

Article

Predicting Discrimination in L3 Portuguese by Hungarian Speakers: The Effect of Perceptual Overlap

Gabriela Tavares ^{1,*}, Andrea Deme ^{2,3,4} and Susana Correia ¹¹ Linguistics Research Centre, NOVA University, 1069-061 Lisbon, Portugal; susanacorreia@fcsh.unl.pt² Department of Applied Linguistics and Phonetics, ELTE Eötvös Loránd University, 1088 Budapest, Hungary; deme.andrea@btk.elte.hu³ MTA-HUN-REN NYTK Lendület “Momentum” Neurophonetics Research Group, 1068 Budapest, Hungary⁴ HUN-REN Hungarian Research Centre for Linguistics, 1088 Budapest, Hungary

* Correspondence: gabrielatavares@fcsh.unl.pt

Abstract: Perceptual overlap has been attested as significantly contributing to difficulties in L2 speech perception. The current study aims at investigating whether this effect is also observable in the context of L3, specifically in the perception of European Portuguese oral vowels by Hungarian listeners. We crossed the results of two experiments—a categorization task and a discrimination task—and found that perceptual overlap is also a significant factor in L3 perception. Furthermore, we compared different measures of perceptual overlap as predictors for discrimination abilities of L3 vowel contrasts. Namely, we compared perceptual overlap scores calculated on group means and scores based on individual results. None of the measures was conclusively more reliable than another in predicting differences in discrimination difficulties. However, accuracy in perception of EP contrasts or vowels absent from the Hungarian vocalic system was significantly lower than for the other vowels, suggesting that *non-nativeness* can cause difficulties in L3 perception. Additionally, participants who also reported knowledge of German performed more accurately in discrimination of contrasts that included the vowel [e], a vowel absent from their L1 but present in the German vocalic system, indicating a positive effect of knowledge of languages previously acquired on L3 perception.

Keywords: L3 Portuguese; speech perception; vowel perception; perceptual overlap; European Portuguese; Hungarian



Citation: Tavares, Gabriela, Andrea Deme, and Susana Correia. 2024.

Predicting Discrimination in L3 Portuguese by Hungarian Speakers: The Effect of Perceptual Overlap.

Languages 9: 352. <https://doi.org/10.3390/languages9110352>

Academic Editors: Romana Kopečková and Anna Balas

Received: 7 August 2024

Revised: 26 October 2024

Accepted: 30 October 2024

Published: 20 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. L3 Phonology

Similarly to research in Second Language Acquisition (SLA), studies on Third Language Acquisition (TLA) were initially focused on morphosyntactic and lexical topics, with investigation on phonology emerging only later. Currently, there is already a considerable body of research in L3 phonology. However, due to the complexity of the issue—L3 learners form a very heterogeneous group, with knowledge of L2 and/or L3 being difficult to assess—existing studies do not provide us yet with a robust framework. Research in L3 phonology presents the challenge of investigating diverse factors, such as L3 experience, L2 status, or typological proximity, in numerous possibilities of language combinations (Wang and Nance 2023). It is not surprising, then, to observe that there are still important issues left to investigate. For example, studies involving at least one non-Indo-European language (as L1, L2, or L3) are rarer than studies where all the observed languages belong to the Indo-European family (Wang and Nance 2023). Furthermore, the number of studies focusing on perception are still considerably fewer than studies observing production (Stan 2024; Wang and Nance 2023). Finally, when it comes to experiments focusing on segmental features, it is clear that consonants have been more extensively investigated than vowels (Stan 2024).

With respect to L3 research focusing on speech perception, and, more specifically, on experiments observing progressive transfer (i.e., from the L1 or L2 to the L3), most studies point to a combined or hybrid L1/L2 transfer to the L3 (Liu et al. 2019; Liu and Lin 2021; Onishi 2016; Stan 2024; Wrembel et al. 2019). More importantly, some authors mention a general cognitive advantage in L3 phonological acquisition, given that learners have already acquired an additional language and thus make use of linguistic awareness and learning strategies (Cabrelli 2012; Onishi 2016). For example, Onishi (2016) found a correlation between perception skills in L2 and L3; that is, the better participants performed in the L2, the better they were at discriminating L3 contrasts. Although this correlation was not found for all tested contrasts, it was observed in contrasts absent from both the L2 and the L1. The author concluded that “a positive L2 influence comes not only from specific phonological contrasts acquired via the L2, but also from general exposure to non-native speech sounds” (p. 471).

1.2. Non-Native Vowel Categorization and Perceptual Overlap

In the context of L2 phonological acquisition, a number of models have been proposed to explain how non-native speakers perceive and learn the sound structure of a new language: the Speech Learning Model (Flege 1995; Flege and Bohn 2021), the Perceptual Assimilation Model (Best 1995; Best and Tyler 2007), the Second Language Linguistic Perception Model (Escudero and Boersma 2004; Escudero 2005), and the Universal Perceptual Model (Georgiou 2021). An extensive body of research has been conducted to test these models for consonants (Guion et al. 2000; Best et al. 2001; Harnsberger 2001; Aoyama et al. 2004; Chen and Fon 2007; Ingvalson et al. 2011; Bohn and Best 2012), as well as vowels (Flege 2003; Trofimovich et al. 2003; Escudero and Boersma 2004; Larson-Hall 2004; Iverson and Evans 2009; Strange et al. 2009; Elvin et al. 2014, 2021; Tyler et al. 2014; Faris et al. 2016, 2018; Georgiou and Dimitriou 2023; Georgiou et al. 2024). Although the mentioned models present differing points, they all sustain that acoustic and/or articulatory similarities and dissimilarities between the L1 and the L2 phonological systems influence L2 speech perception and acquisition; that is, the extent to which speakers experience difficulties in perceiving and learning certain non-native sounds depends on the degree of (dis)similarity of those sounds with native phonetic categories. One of the key issues being discussed in the field of crosslinguistic speech perception concerns how perceptual overlap between L1 and L2 categories accounts for difficulties in L2 speech perception. Perceptual overlap refers to situations where two or more L2 sounds are perceived as a single L1 category. For example, the Brazilian Portuguese vowel /ɛ/ is absent in the Iberian Spanish inventory, and Spanish speakers identify this vowel as /e/. However, the Brazilian Portuguese /e/ is also identified as /e/ by Spanish listeners (Elvin et al. 2021). Consequently, Spanish listeners do not perceive the Brazilian Portuguese /ɛ/-/e/ contrast (e.g., *este* [ˈeʃtɨ] 'this' vs. *este* [ˈeʃtʃi] 'East'). If two (or more) L2 sounds are perceived as only one L1 category (and there are no other competing categories involved), overlap is total. However, this scenario is rarely observed, and in most cases, we find partial overlap. For instance, in the above-mentioned study, Elvin et al. (2021) calculated a 45% chance of perceptual overlap when the Spanish listeners are presented with the Brazilian Portuguese /ɛ/-/e/ contrast; that is, in 45% of the cases, Spanish listeners will perceive both the Portuguese /ɛ/ and /e/ as /e/. To determine the amount of overlap between L2 and L1 categories, and, based on this, make predictions for discrimination, categorization tasks¹ are commonly used (Flege and Mackay 2004; Tyler et al. 2014; Faris et al. 2018; Elvin et al. 2021).

Research on perceptual overlap has conclusively demonstrated its effect on non-native speech discrimination and its reliability as a predictor for discrimination difficulties. For instance, in a study with Italian native speakers presented with English vowels, performance in contrast discrimination was worse with high values of perceptual overlap and, conversely, contrasts with low perceptual overlap were more accurately discriminated by the participants (Flege and Mackay 2004). Similar results were found in a perceptual study with vowel contrasts from three L2s (French, Norwegian, and Thai) presented to American

English speakers (Tyler et al. 2014). The results showed that participants discriminated vowel contrasts with no overlap or with a very low amount of perceptual overlap significantly better than contrasts with a high amount of perceptual overlap. The relationship between perceptual overlap and L2 vowel discrimination was also observed in Faris et al. (2018), who tested perception of Danish monophthongs and diphthongs by Australian English speakers. In this study, results revealed that discrimination was very good or excellent in trials with no overlapping contrasts, and poor to moderate in contrasts with partial overlap. Consequently, the authors concluded that perceived phonological overlap is a reliable predictor for discrimination difficulties.

Considering the importance of the amount of perceptual overlap in discrimination, how it should be quantified is also frequently discussed. Flege and Mackay (2004) and Faris et al. (2018) assessed perceptual overlap based on data collected in categorization tasks. In Flege and Mackay’s work, the authors calculated overlap scores of a given contrast, such as the English /i/-/ɪ/, by summing the lowest percentages found in the identification of the L2 sounds into the same native vowel category. In this calculation, they included all identification scores, even those below chance level (the classification overlap method). For example, in their study, the Italian listeners identified the English /i/ as /i/ (in 87% of the trials) or as /e/ (in 8% of the trials), while the English /ɪ/ was identified as /i/ (in 65% of the trials) or as /e/ (in 35% of the trials). Accordingly, the overlap score for the /i/-/ɪ/ contrast was (8% + 65% =) 73% (Figure 1)².

