



Jesus, Luis M. T. & Im, Megumi & Veloso, João & Costa, Maria Conceição. 2024. Influence of the European Portuguese phonological system on the perception of close-mid and open-mid vowels. *Glossa: a journal of general linguistics* 9(1). pp. 1–40. DOI: <https://doi.org/10.16995/glossa.10607>



Open Library of Humanities

## Influence of the European Portuguese phonological system on the perception of close-mid and open-mid vowels

**Luis M. T. Jesus**, School of Health Sciences (ESSUA), Institute of Electronics and Informatics Engineering of Aveiro (IEETA), Intelligent Systems Associate Laboratory (LASI), University of Aveiro, Aveiro, Portugal, [lmtj@ua.pt](mailto:lmtj@ua.pt)

**Megumi Im**, Faculdade de Letras da Universidade do Porto (FLUP) and Centro de Linguística da Universidade do Porto (CLUP), University of Porto, Porto, Portugal, [megumiim1004@gmail.com](mailto:megumiim1004@gmail.com)

**João Veloso**, Faculdade de Letras da Universidade do Porto (FLUP) and Centro de Linguística da Universidade do Porto (CLUP), University of Porto, Porto, Portugal, [jveloso@letras.up.pt](mailto:jveloso@letras.up.pt)

**Maria Conceição Costa**, Department of Mathematics (DMat) and Centre of Research and Development in Mathematics and Applications (CIDMA), University of Aveiro, Portugal, [lopecosta@ua.pt](mailto:lopecosta@ua.pt)

This study's purpose was to analyse the influence of neutralisation on the perception of European Portuguese (EP) close-mid and open-mid vowels, and the gradient between the /i/, /e/, /ɛ/ and /a/ phonological categories. A fuzzy boundary between mid-vowels has been described in the phonological systems of various languages including EP, a variety of Portuguese with complex phonological phenomena and exceptional realisations of vowels in different contexts. Fifteen listeners from Lisbon participated in two tasks: Identification and goodness rating of 20 random repetitions of 31 stimuli from a /i/-/e/-/ɛ/-/a/ continuum. The boundaries between categories in the identification task, the internal structures of each category and the gradient of each boundary in the goodness rating task were analysed. Logistic regression was used to analyse the individual responses of listeners and the internal structure of phonemic categories, and prototypes were determined using bubble charts of goodness scores. The gradient of boundaries was analysed using two sets of data: X responses and goodness ratings. Models of the gradients of transition were developed using a mixed effects regression approach. Results from the identification task showed four vowel categories marked by three distinct boundaries. The stimulus with the highest score was approximately in the centroid and the goodness score decreased as the stimulus moved away from the centroid, which constitutes new evidence of an internal structure of the phonemic categories in EP. The main implications of the results obtained are for the description and typology of phonological phenomena in EP, supporting the existence of a non-partial contrast (which does not neutralise in any context) between /ɛ/ and /a/, and revealing less robust /i/-/e/ and /e/-/ɛ/ phonemic oppositions.

*Glossa: a journal of general linguistics* is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

OPEN ACCESS



## 1 Introduction

This study is based on two theoretical assumptions: The perception of speech sounds is influenced by the relationships they have within their language system (Trubetzkoy 1939; Hume & Johnson 2003; Boomershine et al. 2008; Shi et al. 2010); the categories of speech sounds, especially vowels, have an internal structure (Kuhl 1991; Kuhl et al. 1992; Vihman & Croft 2007). The main goal was to explore the boundary between the categories of close-mid and open-mid front vowels as defined by Portuguese listeners and to compare them with the boundaries between other front/ central vowel categories. This work intends to shed a new light on the original idea of Trubetzkoy (1939) that “(t)he psychological difference between constant and neutralizable distinctive oppositions is very great”, that is, “constant distinctive oppositions” are clearly perceived and “neutralizable distinctive oppositions fluctuate”.

This phenomenon of a fuzzy boundary (Zhang et al. 2022) between mid-vowels is not unique to EP; it has been observed in mid vowels in other Romance languages, that have been explored using a variety of methods and approaches: Italian (Renwick & Ladd 2016); Catalan (Nadeu & Renwick 2016); Galician (Amengual & Chamorro 2016); French (Stevenson & Zamuner 2017); Romanian (Renwick 2014).

In Trubetzkoy’s (1939) theoretical model of phonology, phonemes have a distinctive function in each language and are abstract entities, stored as part of the speakers’ knowledge of the language, which are activated in the production and perception of speech sounds (Veloso 2015; Frisch 2018). Hume & Johnson (2003) analysed differences in perception between partial contrasts (which neutralise in certain contexts), and non-partial contrasts (which do not neutralise in any context) in Mandarin tones, by comparing Mandarin native speakers with American English native speakers (a language with none of these contrasts), concluding that there is a reduction in sensitivity in the perception of partial contrasts. In Boomershine et al. (2008), the perceptual differences between allophonic and contrasting sounds in English and Spanish were analysed. The consonants [ð, d, r] were used as stimuli, given that, in English, [d, r] are allophones but [ð] is always contrasting; while, in Spanish, [ð, d] are allophones and [r] is contrasting. The results showed that listeners of each language are less sensitive to the difference between allophones than to the contrasting sounds of their own language.

The way we decode speech signals into phonemic categories is still unclear, but some studies (Lively & Pisoni 1997; Iverson & Kuhl 2000; Eerola et al. 2012; McMurray 2022) show that prototypes play an important role in structuring the categories of speech sounds. According to Lively & Pisoni (1997), 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. When an input is like the prototype (the best representative of a given phonemic category), it is recognised by the listener as a member of that category (Lively & Pisoni 1997; Eerola et al. 2012). 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 through the distance in relation to these prototypes, that the members are perceived, at least partially, as a phonemic category (Iverson & Kuhl 2000). Thus, the idea of an internal structure of category is postulated in which its members do not have the same representative relevance. What is considered to be the internal structure of vowels in this framework is the “graded structure” of phonemic categories “with some stimuli perceived as better exemplars of the category than others”. This “graded internal structure is revealed” by “overt judgements of category goodness” (Miller 1994: 272).

This idea is in line with the results of Pisoni’s (1973) experiment on the speech perception of consonants and vowels with two tasks: Identification and discrimination of a /i/-/ɪ/ continuum. In this experiment the listeners noticed small differences even within a single category, revealing greater difficulty in the act of categorising vowels, compared to that of stop consonants. Thus, they demonstrated their ability to capture more information in vowels beyond the discrete phonemic labels (Iverson & Kuhl 2000). In other words, listeners were more sensitive to the internal structure of the vowel categories.

## 2 The vowel system of European Portuguese

The vowel system of EP can be divided into two subsystems: One of tonic vowels and one of unstressed vowels, due to the close relationship between the quality of the vowel and the stress of the word in this variety of Portuguese (Mateus & Andrade 2000; Correia et al. 2015). The second subsystem (unstressed vowels) can be further divided into three subsets according to their position in the word and the position of the stress. We, therefore, have seven oral vowels in the stressed position and four in the unstressed position in the phonological inventory of EP (Mateus 2003). The phonetic realisations in the tonic position correspond to the seven segmental vowels of the phonological system /i, e, ε, a, ɔ, o, u/. There are oppositions between the close-mid and open-mid vowels in the system of this language, since lexical pairs are found whose meanings are not related (Veloso 2016: 644), such as: <sede> [ˈsɛdi] <headquarters> vs. [ˈsedi] <thirst>; <molho> [ˈmɔlu] <bunch> vs. [ˈmoɫu] <gravy>. In the unstressed position, four vowels [i, ɨ, ɐ, u] can be found in a non-final position and three [ɨ, ɐ, u] in the final position. These vowels in unstressed syllables are derived from one of the seven phonological segments found in the tonic position. Their vowel quality is altered through a phonological process. When morphophonologically compared, the corresponding relationship between the underlying form and the surface form can be observed (Mateus 2003).

The following contexts and processes should be considered when analysing the realisation of EP vowels in unstressed position (Vigário 2003; 2022; Bisol & Veloso 2016; Veloso 2016; Andrade 2020): Word initial non-close unstressed vowels when realised in prosodic word-initial

position, are not reduced (Andrade 2020: 3327–3329) – <ovelha> [o'vɛʎɐ] <sheep>, where /o/ is not reduced to [u] (Vigário 2022: 844); vowel reduction does not occur in complex/branching nuclei (diphthongs) – < vaidade > [vaj'dadi] <vanity>, where /a/ does not reduce to [ɐ] (Vigário 2022: 844) or <oitavo> [oj'tavu] <eighth>, where /o/ does not reduce to [u] (Vigário 2003: 70); /e/ is centralised before palatals, <telha> [tɛ'ʎɐ] <roof tile> (Vigário 2022: 845), where /e/ is centralised to [ɐ]; /ow/ is monophthongised, a process that is responsible for [o] in word-internal stressless position – <roupheiro> [ʁo'pɛjru] <wardrobe><sup>1</sup> (Vigário 2022: 846); there are several cases of vowel opening in prosodic word-final position (some with further morphological specification) – <júnior> [ʒunjɔɾ] <junior>, where /o/ opens to [ɔ] (Vigário 2022: 848) or <líder> [li'dɛɾ] <leader>, where /e/ opens to /ɛ/ (Vigário 2003: 86); /i/ produced in the final position of verbs and in palatal contexts patterns with non-close/ non-back vowels (Mateus & Andrade 2000) – <parte> [partɨ] <he/ she leaves>, where /i/ in the verb <partir> [par'tir] <to leave> is centralised to [i] (Vigário 2003: 69–70; Andrade 2020: 3311–3313). These phonological processes do not apply uniformly across EP dialects (Rodrigues & Martins 1999; Vigário 2003; 2022; Segura 2013).

There are occurrences contributing to the complexity of the relationships that the vowels establish within EP phonological system, in the pre-tonic unstressed position such as (Veloso 2016): 1) < europeu > [ɛwru'pew] <European>, < açoitar > [ɛso'itar] <to flog>; 2) < inflação > [ĩfla'sɛw] <inflation>, < economia > [ekɔnu'mie] <economy>; 3) < ecologista > [ɛkulu'zi'tɛ] <ecologist>, < obrigado > [ɔbri'gadu] <thank you> and in the post-tonic as 4) < reporter > [ʁi'pɔɾtɛɾ], < plâncton > [plɛktɔn] <plankton>, being 1) the result of the post-lexical process of constructing diphthongs in EP, and not of lexical forms; 2) lexically marked exceptions; 3) and 4) result from exceptional rules that prevent the reduction of the unstressed vocalism. There are also exceptional terminations of words with realisations of unstressed [i] and [ɛ] (Veloso 2016): < táxi > [taksi] <taxi>, < biquíni > [bi'kini] <bikini>, < inclusive > [ĩklu'zive] <including> and < exclusive > [ɛʃklu'zive] <excluding>. Although these can suffer changes due to phonological regularisation, the strength of the neologism currently reduces the productivity of the vocalic reduction of EP. The reduction of unstressed vowels is less productive in contemporary EP (in words that entered the EP lexicon in recent years), particularly in neologisms, than in the realisations of words from the EP heritage lexicon, or even most of the lexicon introduced until the mid-20th century (Veloso 2016: 655–656): < telefone > [tɛlɨ'fɔni] <telephone> (/e/ reduced to [i]) and < telemóvel > [tɛlɛ'mɔvɛl] <mobile phone> (produced without vowel reduction).

---

<sup>1</sup> Although /o/ appears in unstressed position, this is not a case of a true exception to vowel reduction, because in the underlying form there is a diphthong, and vowel reduction does not apply to vowels in a branching nucleus.

The proposal with seven phonological vowels is widely accepted, but the phonological inventory of Portuguese vowels is still under discussion (Veloso 2016). For example, Carvalho (2011) and Wetzels (1992; 2011) proposed a phonological vowel system with the five segments /i, E, a, O, u/, which represents, in a more abstract manner, the pairs of the mid vowels. On the other hand, Veloso (2016) proposed the simultaneous existence of lexical opposition between the mid vowels and the existence of an underlying vowel (/E/ and /O/) in the phonological system of EP, thus explaining two types of opposition of this language: Lexical opposition and grammatical opposition (Trubetzkoy 1939; Ladd 2006; Veloso 2016). Lexical oppositions refer to lexical pairs, which contrast only by their degree of openness at the underlying level. As for grammatical oppositions, they refer to pairs of words with radicals that have the same meaning, that is, in these pairs there is no contrast of meaning, but of grammatical category (e.g., noun or verb). Veloso (2012; 2016) also proposes the inclusion of vowels [i] and [ɐ] in the phonological system of EP. The author argues for the existence of the theoretical forms /i/ and /ɐ/ in nominal endings (e.g., bas[i] and cas[ɐ],) and in clitics (e.g., qu[i], d[i], m[i], lh[i] and s[i]) which are not forms derived from /e, ε/ (or /E/) and /a/. Therefore, Veloso's (2012; 2016) proposal for the phonological description of vowels includes /i, e, ε, a, o, ɔ, u, ɐ, i/, lexically specified and multivocally related to the surface forms in the sense that: The same segment can be produced as distinct sounds (phonetic level) – <dedo> {/e/ → [e] or [i]}; the same sound can be related to two different segments at the underlying level – <fogueira> {[u]→/o/ or /u/}; the inclusion of the vowels /i, ɐ/ in the underlying level, makes the relations between the phonetic realisations of unstressed vowels with the theoretical forms more complex than in the previous proposals – [ɐ] → /ɐ/; [i] → /i/; [u] → /u/ or [ɐ] → /a/; [i] → /e, ε/; [u] → /o, ɔ/.

