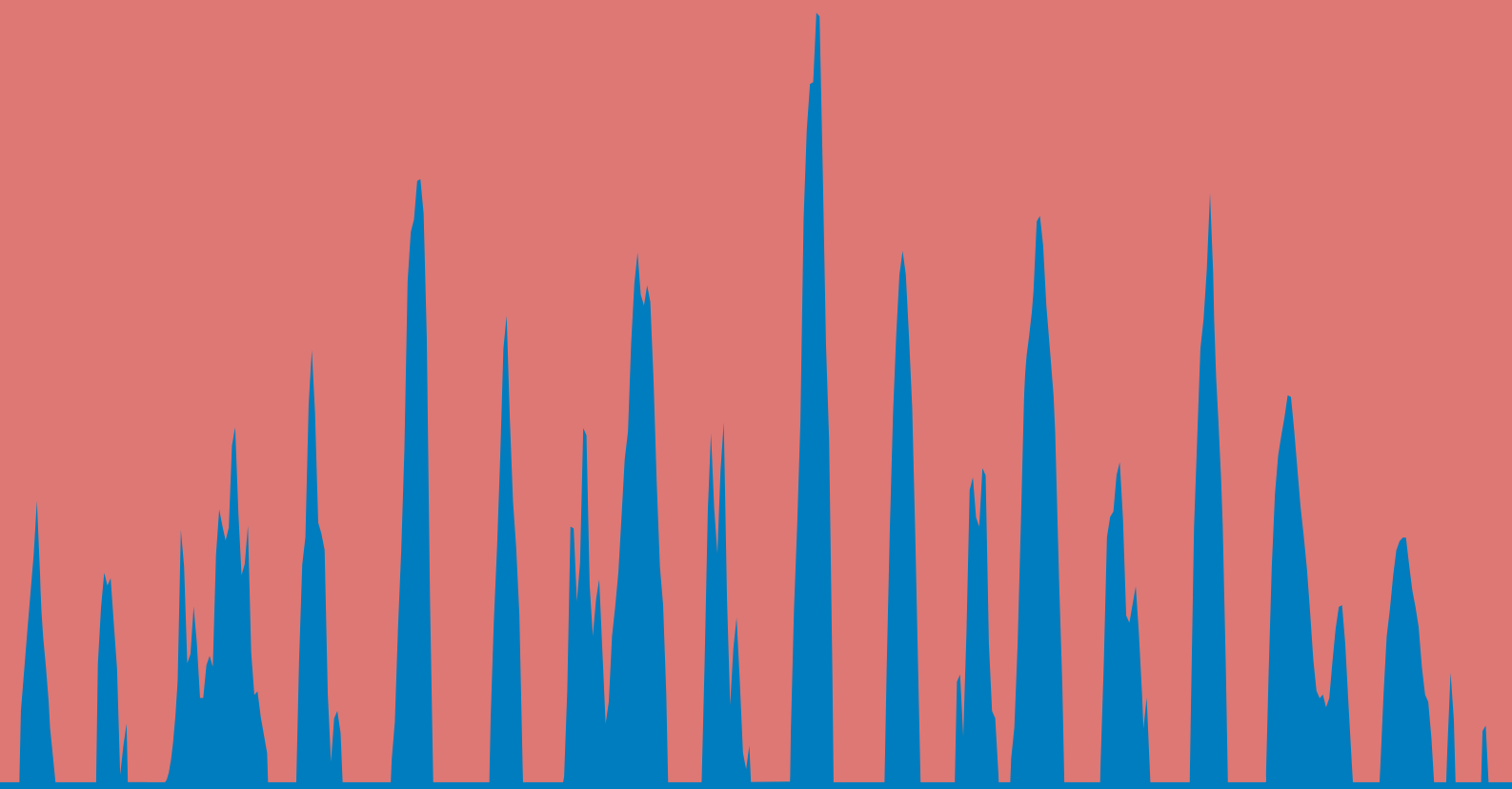
[25] W. L. M. Wetzels, "Mid vowel neutralization in Brazilian Portuguese," *Cadernos de Estudos Lingüísticos*, vol. 23, pp. 19–55, 1992.

[26] W. L. M. Wetzels, "The representation of vowel height and vowel height neutralization in Brazilian Portuguese (Southern Dialects)," in *Tones and Features: Phonetic and Phonological Perspectives*, J. A. Goldsmith, E. Hume, and W. L. M. Wetzels, Eds. 2011, pp. 331–360.

August 7–11, 2023 — Prague, Czech Republic

ICPhS 2023

CONGRESS PROCEEDINGS



ICPhS 2023
August 7–11, 2023
Prague Congress Center, Czech Republic

20th International Congress of Phonetic Sciences (ICPhS)

August 7–11, 2023
Prague Congress Center, Czech Republic

ELSEVIER

Title:Proceedings of the 20th International Congress of Phonetic Sciences, Prague 2023
Publisher: GUARANT International spol. s r.o.
Edited by Radek Skarnitzl and Jan Volín
Edition: first
Prague, August 2023
ISBN 978-80-908 114-2-3

To cite papers in this volume:
Author Name(s) (2023). Paper title. In: Radek Skarnitzl & Jan Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. XX–YY). Guarant International.