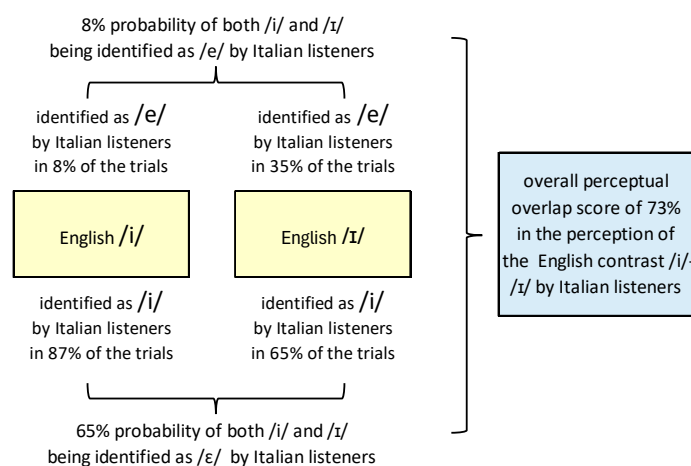


Figure 1. Example for the calculation of perceptual overlap scores, according to the classification overlap method, retrieved from Flege and Mackay (2004).

Faris et al. (2018) proposed instead the phonological overlap method. In their study, the authors determined the amount of overlap similarly to Flege and Mackay (2004), but they excluded identification percentages below chance level, since these “constitute anything more systematic than random responding” (p. 4). For example, in line with the phonological overlap method, the overlap score for the English /i/-/ɪ/ contrast shown in Figure 1 above would be only of 65%. Since /i/ was identified as /e/ below chance level (9%), the /i/-/ɪ/ contrast perceived as /e/ would not be pertinent. Consequently, only the perception of /i/-/ɪ/ as /i/ would be considered, thus the overlap score of 65%. The methodological difference between the classification and the phonological overlap methods has consequences in the prediction of discrimination difficulties. According to the first method, the English /i/-/ɪ/ should present a high level of difficulty in discrimination, since the overlap score is 73%. However, in the study of Flege and Mackay (2004), the Italian participants achieved an accuracy of 0.8 (A-prime score)—a very positive score—in an oddity discrimination task. Consequently, the phonological overlap method, with a lower perceptual overlap value, seems more in line with the accuracy in discrimination in the mentioned study.

In turn, [Elvin et al. \(2021\)](#) compared the effectiveness of two perceptual overlap parameters as predictors for L2 discrimination: perceptual overlap scores obtained from a categorization task (the perceptual overlap method) and perceptual overlap scores calculated based on L1–L2 acoustic comparisons of production data collected for each participant (the acoustic overlap method). The authors recruited Australian English and European Spanish native speakers and presented them with Brazilian Portuguese oral vowels as L2 stimuli. Comparisons of the results revealed that, although both methods effectively predicted discrimination difficulties, perceptual overlap scores were a more accurate predictor in the case of the Spanish listeners, whereas the results from the Australian participants were more in line with the acoustic overlap values. The authors suggest that these differences may be due to the size of L2 vowel inventories, which is considerably smaller in Spanish than in Australian English.

In addition to investigating perceptual overlap, [Tyler et al. \(2014\)](#) and [Elvin et al. \(2021\)](#) accounted for individual variability in their studies. In the former, individual variability was analyzed by determining each participant's assimilation pattern for each contrast in the categorization task. Discrimination of the non-native vowels was investigated by grouping participants' scores according to assimilation patterns. In the latter, predictions for categorization and discrimination were tested based on individual L1–L2 acoustic comparisons from production and perception data collected for each participant. Surprisingly, based on the results, [Elvin et al. \(2021\)](#) concluded that the first approach proved more efficient than the second in predicting discrimination performances; that is, group means were more reliable predictors of discrimination abilities than individual data. The authors purported that this is a by-product of the method of the analysis, since outliers have a weaker effect on the results in the case of group means than in the individual-level analysis, and they pointed out the need for further research.

The present study aims to investigate whether assumptions on L1–L2 speech perception, namely, the predictive role of perceptual overlap on discrimination abilities, hold up in an L3 speech perception context. In the experiments presented below, we test Hungarian native speakers on European Portuguese vowel perception. These experiments are part of a research project that observes the acquisition of European Portuguese phonology by Hungarian learners at the onset of acquisition. In the moment these learners start to acquire Portuguese, they already have experience with other languages, namely, English, German, and Spanish. Consequently, to them, European Portuguese is an L3, where L3 refers to a non-native language in the process of acquisition, considering that the speaker has previously acquired other L2(s) in addition to the L1(s) ([Hammarberg 2010](#)).

1.3. The European Portuguese and Hungarian Vowel Systems

European Portuguese (EP) is a national variety of Portuguese, an Indo-European, Romance, and Iberian language ([Ilari 2013](#)). Its standard dialect comprises a sizeable vowel inventory of oral and nasal vowels and diphthongs ([Andrade 2020](#)). In respect to oral vowels—/a/, /ɛ/, /e/, /i/, /ɔ/, /o/, and /u/—the quality of these phonemes' realization may depend on word stress, which in EP is a lexical feature that can fall on the last, penultimate, or antepenultimate syllable of the prosodic word ([Pereira 2020](#)). With the exception of a few phonological contexts or lexically marked cases, in an unstressed position /a/ undergoes reduction and surfaces as [ɐ], /ɛ/ and /e/ surface as [i], and /ɔ/ and /o/ as [u] ([Andrade 2020](#)). Previous studies have attested to the importance of vowel reduction in EP speech perception by native speakers ([Delgado-Martins 1986](#); [Correia et al. 2015](#); [Castelo and Santos 2017](#)), as well as difficulties in the acquisition of EP vowel reduction and reduced vowels in L3 Portuguese ([Castelo and Santos 2017](#); [Oliveira 2006](#)).

Figure 2 displays the EP oral vowel space.

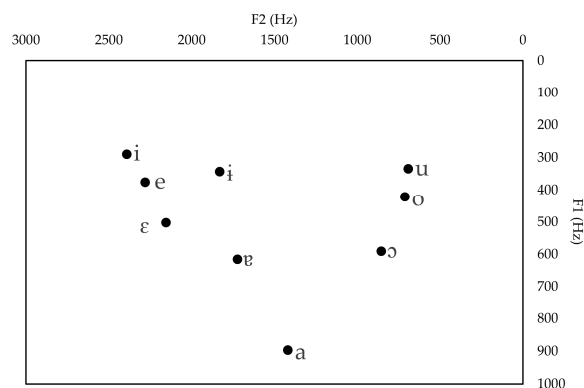


Figure 2. EP oral vowel space. Values retrieved from Andrade (2020, p. 3251).

Hungarian (“magyar”) is a non-Indo-European language from the Finno-Ugric language family (Siptár and Törkenczy 2000). The Hungarian vowel system includes fourteen phones: [ɒ], [a:], [ɛ], [e:], [i], [i:], [o], [o:], [ø], [ø:], [u], [u:], [y], and [y:] (Figure 3). These phones correspond to fourteen phonemes: /ɒ/, /a:/, /ɛ/, /e:/, /i/, /i:/, /o/, /o:/, /ø/, /ø:/, /u/, /u:/, /y/, and /y:/, respectively. The 14 vowels include only 9 different vocalic qualities, /ɒ/, /a:/, /ɛ/, /e:/, /i/, /i:/, /o/, /ø/, /u/, and /y/, and they are organized into seven short–long pairs, five contrasting only in length ([i]-[i:], [o]-[o:], [ø][ø:], [u]-[u:], and [y]-[y:]), and two contrasting in both length and quality ([ɒ]-[a:] and [ɛ]-[e:]) (Siptár and Törkenczy 2000; Markó 2017). In Hungarian, length is distinctive (e.g., *örül* [ˈøryl] ‘rejoiced’ vs. *örül* [ˈø:ryl] ‘getting crazy’) (Siptár and Törkenczy 2000; Markó 2017), which means that Hungarian listeners are sensitive to duration at the segmental level. Contrary to EP, word stress is not contrastive in Hungarian; consequently, speakers of this language are insensitive to word stress contrasts (Honbolygó et al. 2017; Peperkamp and Dupoux 2002; Peperkamp et al. 2010). While vowel reduction is traditionally not accounted for in Hungarian, reliable research on this topic is in fact absent from Hungarian linguistic studies (Markó et al. 2019).

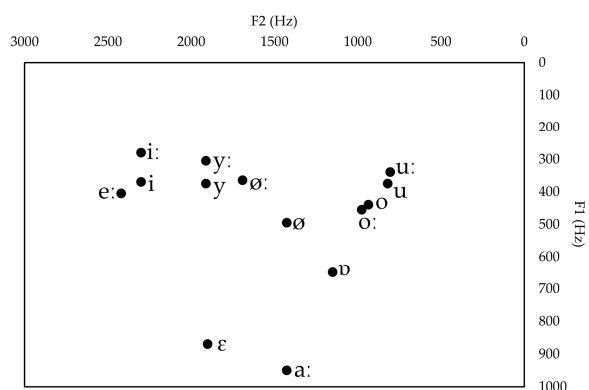


Figure 3. Hungarian vowel space. Values retrieved from Markó (2017, pp. 153–55).