The phonetics of Portuguese [i, e, ε, a, ɔ, o, u] were the focus of Escudero's et al. (2009) speech production study of 20 young adults from Lisbon (females = 10) and 20 young adults from São Paulo (females = 10). The seven vowels were produced in a stressed syllable in CVCV context, and five acoustic parameters were estimated: Fundamental frequency ( $f_0$ ), formant frequencies ( $F_1$ ,  $F_2$ , and  $F_3$ ) and duration. There was considerable intra-speaker and inter-speaker dispersion of  $F_1$  and  $F_2$  frequencies, but the results showed that the /e/ and /ε/ first formant frequencies ( $F_1$ ) are closer in EP than in Brazilian Portuguese (BP). It was concluded that the approximation of the  $F_1$  values of /e/ and /ε/ was mainly due to the elevation of /ε/ and not so much to the lowering of /e/. The authors considered this phenomenon of imminent convergence of /ε/ and /e/ to be a process of linguistic change (Escudero et al. 2009).

The production experiments of Andrade (2020), based on 6 female and 6 male adult speakers of standard EP, aged between 25 and 35 years, revealed that if only  $F_1$  and  $F_2$  formant frequency values were considered, there was considerable variability and overlap of the 12 speakers, and that  $F_3$  and  $F_4$  frequencies should be considered as discriminative segmental cues for front vowels /i, e, ε/. Andrade (2020) also used synthetic stimuli to run vowel perception experiments

involving the identification of vowels by 22 EP listeners. Production results were compared with the listener's responses revealing that the  $F_3$  frequencies, the relative  $F_3$ - $F_2$  and  $F_3$ - $F_4$  frequencies, play a central role in differentiating the front vowels.

The difference in the perception of the neutralizable opposition of the mid vowels and the constant opposition of other vowels was also observed in BP studies (Silva & Neves 2009; 2016). As in 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 (Câmara 1970; Wetzels 1992; Bisol 2001). Silva & Neves (2009; 2016) carried out two experiments of categorical perception to study the potential differences in the way contrasts between mid-back vowels [o] and [ɔ] (Silva & Neves 2009) and mid front vowels [e] and [ɛ] (Silva & Neves 2016) are represented in the perceptual system of BP speakers.

Twelve speakers (females = 6) of BP from Belo Horizonte, aged between 18 and 27 years, participated in the first study (Silva & Neves 2009). The experiment included a classification task of a continuum of sounds between [u] and [ɔ] (two alternatives forced choice – 2AFC), and two discrimination tasks with two subcontinua between [u] and [o], and between [o] and [ɔ] (four intervals two alternatives forced choice – 4I2FC). In the classification task, a more abrupt curve was observed in the category transition between [u] and [o] than in the transition between [o] and [ɔ]. Regarding the discrimination tasks, there was a correlation with the classification result only in the 2AFC task, by calculating the discriminability value of this task and using it as a sensitivity measure (Silva & Neves 2009). This value was higher in the continuum between [u] and [o] than between [o] and [ɔ]. Therefore, the distinction between [o] and [ɔ] was less stable and less relevant than that between [u] and [o] at the level of the categories stored in long-term memory and used in the perception of BP.

The second study (Silva & Neves 2016) involved forty-two participants (females = 20) from the central region of Minas Gerais. They performed two identification tasks: 2AFC between /i/ and /e/, and between /e/ and /ɛ/. The results revealed a greater discriminability and faster response in the task that involved the /i/ and /e/ continuum. These results indicated that the boundary between /e/ and /ɛ/ was less defined and required a more complex processing during categorisation, thus justifying the complex phonological descriptions (Harris 1994; Harris & Lindsey 2000; Nevins 2012) of these vowels.<sup>2</sup>

---

<sup>2</sup> The authors based their work on two phonological models of the complexity of the opposition between mid-vowels: Contrastive Hierarchy Theory (Dresher, 2003a) and Lee's (2010) model for BP; Phonology of the elements (Harris, 1994; Harris & Lindsey, 2000), analysed by Nevins (2012) in BP. In both models, the descriptions of these oppositions between close-mid and open-mid are more complex than those between close and open vowels (Silva & Neves, 2009, 2016).

### 3 Hypotheses

Considering the complexity observed in the Portuguese vowel system, the proposals of the studies mentioned above, and the results obtained for BP (Silva & Neves 2009; 2016), we formulated the following hypotheses:

H1. There are four phonemic categories with a gradient structure; there may be intra and inter-listener dispersion, as previously observed in speech production studies (Escudero et al. 2009);

H2. The gradient of the boundary between the mid vowels /e/ and /ɛ/ is less marked/ less steep than in the other boundaries.

To test the first hypothesis (H1) the following questions were posed:

Q1. Can four phonemic categories be identified?

Q2. Do these categories have a gradient internal structure?

To test these hypotheses, we analysed four vowel categories and their boundaries, the presence, in these categories, of an internal structure with prototypes (H2), and the gradient (Aaltonen et al. 1997) of the boundaries (H2); the gradient of the boundary characterises the slope of the perceptual transition, i.e., the abruptness of perceptual transition, between vowels. Postulating an internal structure of the phonetic category and, consequently, distinct boundaries between categories (H1), the gradient of these boundaries was studied using goodness rating tasks (H2).

### 4 Method

The stimuli used in the perception experiments were synthesised with a morphing technique (Kawahara et al. 2009) that is based on natural speech samples since, in previous studies (Greene 1986; Logan & Greene & Pisoni 1989; Ralston et al. 1991; Pisoni 1997), a difference was observed (in the listeners' responses) between natural and synthesised speech, due to the lack of acoustic information (typical of natural speech) in the synthesised speech that influences the outcome of perception experiments. Having observed (Escudero et al. 2009) a convergence in the production of mid front vowels by young speakers from Lisbon and considering the role perception units play in speech production (Newman 2003; Evans & Iverson 2004; 2007), stimuli were based on a male speaker<sup>3</sup> from Lisbon, maintaining the characteristics of natural speech during the synthesis process. Therefore, speech samples of young people from Lisbon were recorded and an online

---

<sup>3</sup> In Escudero et al. (2009), the convergence of the mid vowels was observed for both sexes. Aiming at making a comparison with the results of Silva and Neves (2016), who investigated the same topic in BP, we chose to use a male voice in this research.

survey was used to select one of the speakers. Finally, the selected speaker was recorded again, and stimuli were created based on his productions of the target vowels.

#### **4.1 Selection of the speaker**

Postulating correlations between production and perception, we selected a speaker with the similar characteristics as those in the study by Escudero et al. (2009), that is, a young male university student that had been residing in Lisbon since birth, to generate stimuli from natural speech. Foreign language proficiency was also controlled for using the same strategy as Escudero et al. (2009): It had to be less than 3 on a scale from 0 to 7, where 0 stands for the lowest proficiency in a foreign language. 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, with a TRUST Mico USB microphone connected to a laptop computer. The recordings were made with Praat (Boersma 2001) 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. Participants were asked to speak spontaneously for 2 minutes on a common topic (football), to obtain a speech sample as natural as possible.

The five recordings were placed separately on the SoundCloud website and links to a Google form, entitled “The vowels of Portuguese from Lisbon” were created. We then conducted an online survey aimed at specialists in the phonetics and/or dialectology of EP, with the aim of selecting an exemplary speech sample from Lisbon. The recordings were classified by 8 experts, through a webpage, on a scale of 1 to 5, with 1 corresponding to “totally disagree” and 5 to “totally agree”, as to the following observations: (Q1) The recording has the typical characteristics of the vowels of the Lisbon dialect; (Q2) the quality of the voice on the recording is pleasant. We selected the speaker with the highest average score and the lowest standard deviation for the two criteria (Q1 average = 4.25 and standard deviation = 0.66; Q2 average = 4.38 and standard deviation = 0.48).

#### **4.2 Recording of the target-vowels**

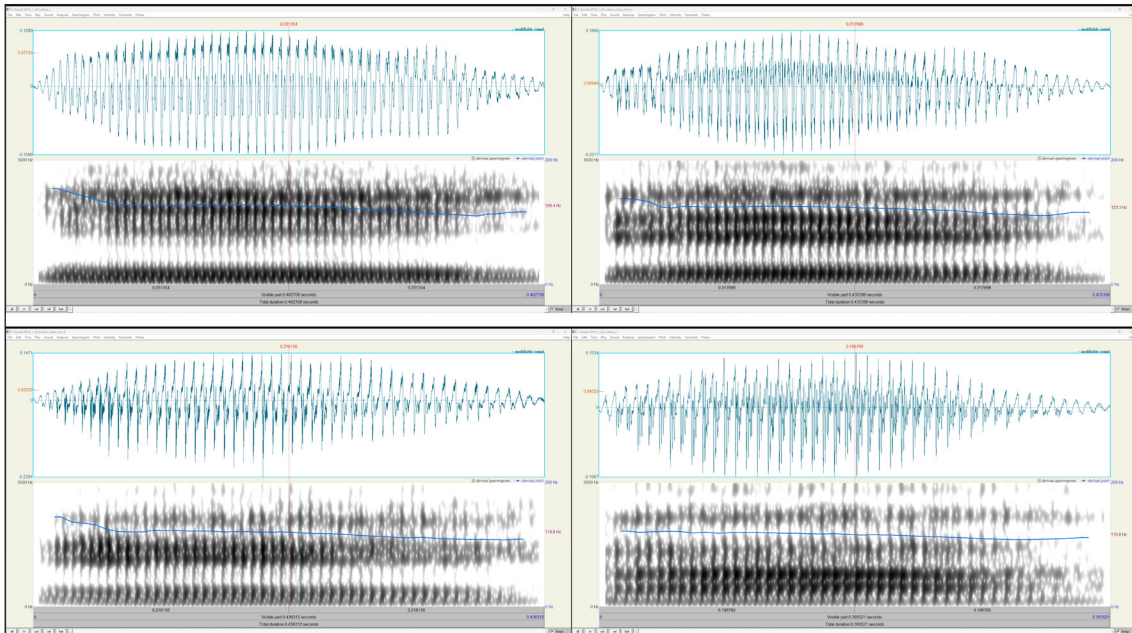
New recordings were made in Lisbon with the selected speaker (21 years of age; university student of the 3rd year of an Undergraduate Degree). We used an AKG Perception 120 USB condenser microphone connected to a laptop computer. The recordings were made with Praat version 6.0.37, at a sampling frequency of 48000 Hz, with 16 bits per sample, and the data stored in mono .wav format (Windows PCM) with no compression. In order to get familiar with the speech materials and to facilitate pronunciation, the speaker was asked to produce the following two-syllable words containing the vowel under study in the tonic syllable: < pico >, < medo > ,



<teto> and <pato>. After pronouncing the two-syllable words, the speaker produced the target vowels [i, e, ε, a] in isolation three times. We chose to record samples with a drop in tone at the end, a procedure like that used in the study by Silva & Neves (2016), thus seeking to generate the stimuli in the most natural fashion possible. We used a duration of 400 ms for the vowels, like that of the study by Masapollo et al. (2017).

### 4.3 Creation of stimuli

We selected one repetition of the four vowels, with the following criteria: Descent from  $f_0$  at the end; values of the frequency of  $F_1$  and  $F_2$ , as close as possible to those of Escudero et al. (2009). The values of the frequencies of the formants were extracted at midpoint, using the following Praat version 6.0.37 function (parameter values chosen by default in SoundEditor): To Formant (burg)... 0.01 5 5500; 0.025 50; [Time step(s), Max. number of formants, Maximum formant (Hz), Window length(s), Preemphasis from (Hz)] – split Levinson algorithm. The spectrogram (values of the parameters chosen by default in SoundEditor) was also used to check the extracted values. A drop in the  $f_0$  of 20-25 Hz from beginning to end can be observed in **Figure 1**.



**Figure 1:** Waveforms (top), spectrograms (bottom) and  $f_0$  (in blue superimposed on the spectrogram) of the four samples of the oral vowels produced by the male speaker: (top left) sample of [i] with a duration of 463 ms,  $f_0$  123 Hz,  $F_1$  319 Hz,  $F_2$  2245 Hz and  $F_3$  2891 Hz. (top right) sample of [e] with a duration of 440 ms,  $f_0$  120 Hz,  $F_1$  376 Hz,  $F_2$  1918 Hz and  $F_3$  2599 Hz. (bottom left) sample of [ε] with a duration of 440 ms,  $f_0$  117 Hz,  $F_1$  440 Hz,  $F_2$  1860 Hz and  $F_3$  2391 Hz. (bottom right) sample of [a] with a duration of 440 ms,  $f_0$  114 Hz,  $F_1$  709 Hz,  $F_2$  1261 Hz and  $F_3$  2253 Hz.

These four samples were used as anchors to generate 31 stimuli with the morphing procedure in the TANDEM-STRAIGHT (Kawahara & Morise 2011) 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 stimulus001, stimulus011, stimulus021 and stimulus031 correspond to the anchors of the vowels /i/, /e/, /ɛ/ and /a/, respectively. In Table A.1 of Appendix A, it can be seen that, as the value of  $F_1$  increases, the values of  $F_2$  and  $F_3$  decrease, thus reflecting the characteristics of the stimulus filter.<sup>4</sup>

#### 4.4 Perception tests and participants' characteristics

A questionnaire regarding the biographical information and linguistic profile (appendixes B and C) was used to select a relatively homogeneous group, comparable with the group of EP speakers in the study by Escudero et al. (2009). Therefore, the participants were selected according to the following inclusion criteria: EP as L1; Portuguese parents; born in Lisbon and having resided in the same city all their lives; no reported hearing impairment; young adult; degree of academic qualifications above undergraduate or attendance of an undergraduate degree; no phonetic training. Regarding the linguistic profile, the participants self-assessed<sup>5</sup> their proficiency in foreign language(s), in accordance with the Common European Framework of Reference for Languages (CEFR). All participants of the pre-test and the main test signed an informed consent.