In sum, EP and Hungarian vowel inventories crucially differ in two aspects: (i) The Hungarian vowel system obeys short–long phonological contrasts, whereas EP vowels are ruled by a stressed–unstressed pattern, and (ii) when comparing the two vowel inventories, we observe the absence of [ɐ] and [i] from the Hungarian vowel inventory, which are present in EP, and the absence of [y] and [ø] from the EP oral vowel inventory, which are present in Hungarian. It must also be pointed out that EP exhibits contrastive stress, and vowel quality is a relevant segmental cue to stress perception and lexical access (e.g., *ejército* [iˈzɛɾsɪtu] ‘army’ vs. *ejército* [izɪɾˈsɪtu] ‘I exercise’). In Hungarian, word stress is fixed on the first syllable, and native speakers seem to exhibit a robust stress “deafness”.

Our study aims at investigating how Hungarian learners of EP at the onset of learning perceive the unstressed EP vowels [ɐ] and [i] by observing how these speakers map the two target EP vowels into their vowel system (Experiment 1), and the discrimination difficulties they face connected to these vowels (Experiment 2). Additionally, we aim to observe whether knowledge of languages other than the L1 influences the perception of the two EP vowels.

2. Experiment 1

The first experiment consisted of a categorization task with goodness-of-fit rating. This experiment was designed to provide information on how Hungarian listeners map the Portuguese [ɐ] and [i] into their L1 system. To obtain a complete picture and identify possible cases of perceptual overlap, we included the nine EP oral vowels: [a], [ɐ], [ɛ], [e], [i], [i], [ɔ], [o], and [u]. Hungarian participants were presented with auditory tokens of these vowels and asked to identify each one of them as instances of Hungarian vowel categories, represented orthographically. As described before, Hungarian contains seven short vowels and seven long vowels, which correspond to only nine different vocalic qualities: [ɒ], [a:], [ɛ], [e:], [i], [i:], [o], [ø], [u], and [y]. Considering that our goal was to observe perception based on vowel quality, and not on quantity, we only included the nine Hungarian vowels that differ in this respect.

2.1. Method

2.1.1. Participants

We recruited Hungarian native speakers with no previous or present contact with EP as listeners in our experiment through a link shared in Hungarian universities. To select participants, we collected sociolinguistic data by means of a questionnaire. We included 72 Hungarian speakers in the data analysis. These participants were aged between 18 and 45, and 47 were female speakers. All participants reported intermediate or advanced proficiency in English. German was the second most frequently reported additional language, also at an intermediate or advanced level ($n = 22$), followed by French ($n = 10$), Spanish ($n = 9$), Italian ($n = 6$), and Russian ($n = 3$).

To control for stimuli nativeness, we ran the tests with 30 EP native speakers (21 female) as well. They were all from the standard dialectal area³ and had no previous contact with Hungarian. Participants in this group were also aged between 18 and 45.

Neither the Hungarian nor the Portuguese speakers reported hearing problems.

2.1.2. Stimuli

Auditory stimuli comprised the target EP oral vowels inserted in a CV structure. We opted to use vowels in context based on the fact that vowels are better perceived in consonantal context than in isolation (Deme 2014; Rakerd 1984) and, crosslinguistically, the CV structure is the most frequent unmarked syllable structure (Rice 2007). To select the consonant, sequential searches were run in the Hungarian corpus “Szószablya” (Halácsy et al. 2003) and the Portuguese corpus “FrePoP” (Frota et al. 2010) in order to find a CV syllable that complies with both the L1 (Hungarian) and the target L3 (EP) phonotactics and has no meaning in either language. This way, we could prevent the influence of word meaning affecting our results. Using this method, we selected the structure [gV]. We also recorded stimuli for the familiarization tasks, produced in the L1 of the participants (Hungarian and EP). For these stimuli, we selected a highly frequent consonant in both languages, [t], and created [tV]-shaped items.

We recruited three female EP native speakers (aged 42, 42, and 45), also from the standard dialectal area of Lisbon, to record the EP stimuli. The target [gV] and [tV] syllables were recorded in carrier sentences to control for vowel quality and intonation. The speakers were also provided examples of real words containing the target syllables, as exemplified in (1). The tokens selected for the experiment were the last in each phrase so that they had similar intonation patterns (a falling pitch contour) and speech rate.

- (1) a. GUERRA [gɛ]. Diga [gɛ], por favor: [gɛ].
"WAR [gɛ]. Say [gɛ], please: [gɛ]".
- b. TELA [tɛ]. Diga [tɛ], por favor: [tɛ].
"CANVAS [tɛ]. Say [tɛ], please: [tɛ]".

Recording sessions took place individually and at each speaker’s home, due to the lockdown restrictions imposed at the time of the recordings. For the purposes of the recordings, a TASCAM DR-05 V2 digital recorder and a Beyerdynamic MCE 85 BA condenser microphone were used. The file format was set to .wav, with the sampling frequency set to 44,100 Hz, mono, and 32-bit depth. Background noise was eliminated individually for each recording with Audacity (Audacity Team 2020). Special attention was given to normalization of vowel duration, since Hungarian listeners display sensitivity to vowel length. Vowel length was manipulated using Praat (Boersma and Weenink 2020). Mean intensity averaged over the total course of each syllable was equalized across tokens.

Other than the EP stimuli, we also recorded [tV] syllables in Hungarian for the familiarization task presented to the Hungarian participants. These tokens were recorded by three female Hungarian native speakers (aged 27, 35, and 47), and similar preparation steps were taken as for the EP stimuli.

The final set of stimuli consisted of 81 tokens: 54 produced by female Portuguese speakers ((9 [gV] for the main trials + 9 [tV] for the familiarization trials with the Portuguese participants) × 3 speakers) and 27 produced by female Hungarian speakers (9 [tV] for the familiarization trials with the Hungarian participants × 3 speakers).

Additionally, we looked into the EP tokens’ acoustic parameters to investigate the relationship between the Hungarian and the EP acoustic vowel space. Vowel onset was identified at the first voicing bar of a complete formant structure. Intensity, fundamental frequency (f0), and the first two formant frequencies (F1 and F2) of each vowel were calculated from the median of the values measured in a 10% window in the midsection of the vowel. For f0 measures, the pitch floor was set to 100 Hz and the ceiling to 500 Hz, as recommended for female voices. To estimate formant frequencies, we used the Burg algorithm and set the maximum number of formants to 5 below 5500 Hz, while we used the default settings in every other parameter (Boersma and Weenink 2020). A summary of the acoustic data for the stimuli included in the main trials ([gV]) is presented in Table 1.

Table 1. Mean values of acoustic measures for the [gV] tokens’ vocalic segment.

Stimuli (syllable)	Vowel				
	Duration (ms)	Intensity (db)	f0 (Hz)	F1 (Hz)	F2 (Hz)
[ga]	261	85	174	850	1543
[gɛ]	260	88	180	636	1967
[gɛ]	265	88	178	528	2450
[ge]	285	89	181	427	2749
[gi]	244	89	197	381	2236
[gi]	240	91	206	384	2841
[gɔ]	275	89	186	590	1024
[go]	247	90	195	391	760
[gu]	261	90	203	350	718

The orthographic stimuli consisted of sets of real words written in the L1 of the participants. To select the Hungarian words, we carried out a sequential search conducted in the Hungarian corpus “Szószablya” (Halácsy et al. 2003). We opted to use words that had [gV.CVC] structure and were similar in token frequency. To ensure comparable familiarity across words presented to Hungarian listeners, a questionnaire was conducted with 25 Hungarian native speakers (aged 18–45), who rated how familiar they were with different words with a [gV.CVC] structure on a scale from 1 (not familiar at all) to 4 (very familiar). All words selected for the main trials presented a median rating of 4 (very familiar). In the familiarization trials, we used real words with a [tV.CVC] structure to

comply with the words for the main trials. The set of words we presented to the Portuguese participants to test stimuli nativeness included real words with [gV] or [tV] as a first syllable and that were believed to be easily identifiable by Portuguese listeners. Appendix A displays the words presented to both groups of participants.

We placed each set of words in 3×3 grids, with the first (target) syllables highlighted in yellow, as exemplified in Figure 4.



Figure 4. Computer screen of the trials presented to the Hungarian participants in Experiment 1.

2.1.3. Procedure

The experiment was built in PsychoPy, version 2020.2.3 (Peirce et al. 2019), and hosted online by pavlovia.org. Two versions were created, one for the Hungarian listeners and one for the Portuguese listeners. The two versions had similar structure and design, differing only in the language of the instructions and auditory stimuli for the familiarization trials, which were presented in the respective L1.

We asked the participants to match the EP auditory stimuli to L1 real words, displayed in the grids. Participants were presented with each token twice. After the first audition, listeners were asked to choose a word from the grid. Following this, a second audition of the same token followed, and participants had to rate the goodness-of-fit of the auditory stimulus on a four-point scale (1 = very bad example, 2 = bad example, 3 = good example, 4 = very good example). By using a four-point Likert scale, our goal was to force participants to express a clearly positive (3 and 4) or negative (1 and 2) judgment and to avoid the “face-saving don’t know” effect (Sturgis et al. 2014). Before the main task, participants had to complete two familiarization tasks with nine trials each, with one trial for each [tV] stimulus. The first consisted of identification with feedback (“Correct!” or “Incorrect. . . try again”). In the case of an incorrect response, participants repeated the same trial immediately after. No limit for attempts was set. The second familiarization task was similar, with [tV] stimuli and nine trials, but it included a goodness-of-fit rating, and no feedback was given to the participants. The main task consisted of 27 trials with the [gV] stimuli (9 EP oral vowels \times 3 speakers), randomized across participants. No feedback was provided in these trials.

2.1.4. Analyses

Raw data were processed and further analyzed in R (R Core Team 2021). We collected participants’ answers (number of times a real word was chosen for each EP vowel and respective goodness-of-fit rating).

For the confusion matrixes, we calculated mean percentages of identification and medians of rating for each EP vowel per L1 group. We used one-sample Wilcoxon signed-rank tests to determine when these percentages were significantly above chance level (i.e., 11.1%, as calculated by dividing one EP vowel by nine possible Hungarian categories).

We also looked into possible speaker effects in the identification of the EP vowels using chi-square tests.

A further aim of our study was to observe whether knowledge of additional L2s has an effect on how Hungarians perceive the EP [e] and [i] vowels. Considering that all participants had experience with English, it was not possible to assess whether knowledge of this language influenced the results. Consequently, we only looked into the possible effect of knowledge of German, since 31% of the participants reported speaking this language. To this purpose, we also conducted chi-square tests.

2.2. Results

Results collected for the Portuguese listeners (Table 2) revealed that the EP vowels were identified as the expected EP category, except for [o], which was identified as /o/ (50.0%) or /u/ (46.7%). Furthermore, all tokens were perceived by Portuguese participants as being a good or very good example of the given category (median = 3 or 4).

Results for the Hungarian listeners, as we can see in Table 3, were more dispersed. Four EP vowels—[a], [e], [i], and [u]—were systematically identified as four Hungarian categories, /a:/, /e:/, /i/, and /u/, respectively. Similar to what was observed with the Portuguese listeners, [o] was identified as /o/ (57.4%) or /u/ (42.1%). Chi-square tests revealed that the perception of this EP vowel was affected by the identity of the speaker who produced the tokens, in the case of both Portuguese listeners ($\chi^2 = 45.067, p < 0.001$) and Hungarian participants ($\chi^2 = 113.87, p < 0.001$). Regarding the critical target vowels, [e] and [i], the first was identified as /ε/ (68.5%) or /ø/ (30.1%), while the identification of the second was split between /y/ (73.1%) and /ø/ (25.0%). In what concerns the perception of the vowel [ε], Hungarian listeners identified this EP vowel as /ε/ (58.8%) or /e:/ (39.8%). In this case, the goodness-of-fit rating values (median = 2) suggest that the EP [ε] was not considered a good example of either the Hungarian /ε/ or the /e:/. In the remaining EP vowels, ratings for the more systematic choices were positive (median = 3), or even very positive (median = 4), as in the case of [a].

Table 2. Mean percentage identifications and median of goodness-of-fit ratings for the Portuguese listeners (n = 30). Portuguese words are presented in capital letters with the correspondent phonemic notation of the first syllable above. Numbers in boldface represent the stimuli identified above chance level (11.1%), according to Wilcoxon signed-rank tests. The rectangles identify the expected answers.

		Orthographic Stimuli								
		/ga/ GATO	/ge/ GAVETA	/ge/ GUERRA	/ge/ GUÊ	/gi/ GUERREAR	/gi/ GUITO	/go/ GOLA	/go/ GOTA	/gu/ GULA
Auditory stimuli	[ga]	98.9% (4)	1.1% (2)							
	[ge]	2.2% (3)	95.6% (4)			2.2% (2.5)				
	[ge]			90.0% (3)	8.9% (3)	1.1% (2)				
	[ge]			4.4% (3)	87.8% (3)	5.6% (2)	2.2% (2)			
	[gi]			2.2% (3.5)	5.6% (2)	92.2% (3)				
	[gi]				8.9% (2)	1.1% (2)	90.0% (3)			
	[go]							94.4% (4)	5.6% (3)	
	[go]					1.1% (4)		2.2% (3)	50.0% (4)	46.7% (3)
	[gu]	1.1% (1)			1.1% (3)				4.4% (3)	93.3% (3)

Table 3. Mean percentage identifications and median of goodness-of-fit ratings for the Hungarian listeners (n = 72). Hungarian words are presented in capital letters with the correspondent phonemic notation of the first syllable above. Numbers in boldface represent the stimuli identified above chance level (11.1%), according to Wilcoxon signed-rank tests.

		Orthographic Stimuli							
		<i>/ga/</i> GÁBOR	<i>/gɛ/</i> GARÁZS	<i>/gɛ/</i> GERELY	<i>/gɛ/</i> GÉPÉSZ	<i>/gi/</i> GITÁR	<i>/gi/</i> GONOSZ	<i>/gɔ/</i> GÖRÉNY	<i>/gol/</i> GULYÁS
Auditory stimuli	[ga]	97.7% (4)	2.3% (3)						
	[gɛ]	0.5% (2)		68.5% (3)			30.1% (3)	0.5% (2)	0.5% (4)
	[gɛ]			58.8% (2)	39.8% (2)		1.4% (1)		
	[ge]			1.4% (3)	81.9% (3)	16.2% (2)	0.5% (2)		
	[gi]				1.9% (1.5)		25.0% (2)		73.1% (3)
	[gi]				1.9% (1)	98.1% (3)			
	[gɔ]		78.7% (3)				21.3% (3)		
	[go]						57.4% (3)	0.5% (3)	42.1% (3)
	[gu]						10.6% (3)		89.4% (3)

To investigate the relationship between the results for the Hungarian listeners and the acoustic distance between L1 and EP vowels, we plotted the F1 × F2 vowel space for Hungarian vowels in perception; EP [ɛ], [ɛ], [e], and [i] tokens produced for the experiment; and EP reference values obtained from a standard variety of EP for production (Figure 5).

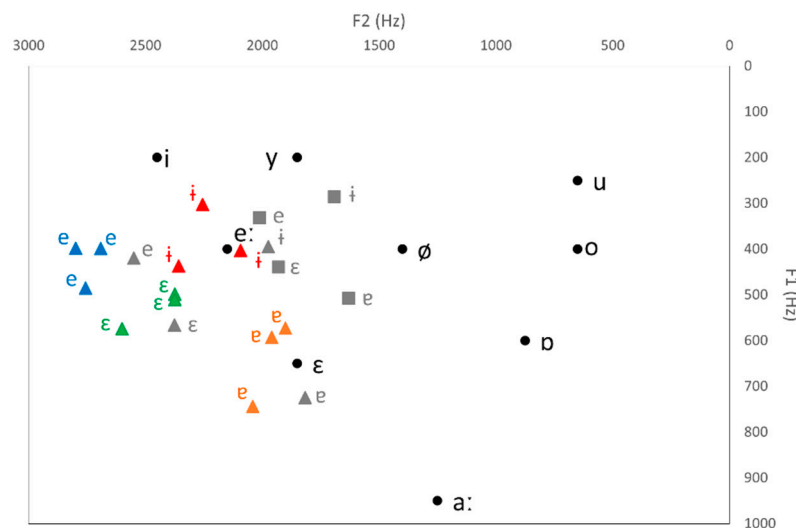


Figure 5. F1 × F2 vowel space for Hungarian vowels in perception (in black); EP [ɛ], [ɛ], [e], and [i] tokens used in the experiment (colored); and EP reference values for production (grey). Hungarian data were retrieved from Kiss (1985), with phonetic notations adapted from Markó (2017); EP reference values data were retrieved from Andrade (2020).

As we can see from Figure 5, [ɛ] tokens used in the experiment (colored triangles) are closer to the Hungarian /ɛ/, which explains this L1 category as first choice (68.5%) for the Hungarian listeners when perceiving the EP [ɛ]. However, the identification of [ɛ] as /ø/ in 30.1% of cases cannot be explained by the distance between the two vowels. Furthermore, even though the EP [i] seems to be located near the Hungarian /e:/ in the acoustic vowel space, it was identified as /y/ in 73.1% of cases and as /ø/ in 25% of cases.