##### 4.4.1 Pre-test

The pre-test was used to verify the experimental procedures and if there was any issue with the stimuli generated. In this pre-test, the area of residence was not considered and only the origin of the participants was questioned. We obtained 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), who met the inclusion criteria, participated voluntarily in the pre-test.

The pre-test was used to debug the Praat scripts that were developed for the project, to optimise the data extraction procedures and to estimate the total duration of each experiment. It was also used to explore the possibility of dialectal specificity, i.e., results based on speakers from the North of Portugal could be compared with those of Lisbon that participated in the main test.

---

<sup>4</sup> 15% (14/93) of the formant values, automatically obtained, were manually corrected, based on the spectrogram and the LPC spectrum.

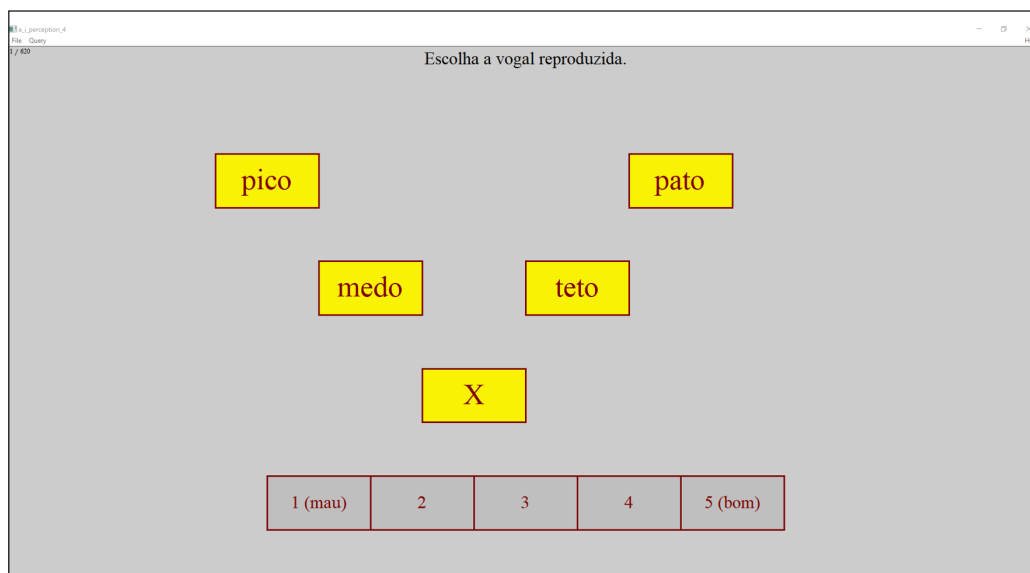
<sup>5</sup> English C2=1, C1=3, B1=1, A2=1, Japanese A2=2, A1=1 in the pre-test and, English C2=5, C1=3, B2=1, A2=1, English C1 and Spanish B1=1, English B1 and Spanish B1=1, English C1 and French A1=1 in the main test (level=no. of participants).

#### 4.4.2 Main test

Fifteen (females = 9) undergraduate and postgraduate participants, EP speakers from Lisbon, aged between 19 and 34 (average = 23.7 years of age) who met the inclusion criteria mentioned above, 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. Stimulus presentation was self-paced, there was no possibility of repetition and lasted an average of 76 minutes. The experiment required two responses: 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). A practice test was carried out with 21 continuous stimuli of /u/-/o/ vowels not targeted in the study, in a random order and with no repetition, to determine if the participants understood the procedure. In this practice test, the volume of the sounds was individually adjusted.

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 (see **Figure 2**).



**Figure 2:** Screen capture of the identification and goodness rating task.

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.

## 4.5 Data processing

All graphs and statistical analysis in this paper were generated using R (RCoreTeam 2023) version 4.3.1 running in RStudio (RStudioTeam 2023) 2023.06.1 + 524 Release and Excel 2016. The individual responses of listeners were modelled using logistic regression analysis. Three mixed logistic regression models of all responses were developed (one for each boundary /i/-/e/, /e/-/ε/ and /ε/-/a/) using the `glmer` function from the `lme4` version 1.1-31 package, with `correct/incorrect` category answers as outcome variable, considering `stimulus` number as a fixed effect, and `listener` as random effect. The resulting models' regression lines and grey shading spanning the 95% confidence interval were drawn using the `sjPlot` 2.8.14 package.

Six mixed effects regression models were developed (from the centroid of each category to the boundary with the neighbouring category) using the `lmer` function also from the `lme4` version 1.1-31 package, with the `goodness` score as outcome variable, considering `stimulus` number as a fixed effect, and `listener` as random effect. Results from likelihood ratio tests of the model with the `stimulus` effect against the model without the `stimulus` effect were also analysed.

## 5 Results

This section presents the analysis of the results collected throughout the study. We begin with the results obtained in the identification task and then with the data from the goodness rating task, to address the two hypotheses of our study: H1 – There are four phonemic categories with a gradient structure; H2 – The gradient of the boundary between the mid vowels is smoother than other boundaries.

### 5.1 Phonemic categories

The first research hypothesis (H1) has to do with the existence of speech perception boundaries between the phonemic categories /i/-/e/, /e/-/ε/ and /ε/-/a/. In order to determine the boundaries between two categories, data from the identification task were used, in which participants categorised 31 stimuli repeated 20 times, in a random order (620 stimuli in total), into four vowel categories. They could also choose option X should they decide that the sound heard did not belong to any category. The total number of valid responses was: Boundary 1 – 3816 / 4042 (94 %); boundary 2– 5157 / 5661 (91 %); boundary 3 – 3735 / 3815 (98 %).

The data obtained in this task was initially explored using logistic regression analysis, as used in a previous study of Portuguese by Silva & Neves (2016)<sup>6</sup>. The logistic regression model was chosen because, first, the answers to the identification task are not continuous variables, but categorical variables. Secondly, by processing the identification data as binary data (that is, there are only two possible values for the answer – 0 if it belongs to a category and 1 if it does not) we are able to represent, with this model, the probability of the transition of a category to another with high robustness. This model also facilitates the interpretation of the results and highlights the individual variations (idiosyncrasies between listeners).

The logistic regression curves were used to estimate the boundary between two categories, which corresponds to the probability of response  $p = 0.5$ . This boundary point can be found using the logistic regression equation, when  $p$  takes the value 0.5:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$$

in which the parameter  $p$  represents the probability of the response not falling into a given category,  $x$  is the independent variable that corresponds to the identity represented by the number of stimuli, and  $\beta_0$  and  $\beta_1$  are the ordinate at the origin and the slope, respectively (Silva & Neves 2016). In our identification task, there were five options to choose from, including X if there was no correspondence with the sound heard. That way,  $p$  does not represent the probability of one of the alternative answers, as used in Silva & Neves (2016), but the probability of not being a given answer. Thus, the value of  $x$  which corresponds to the boundary between categories was calculated by replacing  $p$  with 0.5 according to:

$$\ln\left(\frac{0.5}{1-0.5}\right) = \beta_0 + \beta_1 x_{boundary}$$

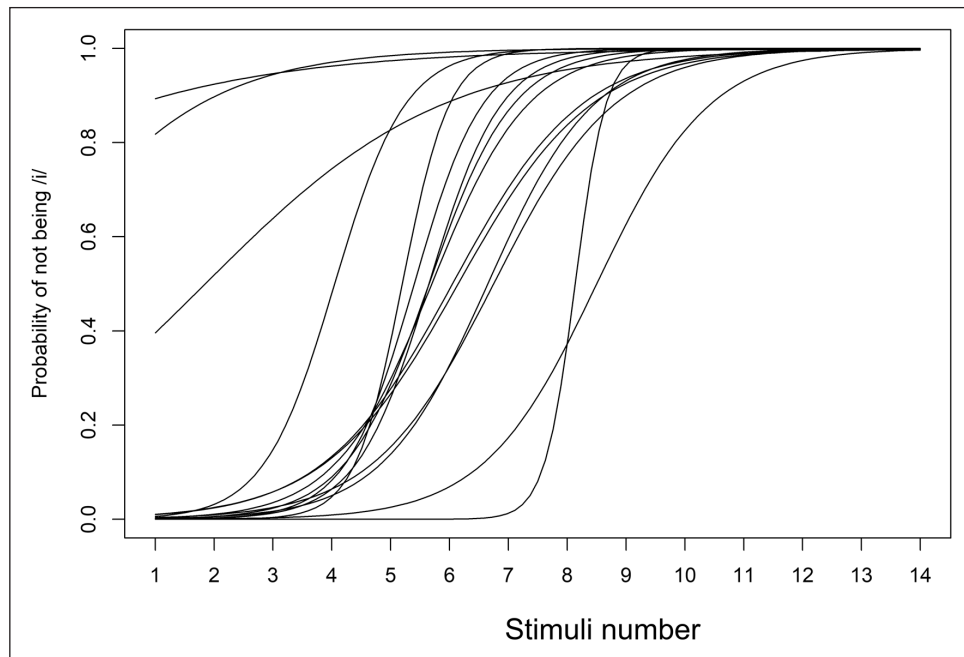
$$\Leftrightarrow 0 = \beta_0 + \beta_1 x_{boundary}$$

$$\Leftrightarrow x_{boundary} = -\frac{\beta_0}{\beta_1}$$

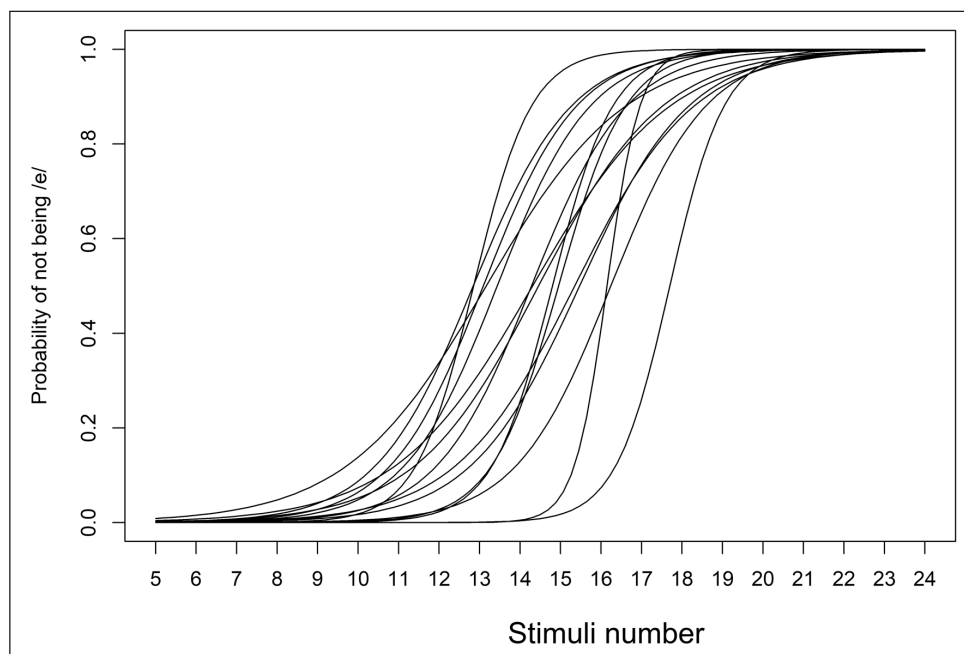
Three boundaries between categories were estimated for each participant. In **Figures 3, 4 and 5** the results of the logistic regression of all participants for the transition of the categories /i/, /e/, /ε/ and /a/, can be observed.

---

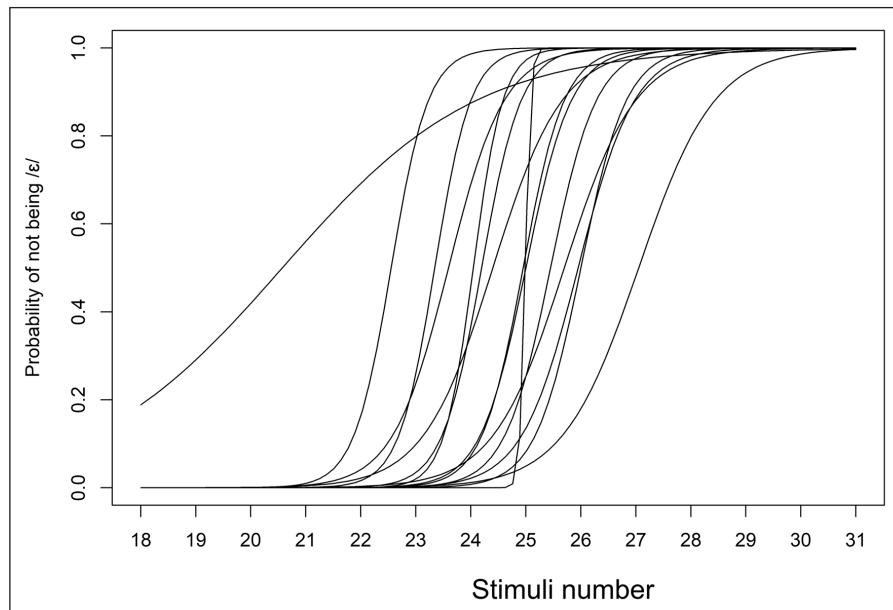
<sup>6</sup> The tasks used in Silva & Neves (2016) were the 2AFC (forced choice between two options) between /i/ and /e/, and between /e/ and /ε/.