A possible explanation may be that tokens recorded for the experiment (colored triangles) were deviant from the EP vowels we expect in female speakers (grey triangles). However, as shown in Figure 5, this explanation does not hold up, as the vowels in our tokens align well with the reference vowels in the acoustic vowel space. An alternative explanation may be related to the Hungarian perceptual map. The F1 and F2 frequency

values for the tokens used in the experiment (produced by female speakers) seem to be shifted towards higher F1 and F2 frequency values compared to values for the Hungarian vowels in the reference map we presented in Figure 5. If the values for the Hungarian vowels in perception were estimated based on measures taken from one (or several) male Hungarian speaker(s), this would mean that due to the larger vocal tract males generally have, we would find lower formant frequencies for the Hungarian vowels than for the Portuguese tokens, which might provide an explanation for the results. To test this, we calculated formant dispersion (FD) values (calculated for individual speakers and averaged in the case of the 3 Portuguese speakers), as these are indicators of body size (Fitch 1997).⁴ The results (see Data Availability Statement) support our assumption: The FD value calculated for the Hungarian vowels in perception (1017 Hz) was lower than for the Portuguese tokens (1226 Hz), which indicates a longer vocal tract—that is, a male speaker, in the case of Hungarian. If we place the reference values for vowels produced by male EP speakers in the vowel space in Figure 5 (grey squares), we see that [ɐ] and [i] produced by male EP speakers fall between /ɛ/ and /ø/, and /y/ and /ø/, respectively, which explains the unexpected results we obtained in our study.

The high frequency of [ɛ] in Hungarian may provide an explanation of the finding that the EP [ɐ] was more systematically identified as /ɛ/ than as /ø/ (68.5% vs. 30.1%). According to Gósy (2004), [ɛ] is the most frequent Hungarian vowel, with an occurrence of 11.4%, while [ø] has a frequency of only 1.1%. This suggests that vowel perception of Hungarian speakers may be biased towards [ɛ] (p. 89). Furthermore, contrary to /ø/, neither the EP [ɐ] nor the Hungarian /ɛ/ are rounded; hence, these latter two may be considered more similar to each other (than to Hungarian /ø/).

Lastly, the identification of the EP [ɛ] was balanced between /ɛ/ and /e:/. As shown in Figure 5, reference F1 and F2 frequency values for EP [ɛ] produced by male speakers (grey square) are in the vowel space between the Hungarian values for /ɛ/ and /e:/, but closer to /e:/. Recall that the median goodness-of-fit rating in the identification of the EP vowel [ɛ] was 2, both as /ɛ/ and /e:/, suggesting that Hungarian listeners did not consider EP [ɛ] to be a good fit for the Hungarian /ɛ/ or /e:/, the two most frequently picked candidates.

Considering that 22 participants reported having an intermediate or advanced level in German, we also investigated the possible effect of this knowledge in the categorization of the two EP vowels absent in the participants' L1, [ɐ] and [i]. Additionally, due to the difficulties observed with the perception of the EP /ɛ/ and the acoustic distance between this vowel and the correspondent Hungarian category, we also analyzed the data for the categorization of [ɛ]. Results from chi-square tests run for each vowel's trials revealed no effect of German knowledge in any of the cases (no significant effects were found).

2.3. Perceptual Overlap and Predictions for Discrimination

The data collected in the categorization experiment allowed us to put forward predictions for the discrimination of EP vowel contrasts involving [ɐ] and [i] directly or indirectly by Hungarian listeners. Following Guion et al. (2000), we calculated the fit index (proportion of categorization × median of goodness-of-fit rating) for the identification of each EP vowel (Table 4).⁵ Although the calculation of the fit index presents some methodological questions (Tyler 2021) and thus, the information displayed in Table 4 should be interpreted cautiously, it remains an effective way of showing problematic contrasts.

As shown in Table 4 below, the EP vowels [a] and [i] (and, to a lesser extent, [e]) reached the highest fit indexes. This suggests that these were accepted as good instances of the respective Hungarian category, and conversely, the remaining EP vowels, [ɐ], [ɛ], and [i], which presented the lowest fit index, were perceived as “non-native” or deviant categories. We can also observe several situations of perceptual overlap, that is, when two EP vowels are perceived as a single Hungarian category: [ɐ] and [ɛ] (identified as /ɛ/), [ɛ] and [e] (identified as /e:/), [ɐ] and [i] (identified as /ø/), and [e] and [i] (identified as /i/).

Table 4. Fit indexes for the tested EP vowels.

EP Vowel	L1 Chosen Categories	Proportion of Identification	Mean Goodness-of-Fit Rating	Fit Index
[a]	/a:/	0.98	4	3.91
[ɐ]	/ɛ/	0.69	3	2.06
	/ø/	0.30	3	0.90
[ɛ]	/ɛ/	0.59	2	1.18
	/e:/	0.40	2	0.80
[e]	/e:/	0.82	3	2.46
	/i/	0.16	2	0.32
[i]	/y/	0.73	3	2.19
	/ø/	0.25	2	0.50
[i]	/i/	0.98	3	2.94

With respect to predictions for EP contrast discrimination, we calculated the perceptual overlap scores following Faris et al. (2018). To this purpose, we took the sum of the smaller percentages when both EP vowels in a given contrast were identified as the same L1 category, excluding identification rates below chance level, since these are not systematic choices and might correspond to involuntary responses (due to tiredness, for example). These perceptual overlap scores were calculated based on group means (displayed in Table 3). For example, the EP [ɛ] was identified by the Hungarian listeners as /ɛ/ (in 58.8% of the trials) or as /e:/ (in 39.8% of the trials). On the other hand, the EP [e] was also identified by the Hungarian listeners as /ɛ/ or as /e:/. However, given that identification of [e] as /ɛ/ was below chance level, the probability of both the EP [ɛ] and [e] being perceived as /ɛ/ should not be considered for the calculation of the perceptual overlap score. Therefore, only the probability of the [ɛ]-[e] contrast being perceived as /e:/ was considered. The overlap score for this contrast was 39.8%, corresponding to the lowest value between 58.8% (identification of [ɛ] as /e:/) and 81.9% (identification of [e] as /e:/). In other words, there was a 39.8% probability that the [ɛ]-[e] contrast would be perceived as a unique vowel by the Hungarian listeners: /e:/.

Table 5 displays the perceptual overlap scores for each of the contrasts in focus.

Table 5. Perceptual overlap scores, calculated on group means. Underline, the lowest value between the identifications of each vowel of the EP contrast.

EP Contrast	Identifications Above Chance-Level	Perceptual Overlap Score (%)
[ɐ]-[ɛ]	[ɐ] as /ɛ/: 68.5%; [ɛ] as /ɛ/: 58.5%	58.5%
[ɛ]-[e]	[ɛ] as /e:/: 39.8%; [e] as /e:/: 81.9%	39.8%
[ɐ]-[i]	[ɐ] as /ø/: 30.1%; [i] as /ø/: 25.0%	25.0%
[e]-[i]	[e] as /i/: <u>16.2%</u> ; [e] as /i/: 98.1%	16.1%

Other than observing performance at the group level, we also aimed at investigating how the scores calculated based on group means relate to perceptual overlap at the individual level, and which of the two measures—based on group means or based on individual performances—is a better predictor for L3 vowel discrimination. To this purpose, we grouped Hungarian participants by levels of perceptual overlap scores for each EP contrast. The levels of overlap scores were calculated in the following way: For each pair of EP vowels, each participant completed six trials (2 vowels × 3 tokens). If, in the six trials, the participant identified both vowels of the contrast as one L1 category, it means that 100% overlap occurred. If, in six trials, the two vowels were identified five times as one Hungarian category, then 83.3% overlap occurred, and so on. We then determined the proportion of Hungarian listeners for each level, in each EP contrast. For example, in the case of the EP [ɐ] and [ɛ], only one participant identified these vowels always as /ɛ/;

that is, only in 1.4% of the participants did we observe a total (100%) perceptual overlap. On the other hand, 65 participants (90.3%) identified the two EP vowels as /ε/ in only three of the six trials (i.e., 50% of the cases). Therefore, for most of the participants, a 50% perceptual overlap was observed. Results for perceptual overlap scores based on individual performances are presented in Table 6.

Table 6. Perceptual overlap based on individual results. In grey, the most frequent perceptual overlap levels for each contrast.