**Figure 3:** The logistic regression curves of all participants for the transition from the /i/ to non-/i/ category.



**Figure 4:** The logistic regression curves of all participants for the transition from the /e/ to non-/e/ category.



**Figure 5:** The logistic regression curves of all participants for the transition from the /ε/ to non-/ε/ category.

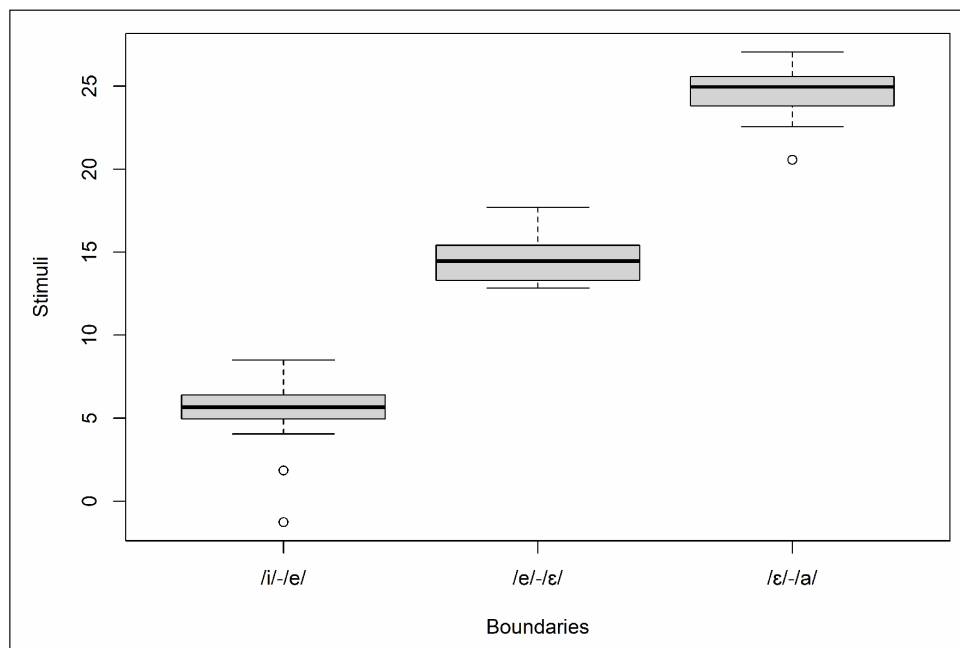
The estimates shown in **Table 1** (mean and median values, with standard deviation in parentheses) for the three boundaries were obtained from the logistic regression curves adjusted for all participants: /i/-/e/ (hereinafter boundary 1), /e/-/ε/ (boundary 2) and /ε/-/a/ (boundary 3). The formant frequency values of the stimuli which correspond to each boundary are shown in the last column of **Table 1**.

	<b>Boundaries mean (<math>\pm</math> standard deviation)</b>	<b>Boundaries median</b>	<b>Formants<sup>7</sup></b>	
/i/-/e/ (boundary 1)	5.28 ( $\pm$ 2.40)	5.65	$F_1$ : $F_2$ : $F_3$ :	344 Hz 2049 Hz 2754 Hz
/e/-/ε/ (boundary 2)	14.61 ( $\pm$ 1.42)	14.46	$F_1$ : $F_2$ : $F_3$ :	406 Hz 1893 Hz 2503 Hz
/ε/-/a/ (boundary 3)	24.52 ( $\pm$ 1.59)	24.95	$F_1$ : $F_2$ : $F_3$ :	551 Hz 1648 Hz 2338 Hz

**Table 1:** Estimated mean boundaries between categories and corresponding formant frequency values.

<sup>7</sup> The values of the formants correspond to the values of the stimuli closest to the medians. Therefore, boundary 1  $\cong$  6, boundary 2  $\cong$  15 and boundary 3  $\cong$  25.

The distributions of these values are represented in the boxplots shown in **Figure 6**. A clear distinction between the three boundaries can be seen in this representation. The reduced variability of the three groups, particularly the group corresponding to boundary 2, which does not exhibit outliers<sup>8</sup> should be noted. Regarding boundary 1, two participants behaved quite differently from the rest of the group. As far as boundary 3 is concerned, there is only one participant with a unique behaviour.



**Figure 6:** The distribution of the values of the three boundaries, /i/-/e/, /e/-/ε/ and /ε/-/a/ among the 15 participants.

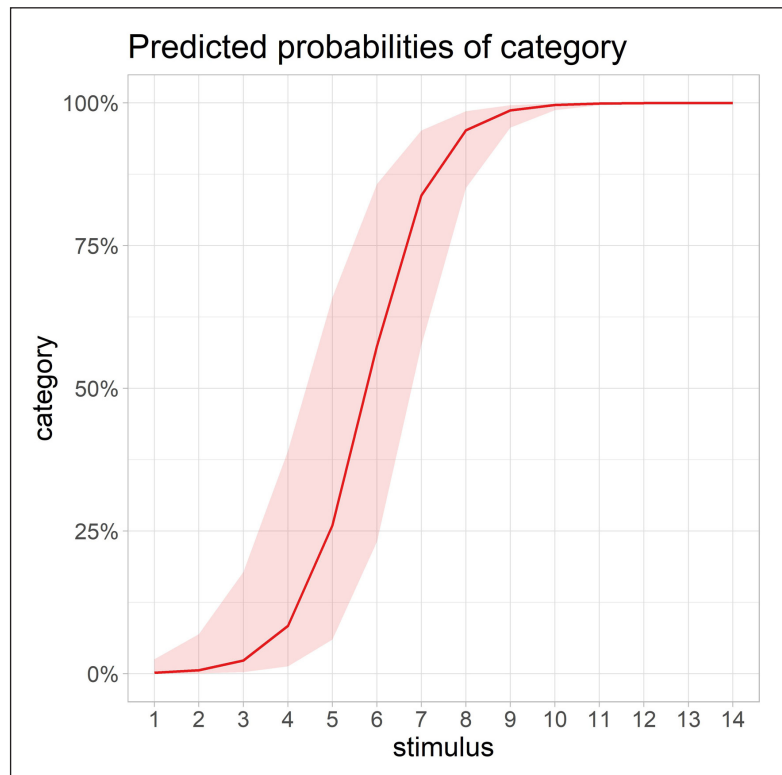
Three additional mixed logistic regression models were developed (one for each boundary /i/-/e/, /e/-/ε/ and /ε/-/a/), satisfying the normality assumption (i.e., its residuals were approximately normally distributed) and the constant variance assumption (homoscedasticity), as assessed by the following visual diagnostics plots: Histogram of residuals; Q-Q plots of residuals; residuals plot.

A mixed logistic regression model of boundary 1 (/i/-/e/) with the lme4 syntax `category ~ stimulus + (1+stimulus|listener)` revealed a SD of 5.7845 for the by-listener varying intercepts. The estimated correlation between varying intercepts and varying slopes

<sup>8</sup> Regarding the two outlier observations of the first set of boundaries, we would have: For boundary /i/-/e/, median = 5.70, /e/-/ε/ = 14.46 and /ε/-/a/ = 24.95, which does not differ much from what is obtained with the inclusion of these two values (-1.26 and 1.85) – see Table 1. We, therefore, chose to keep them in our analysis. In addition, these data provided us with interesting information on the category /i/ which we will discuss below.



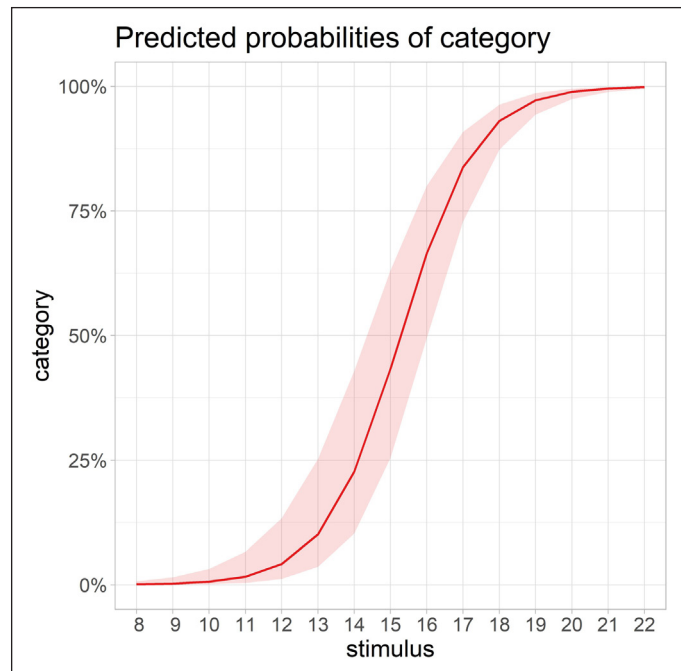
showed that higher intercepts had lower stimulus slopes ( $r = -0.92$ ) for the listener random effect. The resulting model, represented in **Figure 7**, predicted that listeners' choice of category changed from /i/ to /e/ (responses were above chance level) at stimulus 6.



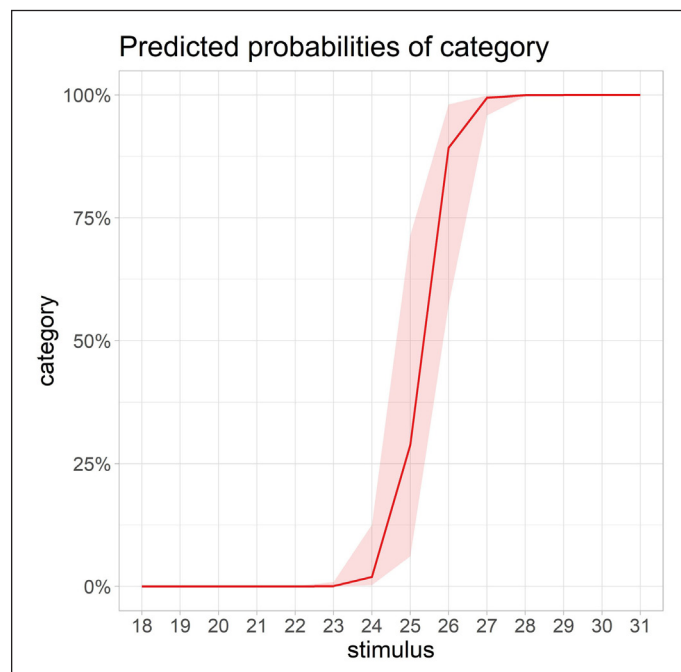
**Figure 7:** Boundary 1(/i/-/e/) model predicted probabilities of correct /e/ responses. The mixed logistic regression model line and shading spanning the 95% confidence interval are shown.

An additional mixed logistic regression model of boundary 2 (/e/-/ε/), with the same lme4 syntax, revealed a SD of 6.7616 for the by-listener varying intercepts. The estimated correlation between varying intercepts and varying slopes also showed that higher intercepts had lower stimulus slopes ( $r = -0.98$ ) for listener random effect. The resulting model, represented in **Figure 8**, predicted that listeners choice of category changed from /e/ to /ε/ at stimulus 16.

A final mixed logistic regression model of boundary 3 (/ε/-/a/), also with the lme4 syntax `category ~ stimulus + (1+stimulus|listener)`, revealed a SD of 15.5863 for the by-listener varying intercepts. The estimated correlation between varying intercepts and varying slopes also showed that higher intercepts had lower stimulus slopes ( $r = -0.98$ ) for listener random effects. The resulting model, represented in **Figure 9**, predicted that listeners choice of category changed from /ε/ to /a/ at stimulus 25.



**Figure 8:** Boundary 2(/e/-/ε/) model predicted probabilities of correct /ε/ responses. The mixed logistic regression model line and shading spanning the 95% confidence interval are shown.

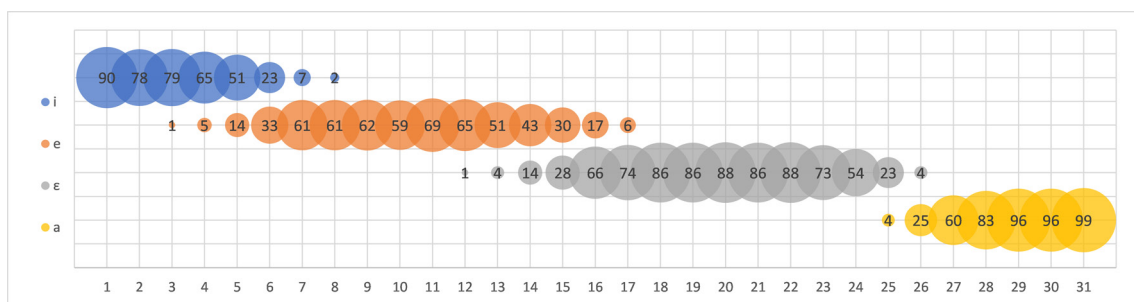


**Figure 9:** Boundary 3(ε /-/a/) model predicted probabilities of correct /a/ responses. The mixed logistic regression model line and shading spanning the 95% confidence interval are shown.

Taking these modelling results into account, we could also observe four vowel categories divided by three different boundaries.

## 5.2 Internal structure of phonemic categories and prototypes

After the identification task, the participants rated the quality of the sound heard on a scale of 1 to 5 according to the vowel chosen in the identification task. Therefore, a stimulus could receive a maximum of 100 points for each participant, should this be ranked 20 times as the best example (5 points). The points added from the goodness rating (1 to 5) of each stimulus were calculated for each participant and represented in a bubble chart. For the representation of the group (all participants), the median value was used instead of the mean (see **Figure 10**), since the result of a normality test showed that the group of observations of the boundary between /i/ and /e/ did not follow a normal distribution, and, in this case, the median is the central tendency measure that best represents the data.