	Levels of Perceptual Overlap	Participants	
		Number	Percentage
[ɐ] and [ɛ] identified as /ε/	100.0%	1	1.4%
	83.3%	2	2.8%
	66.7%	2	2.8%
	50.0%	65	90.3%
	33.3%	2	2.8%
[e] and [ɛ] identified as /e:/	100.0%	5	6.9%
	83.3%	31	43.1%
	50.0%	24	33.3%
	33.3%	9	12.5%
	16.7%	3	4.2%
[ɐ] and [i] identified as /ø/	50.0%	3	4.2%
	33.3%	45	62.5%
	16.7%	20	27.8%
	0.0%	4	5.6%
[e] and [i] identified as /i/	83.3%	3	4.2%
	66.7%	28	38.9%
	50.0%	38	52.8%
	33.3%	3	4.2%

As shown in Tables 5 and 6, the values obtained from individual scores are not entirely in line with the scores obtained from group mean results. First, according to the overlap scores based on group means, we obtained the value of 58.8% for the [ɐ]-[ɛ] contrast, which was the highest perceptual overlap score based on group means. However, according to individual results, 90.3% of the participants identified both [ɐ] and [ɛ] as /ε/ only in half of the trials. Regarding the [ɛ]-[e] contrast, when both vowels were identified as /e:/, we obtained a 39.8% score based on group means. However, 33.3% of the participants revealed a 50% perceptual overlap, and 43.1% of the participants showed an 83.3% overlap score. This suggests that, if we consider individual answers, the EP [ɛ]-[e] contrast may be, in fact, more problematic than the [ɐ]-[ɛ] contrast. A similar situation was observed for the [ɐ]-[i] and [e]-[i] contrasts. If we consider overlap scores based on group means, we can conclude that the [ɐ]-[i] contrast should pose more problems to Hungarian listeners than the [e]-[i] contrast, since the former exhibited a higher perceptual overlap score than the latter (25.0% against 16.1%). However, when we look at the proportion of participants for each level of overlap, we should conclude that the [e]-[i] contrast will be more problematic for Hungarian listeners than [ɐ]-[i]. In this contrast, for 95.8% of the participants the overlap level was equal to or higher than 50%. As for [ɐ]-[i], only 4.2% of the participants displayed an overlap level of 50%, while the remaining section of participants displayed lower levels of overlap.

In summary, it follows from both measures that the [ɐ]-[ɛ] and [e]-[ɛ] EP contrasts pose more difficulties in discrimination for Hungarian listeners than the [ɐ]-[i] and [e]-[i] contrasts. According to group means' results, we obtained the following hierarchy for the predicted discrimination difficulties (from most difficult to easiest): [ɐ]-[ɛ] > [ɛ]-[e] > [ɐ]-[i] > [e]-[i]. However, according to individual results, the hierarchy is as follows: [ɛ]-[e] > [ɐ]-[ɛ] > [e]-[i] > [ɐ]-[i]. To assess whether perceptual overlap affects EP contrast perception and which perceptual overlap measure—scores based on group means or on individual

results—does more accurately predict discrimination difficulties, we designed an oddity discrimination task, which we describe next (Experiment 2).

3. Experiment 2

Experiment 2 consisted of an oddity discrimination task in which participants were presented with the EP vowel contrasts where perceptual overlap was observed in Experiment 1: [v]-[ɛ], [ɛ]-[e], [v]-[i], and [e]-[i]. To test the effect of perceptual overlap, we added three contrasts in which Hungarian listeners did not display perceptual overlap: [a]-[v], [ɛ]-[i], and [e]-[i]. These contrasts correspond to the relevant EP stressed–unstressed pairs involving the target EP vowels, [v] and [i]. The trials consisted of sequences of three EP vowels, two vowels of the same category and one of a different category (the odd). For example, for the [v]-[ɛ] contrast, the following sequences were created: [v]-[ɛ]-[ɛ], [v]-[v]-[ɛ], [v]-[ɛ]-[v], [ɛ]-[ɛ]-[v], [ɛ]-[v]-[ɛ], and [ɛ]-[v]-[v]. Other than change trials, we included catch trials in the material as well, in which participants were presented with sequences of three tokens from the same category (e.g., [v]-[v]-[v] and [ɛ]-[ɛ]-[ɛ]). Consequently, in each trial, participants had two tasks: decide whether the three tokens belong to a unique category or not, and if not, identify the odd token.

3.1. Method

3.1.1. Participants

For the second experiment, a second group of Hungarian participants was recruited. Considering that this experiment was part of a wider project in perceptual training for L3 Portuguese, participants were recruited from beginner Portuguese language courses in different universities across Hungary at the onset of learning. Seventy participants completed the experiment and reported not having attended a Portuguese course before or having any significant (previous or present) contact with Portuguese. Participants completed the experiment in the first week of the course. They were aged 18 to 44 (mean age = 22.1), and 51 were female speakers. All participants reported intermediate or advanced knowledge of English. Other languages reported at an intermediate or advanced level were Spanish (n = 21), German (n = 18), Italian (n = 12), French (n = 7), Romanian (n = 3), Slovak, Catalan, and Dutch (n = 1 for each of these languages).

Furthermore, a second group of EP native speakers was also recruited to serve as a baseline group. The Portuguese group included 13 participants (10 female), all from the dialectal area of Lisbon, and aged 21 to 43 (mean age = 25.1).

Neither the Hungarian nor the Portuguese participants selected reported hearing problems.

3.1.2. Stimuli

The stimuli were the same as in Experiment 1, and each trial included a sequence of three tokens, one from each of the three Portuguese female speakers. Following previous studies (Escudero and Wanrooij 2010; Flege and Mackay 2004), we set the inter stimulus interval (ISI) to 1200 ms. By combining a longer ISI with speaker variability, we aimed at promoting high-level acoustic processing (categorical processing) of speech sounds, rather than low-level phonetic discrimination of stimuli (that also includes speaker identity) during task completion.

3.1.3. Procedure

The experiment was built and conducted online in Gorilla Experiment Builder (Anwyl-Irvine et al. 2020). Similar to Experiment 1, we built two versions, one presented to the Hungarian participants and another to the baseline group, with Portuguese native speakers. These versions were identical in structure and design, except for the instructions and auditory tokens for the familiarization tasks, which were in the L1 of each group.

The main trials were preceded by a familiarization task, in which participants were presented with eight trials containing [tV] tokens produced in the participant's L1. The fa-

miliarization task included sequences that were easier to discriminate than others (e.g., [to]-[ti]-[to]), as well as sequences that were acoustically more similar than others (e.g., [tɛ]-[te]-[tɛ]). Immediate feedback was provided after each trial in written form (“Correct!” or “Incorrect... try again”). If the answer was incorrect, participants had to repeat the same trial immediately after, until they reached the correct answer. No limit of attempts was set. The main task consisted of the 48 trials with the EP [gV] tokens: 42 change trials (7 contrasts × 6 orders: AAB/ABA/BAA/BBA/BAB/ABB) and 6 catch trials (6 vowels × 1 order). These trials were presented in a single block and were randomized between participants. No feedback was provided in the main trials.

Figure 6 displays a computer screen of one trial.

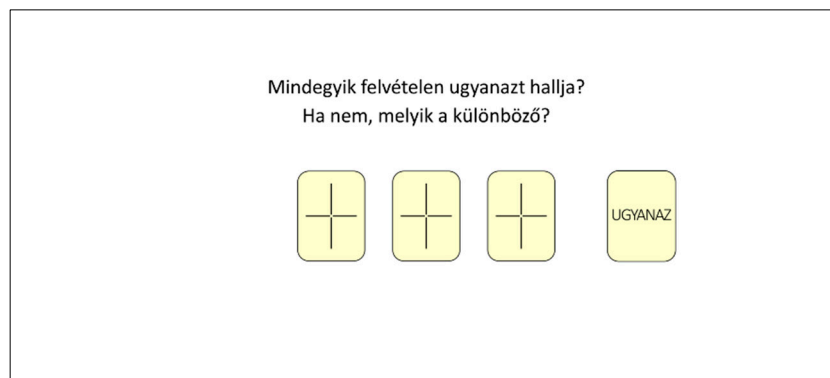


Figure 6. Computer screen of the trials presented to the Hungarian participants in Experiment 2. (“Do you hear the same in each of the recordings? If not, which one is different?,” where the fourth answer is “The same.”)

3.1.4. Analyses

Data were processed and analyzed with the software R version 4.4.1 (R Core Team 2021). We collected correct answers and analyzed the results in two ways. First, we calculated A-prime (A') scores for each participant and each contrast, integrating results from catch trials and change trials. With this, we aimed at obtaining “an unbiased measure of perceptual sensitivity by taking into account the responses to the different trials and the catch trials” (Guion et al. 2000, p. 2718). A' scores were calculated as the proportion of hits and false alarms (correct and incorrect selection of an odd item, respectively), and they ranged between 0 = chance-level discrimination and 1 = perfect discrimination (Makowski 2018)⁶. We then ran linear mixed-effect models with the LMER function (lme4 package, Bates et al. 2015) on these scores. Second, we analyzed responses in change trials and catch trials separately by building linear mixed-effect logistic models (correct answer = 1, incorrect answer = 0) with the GLMER function (lme4 package, Bates et al. 2015). For both analysis— A' scores and responses in change and catch trials—we created null models (with the participant as the random effect) and conducted successive ANOVA tests on the log-likelihood ratio to assess whether adding the fixed effects significantly contributed to explaining the variance. In Appendix B, we report the best fitting models obtained. We also conducted pairwise comparisons of least-square means with the EMMEANS function (emmeans package, Lenth 2024), with Bonferroni corrections.

3.2. Results

A-prime scores are presented in Table 7. While in some contrasts the score was near 1 (indicating a very good discrimination), in other cases it was below 0.7 or even close to 0.5 (indicating limited sensitivity to the contrast, to different extents). Furthermore, although Portuguese listeners outperformed the Hungarian listeners in every contrast, in some cases differences between Portuguese and Hungarian listeners were more pronounced than in others, especially in two contrasts: [ɐ]-[i] and [ɐ]-[ɛ].

Table 7. Mean A' scores for the Hungarian group (n = 70) and the Portuguese group (n = 13).

Contrast Type	Contrast	Participants' L1	Mean A'	sd	se	ci
with perceptual overlap	[ɐ]-[ɛ]	Hungarian	0.68	0.21	0.03	0.05
		Portuguese	0.93	0.13	0.04	0.08
	[ɛ]-[e]	Hungarian	0.69	0.23	0.03	0.05
		Portuguese	0.86	0.15	0.04	0.09
	[e]-[i]	Hungarian	0.82	0.16	0.02	0.04
		Portuguese	0.87	0.13	0.04	0.08
[ɐ]-[i]	Hungarian	0.55	0.24	0.03	0.06	
	Portuguese	0.96	0.04	0.01	0.02	
without perceptual overlap	[a]-[ɐ]	Hungarian	0.90	0.11	0.01	0.03
		Portuguese	0.92	0.13	0.03	0.08
	[ɛ]-[i]	Hungarian	0.82	0.17	0.02	0.04
		Portuguese	0.97	0.05	0.01	0.03
[e]-[i]	Hungarian	0.76	0.17	0.02	0.04	
	Portuguese	0.89	0.19	0.05	0.12	

A' = A-prime score; sd = standard deviation; se = standard error of the mean; ci = confidence interval (95%).

We first compared results while taking into consideration the presence/absence of perceptual overlap—that is, comparing A' scores for the [ɐ]-[ɛ], [ɛ]-[e], [ɐ]-[i], and [e]-[i] contrasts to scores for the [a]-[ɐ], [ɛ]-[i], and [e]-[i] contrasts (Figure 7). We ran a linear mixed-effect model with “contrast type” (contrasts with perceptual overlap vs. contrasts without perceptual overlap) and “L1” (Hungarian vs. Portuguese) as the fixed effects and found a significant contrast type \times L1 interaction ($\chi^2(2) = 30.871, p < 0.001$).

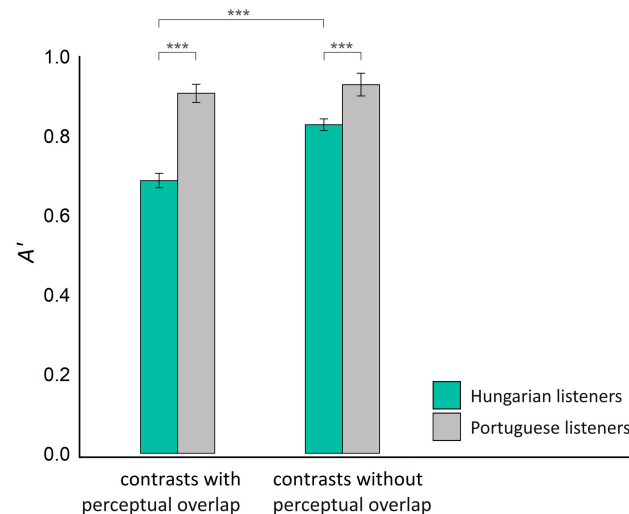


Figure 7. Mean A' scores for Hungarian and Portuguese listeners as a function of presence/absence of perceptual overlap. Error bars represent one standard error. ***: $p \leq 0.001$.

Pairwise comparisons revealed that Hungarian listeners performed significantly worse than Portuguese listeners in both the “contrasts with perceptual overlap” and “contrasts without perceptual overlap” conditions ($p < 0.001$ in both cases). Furthermore, the Hungarian group displayed a significantly lower accuracy in the “contrasts with perceptual overlap” condition than in the “contrasts without perceptual overlap” condition ($p < 0.001$).

Regarding discrimination of vowel contrasts with perceptual overlap (Figure 8), linear mixed-effect modeling revealed a vowel contrast \times L1 interaction ($\chi^2(4) = 42.878, p < 0.001$). The pairwise comparisons showed that in two EP contrasts, [ɐ]-[ɛ] and [ɐ]-[i], Hungarian listeners displayed significantly lower accuracy in discrimination than the Portuguese

participants ($p = 0.0074$ and $p < 0.001$, respectively). As for differences between contrast discrimination, for the Hungarian listeners, pairwise comparisons showed that the [v]-[i] contrast had a significantly lower accuracy than the remaining contrasts ([v]-[i] vs. [v]-[ε]: $p = 0.0075$; [v]-[i] vs. [ε]-[e]: $p = 0.0048$; [v]-[i] vs. [e]-[i]: $p < 0.001$). Moreover, the [e]-[i] contrast was significantly higher in accuracy than the [ε]-[e] and [v]-[ε] contrasts ($p = 0.0058$ and $p < 0.001$, respectively).

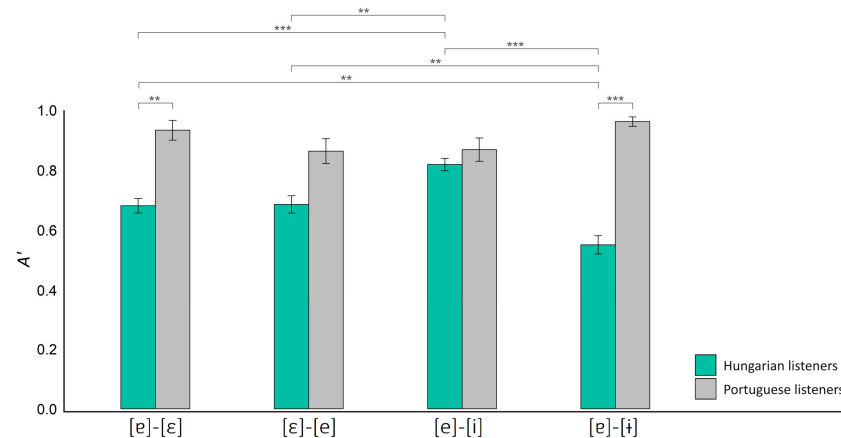


Figure 8. Mean A' scores for Hungarian and Portuguese listeners in vowel contrasts with perceptual overlap as a function of contrast. Error bars represent one standard error. **: $p \leq 0.01$; ***: $p \leq 0.001$.

Other than looking into the A' scores, we also investigated the mean accuracy of Hungarian listeners for change and catch trials separately (Figure 9).

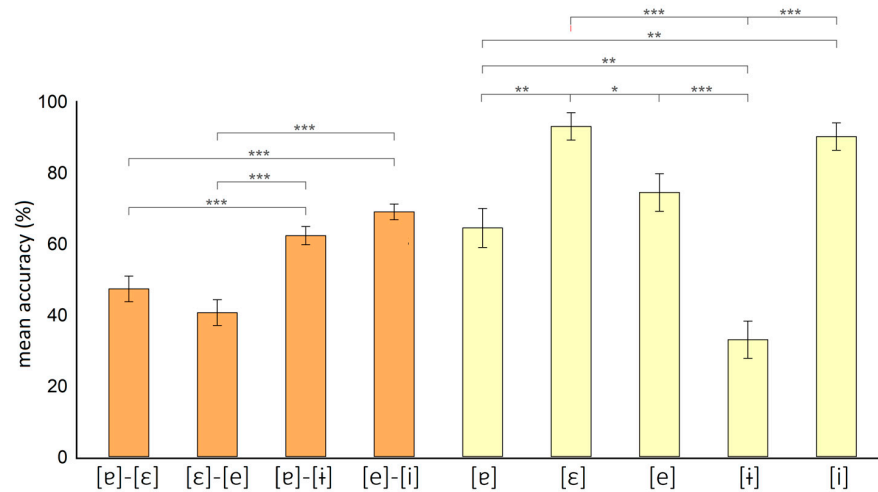


Figure 9. Mean accuracy for Hungarian listeners in vowel contrasts with perceptual overlap as a function of vowel contrast/vowel (change trials in orange, catch trials in yellow). Error bars represent one standard error. *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$.

A linear mixed-effect logistic model showed that the trial type (change vs. catch) significantly affected mean identification accuracy (%) ($\chi^2(1) = 32.216$, $p < 0.001$), with participants showing more difficulties in the change trials than in the catch trials. In the change trials, Hungarian listeners displayed a significantly lower accuracy in the [v]-[ε] and [ε]-[e] contrasts compared to [v]-[i] and [e]-[i] ($p < 0.001$ in all pairwise comparisons). Regarding the catch trial results, we observed that accuracy in the perception of the vowel [i] was significantly lower (below 40%) than in the remaining vowels ([i] vs. [v]: $p = 0.0026$; [i] vs. [ε]: $p < 0.001$; [i] vs. [e]: $p < 0.001$; [i] vs. [i]: $p < 0.001$). The EP [v] also posed problems

to the Hungarian participants, although to a lesser extent ([ɛ] vs. [e]: $p = 0.0018$; [ɛ] vs. [i]: $p = 0.0026$; [e] vs. [i]: $p = 0.0062$).