**Figure 10:** 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). The colour is determined by the individual listener's judgment in the identification task.

The second hypothesis (H2) of this study has to do with the existence of a gradient internal structure of the category and observing a sound (or sounds) representative of each phonemic category, the prototype. As shown in **Figure 10**, a category structure in which the prototype is in the centroid with the highest score, is observed. As the stimulus moves away from the centroid, the goodness score decreases. It should be noted that not all participants responded in the same manner, that is, there was great individual variability in the identification of the category of stimuli and, consequently, great variability in individual prototypes, as shown, for example, in **Figures 11** and **12**. However, the listeners, in general, used four categories with three boundaries, also exhibiting an internal structure of each category, with their prototype.

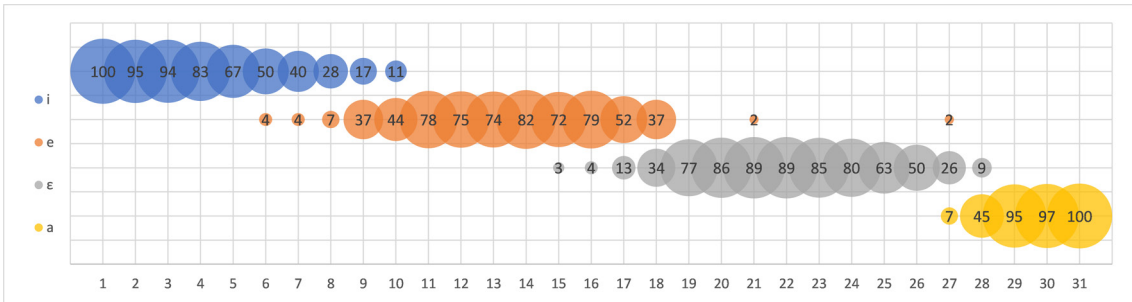


Figure 11: Goodness rating of listener CC.

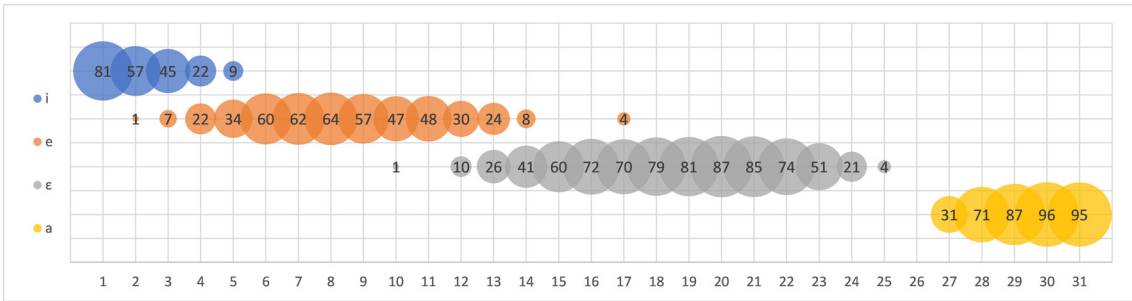


Figure 12: Goodness rating of listener LP.

There were two listeners (BM and AC) who performed very differently from the others in identifying the category /i/, also referred to in the results of the identification task (footnote 8). These data, however, were not excluded, as the internal structure pattern was observed in other categories (as shown in Figures 13 and 14). These data lead us to conjecture a more peripheral location of the /i/ prototype for these listeners. We shall discuss this issue in section 6.

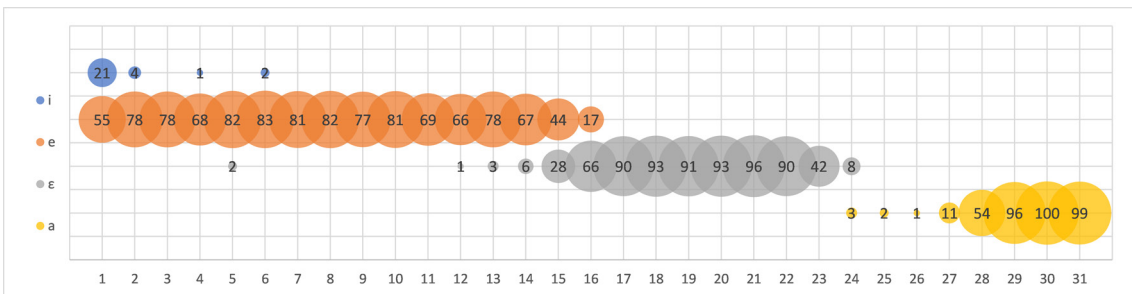


Figure 13: Goodness rating of listener BM.

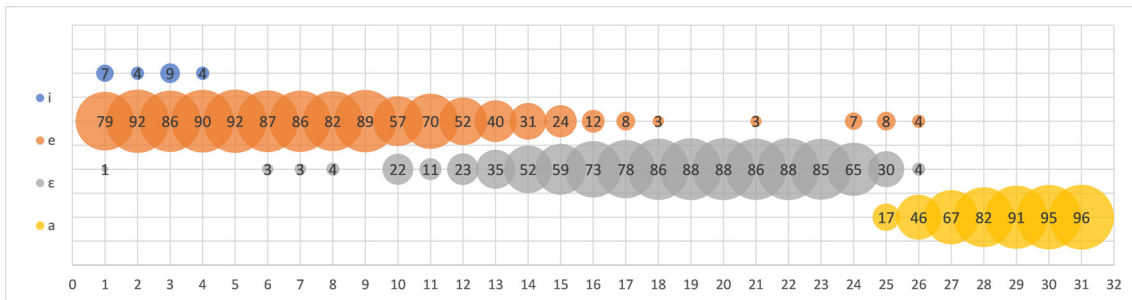
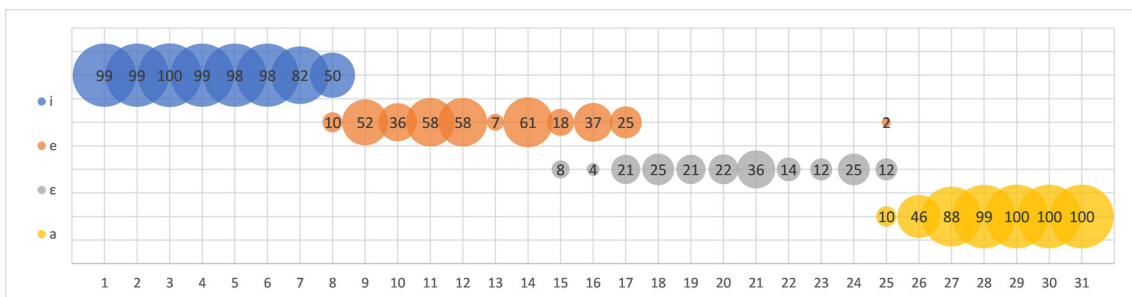
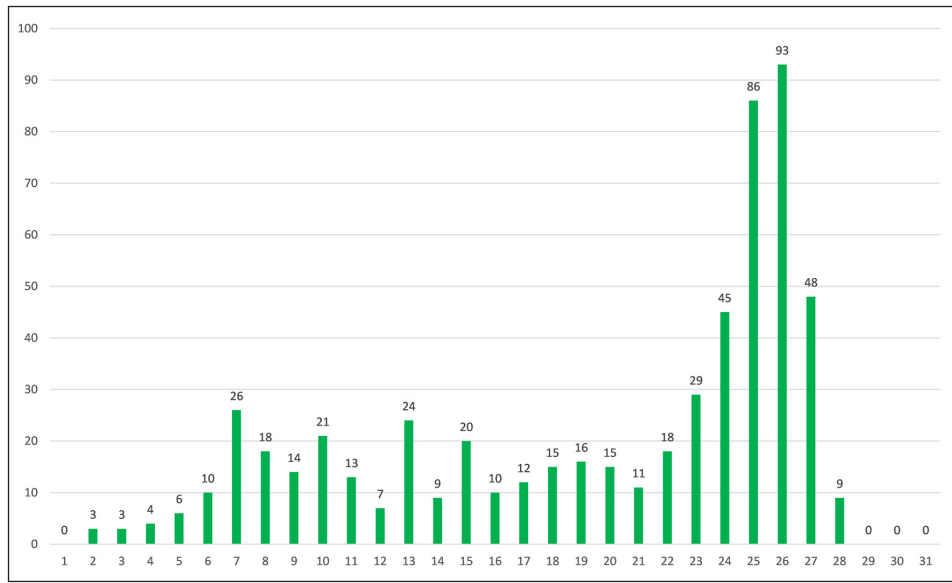


Figure 14: Goodness rating of listener AC.

It should also be noted that the listener CL seems to show some hesitation in the goodness rating for the categories of /e/ and /ɛ/, but not for those of /i/ and /a/ (see Figure 15). Although this is the only listener who has behaved in this manner, it will be nevertheless interesting to discuss the characteristics of these categories which would cause such phenomenon.



did not belong to any category. The presence of many X responses among stimuli representing open-mid and open vowels suggests that, in the perceptual space, / $\epsilon$ / and / $a$ / are far apart. In fact, they are perceptually distant enough that tokens falling between these two vowels are not acceptable examples of either category.



**Figure 16:** The values represented on the y-axis correspond to the absolute frequency of X responses for each of the 31 stimuli on the horizontal axis.

There is a conflict between two categories during the identification process, for vowels that overlap in perceptual space. To analyse it, one could also compare, the goodness scores as they evolve from the prototype of a certain category to the boundary of the “neighbouring” category and from the same boundary to the prototype of its “neighbouring” category. The three individual boundaries between categories were considered as the median of the 15 listeners’ boundaries (estimated using the models described in section 5.1) as shown in **Table 2**. The listeners are ordered from lowest to highest stimulus value for the first/ second boundary, with the following pattern emerging, which must be investigated further: As the first boundary number increases the other two boundary numbers also increase (move along with it).

Based on this assumption, the transitions between categories were studied using mixed effects regression models, and the slopes of overlapping categories analysed, using the goodness ratings values from the 15 listeners, which could be very distant from the average or median values of the boundaries.<sup>9</sup> We expected that less steep slopes would be observed where two

<sup>9</sup> The individual variability relative to median boundaries may be due to the reduced sample size (only 15 participants).

categories overlap because, for some listeners, there were stimuli around the boundary that still represented a “good example” of the vowel; during the goodness rating task listeners provided a score between 1 and 5 for each stimulus, where 1 represented a bad example and 5 a good example of the vowel.

Listener	/i/-/e/	/e/-/ε/	/ε/-/a/
BM	-1	15	23
MM	2	13	24
LP	4	13	24
AC	5	13	25
CP	5	13	25
IM	5	15	26
AP	6	13	24
MF	6	14	24
BR	6	15	26
GS	6	16	25
FM	6	16	26
ML	7	15	23
MC	7	16	26
CL	8	14	21
CC	9	18	27
Median	6	15	25

**Table 2:** Individual values and the medians of boundaries’ stimulus numbers.

Mixed effects regression models  $\text{goodness} \sim \text{stimulus} + (1|\text{listener})$  predicted a negative effect of `stimulus` in `goodness` values, i.e., that as one progressed from stimulus 1 to 6 (from the /i/ prototype to boundary 1) the `goodness` values decreased (slope  $m_1 = -0.20$ ). Likelihood ratio tests of the model with the `stimulus` effect against the model without the `stimulus` effect revealed a significant difference between models, i.e., there was a significant difference between `goodness` values of stimuli from the two categories:  $\chi^2(1) = 230.97, p < 2.2 \times 10^{-16}$ .

From stimulus 6 to 11 (from boundary 1 to the /e/ prototype) the `goodness` values were very stable around 3.5 ( $m_2 = -0.03$ ). Likelihood ratio tests of the model with the `stimulus` effect against the model without the `stimulus` effect revealed non-significant difference between models, i.e., there was not a significant difference between `goodness` values of stimuli from the two categories:  $\chi^2(1) = 4.00, p = 0.05$ .

From stimulus 11 to 15 (from the /e/ prototype to boundary 2) the `goodness` values were also very stable around 3.5 ( $m3 = 0.02$ ). Likelihood ratio tests of the model with the `stimulus` effect against the model without the `stimulus` effect also revealed a non-significant difference between models, i.e., there was not a significant difference between `goodness` values of stimuli from the two categories:  $\chi^2(1) = 1.55, p = 0.21$ .

From stimulus 15 to 20 (from boundary 2 to the /e/ prototype) the `goodness` values increased ( $m4 = 0.18$ ). Likelihood ratio tests of the model with the `stimulus` effect against the model without the `stimulus` effect also revealed a significant difference between models, i.e., there was a significant difference between `goodness` values of stimuli from the two categories:  $\chi^2(1) = 266.23, p < 2.2 \times 10^{-16}$ .

From stimulus 20 to 25 (from the /e/ prototype to boundary 3) the `goodness` values decreased ( $m5 = -0.34$ ). Likelihood ratio tests of the model with the `stimulus` effect against the model without the `stimulus` effect, revealed, once again, a significant difference between models, i.e., there was a significant difference between `goodness` values of stimuli from the two categories:  $\chi^2(1) = 591.97, p < 2.2 \times 10^{-16}$ .

From stimulus 25 to 31 (from boundary 3 to the /a/ prototype) the `goodness` values increased ( $m6 = 0.38$ ). Likelihood ratio tests of the model with the `stimulus` effect against the model without the `stimulus` effect, revealed, yet another significant difference between models, i.e., there was a significant difference between `goodness` values of stimuli from the two categories:  $\chi^2(1) = 1086.11, p < 2.2 \times 10^{-16}$ .

We finally analysed the gradient of the boundaries, by calculating the absolute difference between the slopes of the regression lines, based on the following assumption: In the process of identifying the continuum of stimuli, there is a conflict at the boundary between two categories. The lower the absolute difference between the slopes for a certain boundary is, the more confusing the identification of the stimuli will be. As shown in **Table 3**, by comparing the three boundaries, it can be concluded that boundaries 1 and 2 are the ones with the lowest values, which suggests a greater degree of confusion. The value for boundary 3, suggests a lower degree of confusion than in the previous two.

	Absolute frequency of X responses	Slopes absolute difference
Boundary 1	10	$ m1-m2  =  -0.20+0.03  = 0.17$
Boundary 2	20	$ m3-m4  =  0.02-0.18  = 0.16$
Boundary 3	86	$ m5-m6  =  -0.34-0.38  = 0.72$

**Table 3:** Absolute frequency of X responses and slopes absolute difference for the three boundaries.



Considering the numerous analyses carried out, the gradient of boundary 3 was the highest. This corresponded to the lowest of the slopes' absolute difference and to the highest number of X responses as shown in **Table 3**. Therefore, there was a considerably higher degree of confusion in the first two boundaries than in the third.

There were also very stable and lower values of goodness scores observed between boundaries 1 and 2 (slopes of the models' regressing lines very close to zero), that correspond to the phoneme category /e/, and a maximum goodness rating of "only" 69 (see **Figure 10**).

## 6 Discussion

In this section we discuss the results obtained in our study, considering the hypotheses (H1 and H2) initially formulated. We discuss whether the boundary of Portuguese mid front vowels (boundary 2) is less steep than the other boundaries studied (boundaries 1 and 3), due to the complexity of the relationships that these vowels establish within EP phonological system. The goal was to relate our results to those of previous studies and theories, to provide a new body of experimental evidence which corroborates or rejects the formulated hypotheses.

### 6.1 The four vowel categories

Regarding the first question (Q1), results from the identification task showed four vowel categories marked by three distinct boundaries. This is clear evidence that young (age range of 19 to 34 years) speakers from Lisbon distinguish four vowel categories in the acoustic stimuli used in this study. Therefore, it is assumed that EP speakers have these four categories in their mental representation. These results correspond to the phonological descriptions of the vowel system in EP (Mateus & Andrade 2000; Veloso 2016; Andrade 2020), in which there are four front/ central phonemes /i, e, ε, a/. This "compliance" with phonological descriptions has also been reported in previous speech production studies (Escudero et al. 2009).<sup>10</sup> Therefore, the phonological description and speech production and perception results, are all in agreement regarding this issue. In addition, as reported in the study by Aaltonen et al. (1997), we observed great individual variability at the boundaries. According to Repp and Liberman (1987), in addition to numerous methodological<sup>11</sup> factors (linguistic variables/ phonetic context; individual

---

<sup>10</sup> Escudero et al. (2009) reported results from an analysis of variance with repeated measures, where the main effect was that of  $F_1$  for the vowel categories of both varieties (EP and BP) of Portuguese. This fact agrees with the proposal by Veloso (2016) for EP and Wetzels (1992) for BP, that is, the phonological vowels of Portuguese are opposed by the 4 degrees of openness. In this study, we did not carry out a detailed analysis to reach the same conclusion, by assuming the existence of the 4 distinct categories in the mental representation of the speakers of the dialect in question.

<sup>11</sup> The authors provided examples of methodological factors that may influence the location of the boundary, such as the phonetic context, speech rate, mixture of acoustic cues and linguistic experience (Repp & Liberman, 1987).

variables/ language experience), the location of the boundary between phonetic categories in the study of speech perception may be influenced by individual factors such as the different mental representation of the categories and/ or a different strategy when identifying stimuli (Aaltonen et al. 1997; McMurray 2022).

## 6.2 The internal structure of the vowel categories

Regarding the second question (Q2), we were able to observe an internal structure of the vowel categories through the goodness rating task, in which the stimulus with the highest score, considered to be the prototype of such category, was approximately in the centroid and the goodness score decreased as the stimulus moved away from the centroid. The results obtained were in line with those of previous studies (Grieser & Kuhl 1989; Kuhl 1991; Miller 1994; Frisch 2018; McMurray 2022) which confirmed an internal structure of the phonemic category.<sup>12</sup> In general, the prototypes were found (Savela & Eerola & Aaltonen 2014), in terms of the number of stimuli, more or less in the centre of the category identified by the participants.<sup>13</sup> However, while in Kuhl's (1991) experiment consistent responses were obtained in all participants, our data show individual variability of the prototypes in each category. The individual variation of the prototype was also observed by Lively & Pisoni (1997) and Aaltonen et al. (1997).<sup>14</sup> It would be interesting to establish a comparison between the individual variability of speech production and the prototype, to ascertain the existence of relations among these.

## 6.3 The gradient of boundaries

Hypothesis H2 was tested using the X responses and goodness ratings. The hypothesis (H2) that the /e/-/ε/ boundary is less steep than the /i/-/e/ and /ε/-/a/ boundaries in speakers from Lisbon was supported by the complexity of the relation that these vowels establish between the underlying forms and their phonetic realisations (Veloso 2016), pointed out as the potential

---

<sup>12</sup> An exploration of the “*perceptual magnet effect*” theory (Iverson & Kuhl, 1995; Kuhl, 1991; Kuhl et al., 1992) did not fit the goals of this study.

<sup>13</sup> In the literature on speech perception there are two different proposals regarding the location of the prototype: The most peripheral location in the category in, for example, the dispersion theory (Johnson, 2000; Johnson et al., 1993), or the centroid (Kuhl et al., 1992). The results obtained by Savela et al. (2014) support the theory of Johnson et al. (1993) and Johnson (2000), by verifying that the location of the perceptual prototype is more peripheral than the arithmetic mean of the category. As the stimuli used in our study do not cover the entire space of a category, we will consider an internal structure that includes the prototype, whose score is the highest, and which is, in terms of the stimuli number, approximately at the centre.

<sup>14</sup> In the study by Savela et al. (2014), the three different measures of the prototypes were assessed using the goodness rating: The absolute prototype; the centroid of the prototype and the weighted prototype, to structure the internal vowel category of German and Finnish. The specificity of each language and a greater individual variability in terms of the location of the prototype were confirmed with the absolute prototype methodology. On the other hand, less variability between subjects and even between languages was obtained in both languages with the weighted prototype methodology.

difficulty in the acquisition of these phonological relationships by children (Freitas 2003) and the observation of the convergence between the mid vowels in the production of the speakers of this dialect (Escudero et al. 2009; Segura 2013; Andrade 2020). There is recent evidence (Tiegs 2023) for a perceptual difference in the mid-vowel contrasts from other vowels, namely, “lower perceptual sensitivity and more difficulty in accessing phonological and lexical representations when dealing with mid-vowel contrasts as compared to the more robust contrasts involving point vowels”.

The differences between the goodness ratings of the two categories at each boundary were analysed to assess the degree of confusion in the identification of vowels, based on the individual boundaries marked in the identification task.

The boundary between / $\epsilon$ / and /a/ was consistently shown to be steeper than the other two boundaries. The result of the analysis of the absolute frequency of X responses support the hypothesis that / $\epsilon$ / and /a/ are perceptually distant enough that tokens falling between these two vowels are not acceptable examples of either category. The highest absolute difference between the slopes of mixed effects regression models' lines was observed at the / $\epsilon$ -/a/ boundary, also indicating that these are the least confusing stimuli.

Goodness ratings modelled using mixed effects regressions at the transitions of overlapping categories predicted the lowest values of goodness scores at boundaries 1 and 2. The highest median goodness rating of the /e/ category stimuli was 69, that could be interpreted as evidence of a less well-defined prototype of this category, resulting from an ongoing neutralisation processes of the underlying oppositions.

It should also be noted that, in EP, / $\epsilon$ / is different from /a/ in terms of height and fronting (/ $\epsilon$ / is higher and more fronted than /a/), whereas /i/ and /e/ are only distinct in terms of height, with expected impact in the way listeners distinguish the two pairs of sounds.

Regarding /i/-/e/ and /e/-/ $\epsilon$ / boundaries, deciding which of the two boundaries has the greatest gradient depends on the method used for the analysis. The study by Silva & Neves (2016), which addressed the categorisation of the mid front vowels of BP, showed a better distinction between /i/ and /e/ than between /e/ and / $\epsilon$ /. Our experimental results show that there is not a significant difference between the gradient of the two boundaries, which differs from the results reported by Silva & Neves (2016). This difference may be due to the following factors: Specificities of the two varieties; the distinct methodologies used in each study.

The first factor has been identified in different phonological and language acquisition studies. Fikkert (2005) and Lee (2010) presented a proposal regarding the order of construction of the vowel system for each variety based on the Contrastive Hierarchy Theory (Dresher 2003a; 2003b). According to their proposals, the gradual construction of the vowel oppositions of two

varieties is the same.<sup>15</sup> Fikkert (2005) bases her arguments on two Portuguese children, that produced variations in [i] and [ɛ] height. As for Brazilian children, Bonilha (2004) observed a later acquisition of open-mid vowels.<sup>16</sup> What differs between the two varieties is the phonetic representation at the surface level. That is, each variety has distinct phonological processes: The process of neutralisation of the mid vowels is a clear example (Bisol & Veloso 2016). Therefore, although the two varieties of Portuguese apparently have the same phonological process, the different types of neutralisations of the mid vowels can affect the perception of L1 listeners of each variety in a different fashion.

It is pertinent to mention the participants' comments, collected after each experiment. Some participants mentioned the difficulty in distinguishing /i/ from /e/. For example, subject LP mentioned that the duration of the vowel /i/ (in the stressed position in Portuguese) is never as long as in our stimuli, so sometimes it caused a "feeling of strangeness".<sup>17</sup>

Regarding the second factor, we opted for experimental methodologies which include the prototype theory standpoint, that is, the goodness rating, while Silva & Neves (2016) used methodologies based on categorical perception (McMurray 2022). Regarding the creation of stimuli, Silva & Neves (2016) resorted to the KLSYN88 synthesizer (Klatt & Klatt 1990), thus manipulating the values of the following parameters:  $f_o$ ,  $F_1$ ,  $F_2$ ,  $F_3$  and duration. In other words, their stimuli have no acoustic information beyond these parameters. In turn, the stimuli of this research preserves, as much as possible, the acoustic information of natural speech.

#### 6.4 Influence of phonological relations on speech perception

The results of this research support Trubetzkoy's (1939) assumption, that is, speech perception is affected not only by the presence and absence of sounds in the listener's L1 system, but also by the relationships that the sounds establish within the system. Our results support the assumption that boundary 2 (/e/-/ɛ/), the underlying opposition that neutralises at surface level, was less clear than boundary 3 (/ɛ/-/a/), the opposition that never neutralises. Boundary 1 (/i/-/e/) was also less steep than boundary 3 in the perception of young speakers from Lisbon, having raised the possibility, with the current results, it is a less marked transition. Considering the neutralisation of the oppositions of openness in the absolute initial position (Andrade 2020: 3282–3285), this perceptual phenomenon at boundary 1 can be the result of such phonological processes (Vigário

---

<sup>15</sup> Fikkert (2005) proposed place features for the first dichotomies, while Lee (2010) used features that describe the position of the tongue. Despite the proposals to define the outline in the dichotomies being different between the two varieties, the gradual construction is the same.

<sup>16</sup> Compared to consonants, vowel acquisition is relatively precocious. In BP, children acquire /a/ and /i, u/ first (Bonilha, 2004). In a second phase (1:2), the close-mid /e, o/ vowels are acquired, and finally the open-mid /ɛ, ɔ/ vowels.

<sup>17</sup> The performance of subject LP is reflected on his goodness scores as shown in Figure 12.

2022: 844–848). Also, the centralisation of /i/ in the last unstressed syllable of some verb forms (Vigário 2003: 69–70) could have contributed to a blurred perception of the /i/-/e/ boundary. In two further contexts unstressed /i/ patterns with non-high, non-back vowels, namely, when followed by a palatal fricative and when followed by a syllable headed by /i/. These phenomena may also be contributing factors for a blurrier perception of the /i/-/e/ boundary.

Regarding the dialectal specificities of EP, the study by Rodrigues & Martins (1999) on the acoustic space of the stressed vowels of the speakers from Braga revealed a broader vowel triangle, compared to the dialect of Lisbon, and distinct acoustic ( $F_1/F_2$ ) clusters. This acoustic difference between dialects may have consequences in the categorisation of the speakers of each dialect. Therefore, it is also important to consider the factor dialect (Segura 2013), so the question regarding boundary 1 remains open.

## 6.5 Perceptual hyperspace and adaptative dispersion

Two participants (BM and AC) hardly classified the /i/ stimuli with a high score. We could have considered these outlier data, but the goodness scores of the other categories revealed a sound/normal structure, with the prototypes in the centroid. Therefore, these subjects seem to prefer the sounds of the more peripheral category /i/, that is, the category /i/ that is not included in the continuum of our stimuli.

Johnson et al. (1993) and Johnson (2000) observed a tendency for the listeners, in their speech perception task, to choose peripheral vowel sounds, that is, the stimuli with more extreme values than stimuli with values closer to the vowels produced by their listeners. The same phenomenon was reported by Lively and Pisoni (1997) in a replication of Kuhl's (1992) experiment. In this experiment, participants did not rate the stimulus with the mean value of male voice production (which was supposed to be the prototype of this category) as the best example of category /i/, but the stimuli with an  $F_2$  frequency value higher than the supposed prototype. Johnson (2000) suggested two alternatives to explain this trend, the so-called perceptual hyperspace effect. The first is based on Lindblom's (1990) proposal: The "hyperspace effect reflects listeners' production targets that are subject to undershoot in production" (Johnson 2000: 182), that is, the failure to achieve the intended goal of production. The second, in turn, considers this trend as evidence, albeit indirect, of the adaptive dispersion (Liljencrants & Lindblom 1972) hypothesis: The "hypothesis that the distinctive sounds of a language tend to be positioned in phonetic space so as to maximize perceptual contrast" (Johnson 2000: 181). Further studies (Prince & Smolensky 2004; P. Boersma 2015) are required to define the reason why our two subjects did choose (or not) the category /i/.<sup>18</sup>

---

<sup>18</sup> For this issue of prototype /i/ see also Boersma (2015) in which a model of "phonology and phonetics in parallel" is proposed based on the *Optimality Theory* (Prince & Smolensky, 2004).

## 7 Conclusions

The general goal of this work was to study the influence of L1 phonological relations on speech perception, using a methodology that encompasses the prototype theory standpoint. For this purpose, the four vowels /i, e, ε, a/ of EP were chosen as the object of this research, with the mid front vowels being considered target-vowels and the other two vowels as comparison.

It was concluded through an identification task that the participants used four distinct vowel categories, which are in accordance with the phonological theory and with what has been observed in speech production studies of this language. However, there were individual variations regarding the boundaries between categories. 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 (in terms of number of stimuli), while also varying individually.

The boundaries between /i/ and /e/, and /e/ and /ε/ were less steep than the boundary between /ε/ and /a/, the latter being non-neutralizable. The gradient of the /e/-/ε/ boundary could be due to the complex phonological relations that these sounds establish in the vowel system of EP. Regarding the comparison between the /i/-/e/ and /e/-/ε/ boundaries, results showed that the transitions between these vowel categories had more or less the same flat slope. The smaller gradient of the /i/-/e/ boundary in relation to that of the /ε/-/a/ boundary, may result from a specificity of EP (/ε/ is different from /a/ in terms of height and fronting, whereas /i/ and /e/ are only distinct in terms of height, as mentioned before), from particularities of the methodology used in this study, or these results constitute new evidence of a novel neutralizable opposition in EP, that of /i/-/e/. There is also a possibility of dialectal specificity (Vigário 2003; 2022; Segura 2013; Andrade 2020). According to the comments of the participants from Lisbon, and by inspecting the result of the northern listeners that participated in the pre-test, this result suggests an influence of the dialectal factor (Rodrigues & Martins 1999) on perception, which is extremely interesting. The question remains open for future work.

This study had its limitations, namely, the number of participants and the characteristics of the stimuli. Regarding the first limitation, the fact that we recruited a small group (N = 15) of participants may have influenced the results. Regarding the stimuli, we point out two crucial issues that may have not allowed us to delve into some phonological phenomena in a more rigorous manner. First, the lack of stimuli with more extreme acoustic values than /i/ (stimulus 01) and /a/ (stimulus 31) which would have covered the entire category of these two vowels. Broadening the scope of the category /i/ would make it possible to confirm the performance of the two subjects who barely identified the category /i/ with the stimuli used in our study. Second, the acoustic distance between the stimuli was also a factor that limited the application of the tests mentioned above. We opted, in our study, for the TANDEM-STRAIGHT speech morphing method to generate stimuli with the rich information of natural speech. Eleven

stimuli were created between two anchors of natural speech, but the absolute range of  $F_1$ ,  $F_2$  and  $F_3$  frequencies between stimuli was naturally different.

Considering the study carried out, we highlight some aspects worthy of future research. Thus, the same study can be extended to the BP variety and to other EP dialects (Segura 2013; Vigário 2022), which allows for comparisons to be made and to verify their specificities in perception, as well as providing evidence for the cognitive and/or phonological theoretical model. Should comparative studies be carried out between EP dialects, it would be equally pertinent to obtain, with the same methodology, the acoustic data not only of each dialect but capturing individual variability. This way, it would be possible to directly compare the relationship between the two domains: Speech perception and speech production.

Differences in the occurrence/ absence of phonological processes related to the realisation of vowels in unstressed position (e.g., complex nucleus reduction exceptions) may be relevant for the understanding of the relations between underlying categories and phonetic categories in EP. It would be interesting to carry out an experimental study with mid vowels that neutralise in the non-stressed context and to compare the gradient of the boundaries with that of other mid front/ central vowels.

The methodologies of this research, that is, the assessment of the boundary between categories with goodness rating scores, were, in a way, quite challenging, so we still need to find a better strategy to analyse the gradient of the boundaries.

---

## Appendix A

No. of stimuli	$f_o$ (Hz)	$F_1$ (Hz)	$F_2$ (Hz)	$F_3$ (Hz)	Duration (ms)
stimulus001	123	319	2245	2891	465
stimulus002	123	319	2192	2860	462
stimulus003	122	323	2167	2854	460
stimulus004	122	329	2134	2788	457
stimulus005	121	335	2074	2775	454
stimulus006	122	344	2049	2754	452
stimulus007	121	352	2030	2720	449
stimulus008	121	360	2001	2699	447
stimulus009	121	366	1972	2668	444
stimulus010	120	371	1936	2609	442
stimulus011	120	376	1918	2599	440
stimulus012	120	381	1906	2580	440
stimulus013	119	389	1899	2547	440
stimulus014	119	397	1895	2533	440
stimulus015	119	406	1893	2503	440
stimulus016	119	415	1891	2491	440
stimulus017	118	423	1885	2462	440
stimulus018	118	430	1872	2452	440
stimulus019	118	434	1867	2420	440
stimulus020	117	440	1864	2393	440
stimulus021	117	440	1860	2391	440
stimulus022	117	468	1842	2348	435
stimulus023	117	492	1771	2343	430
stimulus024	116	524	1720	2339	425
stimulus025	116	551	1648	2338	421
stimulus026	115	578	1586	2337	416
stimulus027	116	611	1496	2297	412
stimulus028	115	637	1468	2290	407
stimulus029	115	669	1373	2288	403
stimulus030	115	693	1315	2280	398
stimulus031	114	709	1261	2253	440

**Table A.1:** Values of  $f_o$ ,  $F_1$ ,  $F_2$  and  $F_3$  (Hz) and duration (ms) of the 31 synthesized stimuli, obtained with the default parameterization of Praat version 6.0.37 SoundEditor.



## Appendix B

Listener	Sex	Age	Place of birth	L1	Social status	Academic degree	L2/L3	Feedback
AB	M	22		EP	University Student	Undergraduate	EN(B1), JP(A2)	
VB	F	21	Braga	EP	University Student	Undergraduate	EN(A2), JP(A1)	
FB	M	20	Braga	EP	University Student	Undergraduate	EN(C1)	Confusion between the vowels [e] and [ɛ]
NB	F	24	Porto	EP	Researcher	Ph.D.	EN(C1), JP(A2)	
JB	M	24	Porto	EP	University Student	Master	EN(C1)	
SB	F	23	Braga	EP	University Student	Master	EN(C2)	

**Table B.1:** Biographical, linguistic and sociodemographic data, and comments regarding the participants in the pre-test.

## Appendix C

Listener	Sex	Age	Place of birth	L1	Social status	Academic degree	L2/L3	Feedback
CC	F	19	Lisbon	EP	University Student	Undergraduate		
AP	M	26	Lisbon	EP	Worker	Undergraduate (graduated in 2012)		[i] and [e] were difficult, especially when identifying [e] in <medo>
BM	F	20	Lisbon	EP	University Student	Undergraduate	EN(C2)	
MM	F	25	Lisbon	EP	Worker	Master	EN(B1), SP(B1)	[i] and [e] were difficult
MF	F	20	Lisbon	EP	University Student	Undergraduate	EN(A2)	[i] and [e] were difficult
ML	M	19	Lisbon	EP	University Student	Undergraduate	EN(B2)	
CP	F	34	Lisbon	EP	Researcher	Ph.D.	EN(C1)	
IM	F	20	Lisbon	EP	University Student	Undergraduate	EN(C2)	
MC	F	21	Lisbon	EP	University Student	Undergraduate	EN(C1)	

**Table C.1:** Biographical, linguistic and sociodemographic data, and comments regarding the participants in the main experiment.

Listener	Sex	Age	Place of birth	L1	Social status	Academic degree	L2/L3	Feedback
GS	M	28	Lisbon	EP	University Student	Undergraduate	EN(C2)	
FM	M	28	Lisbon	EP	Worker	Master	EN(C1)	
AC	M	27	Lisbon	EP	Researcher	Master	EN(C1), ES(B1)	
BR	F	22	Lisbon	EP	University Student	Undergraduate	EN(C2)	
CL	F	19	Lisbon	EP	University Student	Undergraduate	EN(B2), FR(A1)	[i] was difficult because it is usually not that long
LP	M	28	Lisbon	EP	Researcher	Master	EN(C2)	

**Table C.2:** Biographical, linguistic and sociodemographic data, and comments regarding the participants in the main experiment.

## Data Availability / Supplementary Files

All audio stimuli, raw listeners responses, .csv and .xlsx files used for data analysis and to generate the figures, Praat and R scripts associated with this submission are openly available from <https://doi.org/10.17605/OSF.IO/QGMPX>.

## Ethics and consent

Ethical permission (Parecer N° P523-10/2018, dated 21/11/2018) was obtained from an independent ethics committee (Comissão de Ética da Unidade Investigação em Ciências da Saúde – Enfermagem da Escola Superior de Enfermagem de Coimbra, Coimbra, Portugal), and informed consent was collected from all participants prior to data collection. The study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments.

## 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).

This work was developed as part of the M.Sc. in Linguistics at the University of Porto, Portugal: Megumi Im, *Perceção das Vogais Semifechadas e Semiabertas pelos Falantes Nativos do Português Europeu* [Speech Perception of Close-mid and Open-mid Vowels by Native Speakers of European Portuguese], 2019, Master in Linguistics, Faculdade de Letras, University of Porto, Porto, Portugal.

## Competing interests

The authors have no competing interests to declare.

---

## References

- Aaltonen, Olli & Eerola, Osmo & Hellström, Åke & Uusipaikka, Esa & Lang, A. Heikki. 1997. Perceptual magnet effect in the light of behavioral and psychophysiological data. *The Journal of the Acoustical Society of America* 101(2). 1090–1105. DOI: <https://doi.org/10.1121/1.418031>
- Amengual, Mark & Chamorro, Pilar. 2016. The Effects of Language Dominance in the Perception and Production of the Galician Mid Vowel Contrasts. *Phonetica* 72(4). 207–236. DOI: <https://doi.org/10.1159/000439406>
- Andrade, Andrade. 2020. Vocalismo. (E. B. P. Raposo & M. F. B. Nascimento & M. A. C. Mota & L. Segura & A. Mendes & G. Vicente & R. Veloso, Eds.), *Gramática Do Português (Vol. 3)*. Fundação Calouste Gulbenkian.
- Bisol, Leda. 2001. *Introdução a estudos de fonologia do português brasileiro*. Porto Alegre: Edipucsl.
- Bisol, Leda & Veloso, João. 2016. Phonological processes affecting vowels. In Marie Wetzels, Willem Leo & Costa, João & Menuzzi, Sergio (eds.), *The Handbook of Portuguese Linguistics*, 69–85. Hoboken: Wiley. DOI: <https://doi.org/10.1002/9781118791844.ch5>
- Boersma, Paul. 2015. Prototypicality judgments as inverted perception. In Fanselow, G. & Caroline, F. & Schlesewsky, M. & Vogel, R. (eds.), *Gradience in Grammar: Generative Perspectives*,. New York: Oxford University Press.
- Boersma, Paul. 2001. Praat, a system for doing phonetics by computer. *Glott International* 5(9/10). 341–345.
- Bonilha, G. F. G. 2004. Sobre a aquisição das vogais. In Lamprecht, Regina R. (ed.), *Aquisição fonológica do português: Perfil de desenvolvimento e subsídios para a terapia*, 61–71. Porto Alegre: Artmed.
- Boomershine, A. & Hall, K. C. & Hume, E. & Johnson, K. 2008. The impact of allophony versus contrast on speech perception. In Avery, P. & Drescher, E. & Rice, K. (eds.), *Phonological contrast: Perception and acquisition*, 146–172. New York: Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110208603.2.145>
- Câmara, J. M. 1970. *Estrutura da Língua Portuguesa*. Petrópolis: Editora Vozes.

- Carvalho, Joaquim Brandão. 2011. Contrastive hierarchies, privative features, and portuguese vowels. *Linguística: Revista de Estudos Linguísticos Da Universidade Do Porto* 6(1). 51–66.
- Correia, Susana & Butler, Joseph & Vigário, Marina & Frota, Sónia. 2015. A Stress “Deafness” Effect in European Portuguese. *Language and Speech* 58(1). 48–67. DOI: <https://doi.org/10.1177/0023830914565193>
- Dresher, B. E. 2003a. On the acquisition of phonological contrasts. In *Proceedings of GALA 2003*, 27–46.
- Dresher, B. E. 2003b. The Contrastive Hierarchy in Phonology. *Tronto Working in Linguistic* 20. 47–62.
- Eerola, Osmo & Savela, Janne & Laaksonen, Juha-Pertti & Aaltonen, Olli. 2012. The effect of duration on vowel categorization and perceptual prototypes in a quantity language. *Journal of Phonetics* 40(2). 315–328. DOI: <https://doi.org/10.1016/j.wocn.2011.12.003>
- Escudero, Paola & Boersma, Paul & Rauber, Andréia Schurt & Bion, Ricardo A. H. 2009. A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *The Journal of the Acoustical Society of America* 126(3). 1379–1393. DOI: <https://doi.org/10.1121/1.3180321>
- Evans, Bronwen G. & Iverson, Paul. 2004. Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *The Journal of the Acoustical Society of America* 115(1). 352–361. DOI: <https://doi.org/10.1121/1.1635413>
- Evans, Bronwen G. & Iverson, Paul. 2007. Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America* 121(6). 3814–3826. DOI: <https://doi.org/10.1121/1.2722209>
- Fikkert, Paula. 2005. From Phonetic Categories to Phonological Features Specification: Acquiring the European Portuguese Vowel System. *Lingue e Linguaggio* 263–280.
- Freitas, M. J. 2003. The Vowel [i] in the Acquisition of European Portuguese. In *Proceedings of GALA 2003*, 163–174.
- Frisch, Stefan A. 2018. Exemplar theories in phonology. In Hannahs, S. J. & Bosch, Anna R. K. (eds.), *The Routledge Handbook of Phonological Theory*, 553–568. Abingdon: Routledge. DOI: <https://doi.org/10.4324/9781315675428-20>
- Greene, Beth G. 1986. Perception of Synthetic Speech by Nonnative Speakers of English. *Proceedings of the Human Factors Society Annual Meeting* 30(13). 1340–1343. DOI: <https://doi.org/10.1177/154193128603001323>
- Grieser, DiAnne & Kuhl, Patricia K. 1989. Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology* 25(4). 577–588. DOI: <https://doi.org/10.1037//0012-1649.25.4.577>
- Harris, J. 1994. *English Sound Structure*. Cambridge: Blackwell.
- Harris, J. & Lindsey, G. 2000. Vowel patterns in mind and sound. In Roberts, Noel Burton & Carr, Philip & Docherty, Gerard (eds.), *Phonological knowledge: Conceptual and empirical issues*. Oxford University Press. DOI: <https://doi.org/10.1093/oso/9780198241270.003.0008>
- Hume, E. & Johnson, K. 2003. The impact of partial phonological contrast on speech perception. In *Proceedings of the 15th International Congress of Phonetic Sciences*, 2385–2388.

- Iverson, Paul & Kuhl, Patricia K. 2000. Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism? *Perception & Psychophysics* 62(4). 874–886. DOI: <https://doi.org/10.3758/BF03206929>
- Johnson, Keith. 2000. Adaptive Dispersion in Vowel Perception. *Phonetica* 57(2–4). 181–188. DOI: <https://doi.org/10.1159/000028471>
- Johnson, Keith & Flemming, Edward & Wright, Richard. 1993. The Hyperspace Effect: Phonetic Targets Are Hyperarticulated. *Language* 69(3). 505–528. DOI: <https://doi.org/10.2307/416697>
- Kawahara, Hideki & Morise, Masanori. 2011. Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework. *Sadhana* 36(5). 713–727. DOI: <https://doi.org/10.1007/s12046-011-0043-3>
- Kawahara, Hideki & Takahashi, Toru & Morise, Masanori & Banno, Hideki. 2009. Development of exploratory research tools based on TANDEM-STRAIGHT. In *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, 111–120.
- Klatt, Dennis H. & Klatt, L. C. 1990. Analysis, Synthesis, and Perception of Voice Quality Variations Among Female and Male Talkers. *Journal of the Acoustical Society of America* 87(2). 820–857. DOI: <https://doi.org/10.1121/1.398894>
- Kuhl, Patricia K. 1991. Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics* 50(2). 93–107. DOI: <https://doi.org/10.3758/BF03212211>
- Kuhl, Patricia K. & Williams, K. & Lacerda, F. & Stevens, K. & Lindblom, B. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255(5044). 606–608. DOI: <https://doi.org/10.1126/science.1736364>
- Ladd, D. 2006. “Distinctive phones” in surface representation. In Goldstein, L. & Whalen, D. & Best, C. (eds.), *Laboratory Phonology* 8, 3–26. New York: De Gruyter. DOI: <https://doi.org/10.1515/9783110197211.1.3>
- Lee, S. H. 2010. Contraste das vogais no PB e OT [Vowel contrast in Brazilian Portuguese and the Optimality Theory]. *Estudos Linguísticos* 39.
- Liljencrants, Johan & Lindblom, Bjorn. 1972. Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast. *Language* 48(4). 839–862. DOI: <https://doi.org/10.2307/411991>
- Lindblom, Bjorn. 1990. Explaining Phonetic Variation: A Sketch of the H & H Theory. In Hardcastle, William J & Marchal, Alain (eds.), *Speech Production and Speech Modelling*, 403–439. Dordrecht: Kluwer Academic. DOI: [https://doi.org/10.1007/978-94-009-2037-8\\_16](https://doi.org/10.1007/978-94-009-2037-8_16)
- Lively, S. E. & Pisoni, D. B. 1997. On prototypes and phonetic categories: a critical assessment of the perceptual magnet effect in speech perception. *Journal of Experimental Psychology. Human Perception and Performance* 23(6). 1665–79. DOI: <https://doi.org/10.1037//0096-1523.23.6.1665>
- Logan, John S. & Greene, Beth G. & Pisoni, David B. 1989. Segmental intelligibility of synthetic speech produced by rule. *The Journal of the Acoustical Society of America* 86(2). 566–581. DOI: <https://doi.org/10.1121/1.398236>

- Masapollo, Matthew & Polka, Linda & Molnar, Monika & Ménard, Lucie. 2017. Directional asymmetries reveal a universal bias in adult vowel perception. *The Journal of the Acoustical Society of America* 141(4). 2857–2869. DOI: <https://doi.org/10.1121/1.4981006>
- Mateus, Maria Helena Mira. 2003. Fonologia. In Mateus, Maria Helena Mira & Brito, Ana Maria & Duarte, Inês & Faria, Isabel Hub (eds.), *Gramática da Língua Portuguesa*, 987–1033. Lisboa: Caminho.
- Mateus, Maria Helena Mira & Andrade, Ernesto. 2000. *The Phonology of Portuguese*. Oxford: Oxford University Press.
- McMurray, B. 2022. The Myth of Categorical Perception. *The Journal of the Acoustical Society of America* 152(6). 3819–3842. DOI: <https://doi.org/10.1121/10.0016614>
- Miller, Joanne L. 1994. On the internal structure of phonetic categories: a progress report. *Cognition* 50(1–3). 271–285. DOI: [https://doi.org/10.1016/0010-0277\(94\)90031-0](https://doi.org/10.1016/0010-0277(94)90031-0)
- Nadeu, Marianna & Renwick, Margaret E. L. 2016. Variation in the lexical distribution and implementation of phonetically similar phonemes in Catalan. *Journal of Phonetics* 58. 22–47. DOI: <https://doi.org/10.1016/j.wocn.2016.05.003>
- Nevins, Andrew. 2012. Vowel lenition and fortition in Brazilian Portuguese. *Letras de Hoje* 47. 228–233.
- Newman, Rochelle S. 2003. Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *The Journal of the Acoustical Society of America* 113(5). 2850–2860. DOI: <https://doi.org/10.1121/1.1567280>
- Pisoni, David B. 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics* 13(2). 253–260. DOI: <https://doi.org/10.3758/BF03214136>
- Pisoni, David B. 1997. Perception of Synthetic Speech. In van Santen, J. P. H. & Olive, J. P. & Sproat, R. W. & Hirschberg, J. (eds.), *Progress in Speech Synthesis*. New York: Springer.
- Prince, Alan & Smolensky, Paul. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden: Blackwell. DOI: <https://doi.org/10.1002/9780470759400>
- Ralston, James V. & Pisoni, David B. & Lively, Scott E. & Greene, Beth G. & Mullennix, John W. 1991. Comprehension of Synthetic Speech Produced by Rule: Word Monitoring and Sentence-by-Sentence Listening Times. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 33(4). 471–491. DOI: <https://doi.org/10.1177/001872089103300408>
- RCoreTeam. 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Renwick, Margaret E. L. 2014. *The Phonetics and Phonology of Contrast*. Berlin: De Gruyter. DOI: <https://doi.org/10.1515/9783110362770>
- Renwick, Margaret E. L. & Ladd, D. Robert. 2016. Phonetic Distinctiveness vs. Lexical Contrastiveness in Non-Robust Phonemic Contrasts. *Laboratory Phonology* 7(1). 19. DOI: <https://doi.org/10.5334/labphon.17>

- Repp, B. H. & Liberman, A. M. 1987. Phonetic category boundaries are flexible. In Harnad, S. (ed.), *Categorical perception: The groundwork of cognition*, 89–112. New York: Cambridge University Press.
- Rodrigues, C. & Martins, F. 1999. Espaço acústico das vogais acentuadas de Braga. In *Actas do XV Encontro Nacional da Associação de Linguística*. vol. II. Faro: Universidade do Algarve. pp. 301–317. In *Actas do XV Encontro Nacional da Associação de Linguística*, 301–317.
- RStudioTeam. 2023. RStudio: Integrated Development for R. Posit Software, PBC, Boston, USA.
- Savela, Janne & Eerola, Osmo & Aaltonen, Olli. 2014. Weighted vowel prototypes in Finnish and German. *The Journal of the Acoustical Society of America* 135(3). 1530–1540. DOI: <https://doi.org/10.1121/1.4864305>
- Segura, Luisa. 2013. Variedades dialetais dos Português Europeu. In Raposo, Eduardo Buzaglo Paiva & Nascimento, Maria Fernanda Bacelar & Mota, Maria Antónia Coelho & Segura, Luisa & Mendes, Amália & Vicente, Graça & Veloso, Rita (eds.), *Gramática do Português (Vol. 1)*, 85–142. Lisboa: Fundação Calouste Gulbenkian.
- Shi, Lei & Griffiths, Thomas L. & Feldman, Naomi H. & Sanborn, Adam N. 2010. Exemplar models as a mechanism for performing Bayesian inference. *Psychonomic Bulletin & Review* 17(4). 443–464. DOI: <https://doi.org/10.3758/PBR.17.4.443>
- Silva, Daniel Marcio Rodrigues & Neves, Rui Rothe. 2009. Um estudo experimental sobre a percepção do contraste entre as vogais médias posteriores do português brasileiro. *DELTA* 25(2). 319–345. DOI: <https://doi.org/10.1590/S0102-44502009000200005>
- Silva, Daniel Marcio Rodrigues & Neves, Rui Rothe. 2016. Perception of height and categorization of Brazilian Portuguese front vowels. *DELTA* 32(2). 355–373. DOI: <https://doi.org/10.1590/0102-4450984064164376868>
- Stevenson, Sophia & Zamuner, Tania. 2017. Gradient phonological relationships: Evidence from vowels in French. *Glossa: A Journal of General Linguistics* 2(1). DOI: <https://doi.org/10.5334/gjgl.162>
- Tiegs, Jessica. 2023. *Processing Consequences of Marginal Contrastivity in Romance Phonology*. Tucson: University of Arizona, USA, <http://hdl.handle.net/10150/668106> dissertation.
- Trubetzkoy, N. S. 1939. *Grundzuge der Phonologie*. Prague: Travaux du Cercle Linguistique de Prague 7.
- Veloso, João. 2012. Vogais centrais do português europeu contemporâneo: uma proposta de análise à luz da fonologia dos elementos. *Letras de Hoje* 47(3). 234–243.
- Veloso, João. 2015. *Introdução à Fonologia: Nível fonético e nível fonológico*. Porto: University of Porto, Portugal.
- Veloso, João. 2016. O sistema vocálico e a redução e neutralização das vogais átonas em português europeu contemporâneo. In Martins, Ana Maria & Carrilho, Ernestina (eds.), *Manual de Linguística Portuguesa*, 636–662. Berlin: De Gruyter. DOI: <https://doi.org/10.1515/9783110368840-026>
- Vigário, Marina. 2003. *The Prosodic Word in European Portuguese*. Berlin: De Gruyter. DOI: <https://doi.org/10.1515/9783110900927>

Vigário, Marina. 2022. Portuguese. In Gabriel, Christoph & Gess, Randall & Meisenburg, Trudel (eds.), *Manual of Romance Phonetics and Phonology*, 839–881. Berlin: De Gruyter. DOI: <https://doi.org/10.1515/9783110550283-027>

Vihman, Marilyn & Croft, William. 2007. Phonological development: toward a “radical” templatic phonology. *Linguistics* 45(4). DOI: <https://doi.org/10.1515/LING.2007.021>

Wetzels, Willem Leo Marie. 1992. Mid vowel neutralization in Brazilian Portuguese. *Cadernos de Estudos Lingüísticos* 23. 19–55.

Wetzels, Willem Leo Marie. 2011. The representation of vowel height and vowel height neutralization in Brazilian Portuguese (Southern Dialects). In Goldsmith, J. A. & Hume, E. & Wetzels, W. L. M. (eds.), *Tones and Features: Phonetic and Phonological Perspectives*, 331–360. De Gruyter. DOI: <https://doi.org/10.1515/9783110246223.331>

Zhang, Jennifer & Graham, Lindsey & Barlaz, Marissa & Hualde, José Ignacio. 2022. Within-Speaker Perception and Production of Two Marginal Contrasts in Illinois English. *Frontiers in Communication* 7. DOI: <https://doi.org/10.3389/fcomm.2022.844862>