Additionally, we looked at the effect of non-nativeness (absence or presence of the vowels in the L1 vocalic system), comparing accuracy in discrimination between trials that included the EP vowels [ɛ] or [i] with the remaining trials. In the analysis of the A' scores, we found a significant effect of non-nativeness ($\chi^2(1) = 22.821$, $p < 0.001$); that is, Hungarian listeners had significantly more difficulties in the perception of contrasts that included [ɛ] or [i] than in other contrasts. As for the results of the separate change trials and catch trials, we did not find any significant effect of non-nativeness when analyzing change trials only. However, in the catch trials, comparing the accuracy rates for [ɛ] and [i] with the accuracy rates for the vowels [ɛ], [e], and [i], we found a significant effect ($\chi^2(1) = 56.182$, $p < 0.001$).

Finally, similar to the analysis of the results collected in Experiment 1, we also investigated a possible effect of other languages spoken by the participants. In the present experiment, 30% of the participants reported knowledge of Spanish and 26% knowledge of German. Regarding the effect of knowledge of Spanish, we did not find any significant result. As for the effect of German, results from a linear mixed-effect model in the A' scores showed a significant interaction, L3 German knowledge \times non-nativeness, in discrimination of the EP contrasts ($\chi^2(2) = 22.861$, $p < 0.001$). Pairwise comparisons showed that while participants without knowledge of German had significantly more difficulties in the contrasts with vowels that are absent from the Hungarian vowel system than in contrasts with familiar vowels, this was not observed in the results of participants with German (Figure 10, on the left). Furthermore, we also investigated the results of the catch trials for the EP [ɛ] and [i]. We found a significant effect of L3 German knowledge in the perception of [ɛ] ($\chi^2(1) = 4.1741$, $p = 0.041$; Figure 10, on the right side), but not in the case of [i].

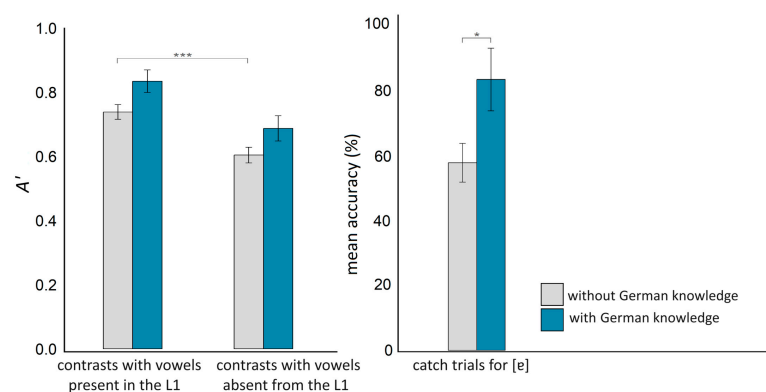


Figure 10. On the left side, A' scores as a function of knowledge of German and presence/absence of unfamiliar vowels in the trials. On the right side, accuracy rates for the perception of [ɛ] in the catch trials as a function of knowledge of German. Error bars represent one standard error. *: $p \leq 0.05$; ***: $p \leq 0.001$.

4. General Discussion

The connection between the findings from the categorization and the discrimination tasks conducted in the present study provides insight into how perceptual overlap can predict discrimination in L3. Our first aim was to observe whether perceptual overlap affects discrimination in L3 Portuguese. The data analysis revealed that Hungarian listeners display significantly more difficulties when discriminating EP contrasts that are perceived with overlap than contrasts in which perceptual overlap is absent. Such results are in line with previous research on L2 speech perception (Flege and Mackay 2004; Tyler et al. 2014; Faris et al. 2018; Elvin et al. 2021). Additionally, we aimed at investigating the extent to which the amount of perceptual overlap relates to difficulties in the discrimination of L3 Portuguese contrasts. To this purpose, we compared two measures for perceptual overlap scores: (i) We calculated the overlap scores considering group means (Table 5), and (ii) we determined the possible levels of overlap scores for each EP contrast, calculating

the proportion of participants for each case (Table 6). With the latter, we aimed to account for intersubject variability. According to both approaches, the [v]-[ɛ] and [ɛ]-[e] contrasts displayed higher perceptual overlap than [v]-[i] and [e]-[i], and consequently, we expected the former two contrasts to cause more discrimination difficulties to Hungarian listeners. However, comparison between the two approaches presented the following discrepancy: If we consider perceptual overlap scores that are measured based on group means, [v]-[ɛ] should present more difficulties than [ɛ]-[e], and [v]-[i] should present more difficulties than [e]-[i]. If our predictions are based on the measures of perceptual overlap calculated from individual results, [ɛ]-[e] should present more difficulties than [v]-[ɛ], and [e]-[i] should present more difficulties than [v]-[i].

A-prime scores obtained in the oddity discrimination task were not completely in line with either measure. First, based on the data collected, we obtained the following hierarchy of discrimination difficulties (from most difficult to easiest): [v]-[i] > [v]-[ɛ] = [ɛ]-[e] > [e]-[i]. Accuracy for [ɛ]-[e] was not statistically different from that obtained in [v]-[ɛ], and [e]-[i] was discriminated significantly better than [v]-[i]. Second, in the [ɛ]-[e] and [e]-[i] contrasts, Hungarian listeners did not perform differently than Portuguese listeners. In summary, the predictions for discrimination based on perceptual overlap scores calculated from the categorization task results were not confirmed by the A' scores.

The analysis of mean accuracy for separate change and catch trials may provide further insight into the results. Accuracy rates for the change trials were in line with the predictions established from the categorization task, considering both group means and individual performances: [v]-[ɛ] and [ɛ]-[e] form one group, with higher perceptual overlap scores and a lower accuracy in discrimination, while [v]-[i] and [e]-[i] constituted another group, with a lower level of perceptual overlap and higher accuracy rates. The results for the catch trials are less straightforward. The high accuracy rate for [i] was in line with the high categorization score obtained in Experiment 1 (98.1%). After [i], [ɛ] reached the highest accuracy, followed by [e]. These results contrast with the poor discrimination abilities in the [ɛ]-[e] contrast in the change trials. Considering that, in the categorization task, nearly 40% of the trials with [ɛ] and [e] were identified by the Hungarian participants as /e:/, low accuracy in the change trials and high accuracy in the catch trials was to be expected. As for the two remaining vowels, [v] and [i], they exhibited the lowest discrimination accuracy rates in the catch trials. Recall that these two vowels are the EP vowels that are absent from the Hungarian listeners' L1, and, in our analyses, there was a significant effect of non-nativeness. Moreover, in the case of [i], the mean accuracy was only 32.9%, indicating a categorization problem. Although this vowel was identified as /y/ in 73.1% of the trials, this result merely informs us that, for Hungarian listeners, /y/ was the closest category to the EP vowel. If participants had been presented with a "not a speech sound" option, for example, identification as /y/ may not have been so robust.

In addition to observing the relation between perceptual overlap, categorization, and discrimination, we also aimed at investigating the effect of knowledge of other languages in the perception of the EP [v] and [i] vowels. Other than English, a considerable number of participants in our study reported knowledge of two other languages. In Experiment 1, German was the second most frequently spoken additional language. The analysis of the results showed that the identification of EP vowels into Hungarian categories was not influenced by knowledge of German. Regarding Experiment 2, Spanish was the most frequently reported additional spoken language, followed by German. The analysis revealed that knowledge of Spanish did not affect the A' scores. However, we found that knowledge of German positively influenced the perception of the EP contrasts that included non-native vowels ([v]-[ɛ] and [v]-[i]). The vowels [v] and [i] are absent from participants' L1 system, as well as from English (Carr 2019) and from Spanish (Torres-Tamarit 2020). However, [v] is present in the German vowel inventory (Wiese 2000), and better accuracy in the perception of vowel [v] was observed in the participants with German knowledge (Figure 10), suggesting that these participants may have used their previous experience with this vowel. It is, however, important to keep in mind that, even though specific vowels

are transcribed with the same IPA symbol, they may have slightly different realizations in different languages. Likewise, vowel qualities transcribed using different symbols may be more similar in their quality than the phonetic notation may suggest. As a result, we should be cautious while trying to draw conclusions based on the IPA symbols. Furthermore, since we did not collect data on the perception of German vowels by the Hungarian participants, we cannot assume that participants who reported knowledge of German have in fact established a target-like perceptual representation of the German [ɐ]. Nevertheless, the improved performance in discrimination of the [ɐ]-[ɛ] and [ɐ]-[i] contrasts, as well as for the vowel [ɐ], suggests that Hungarian participants with some knowledge of German may have already created a new category for a sound close to the EP [ɐ]. This, in turn, helped these participants to better identify and distinguish EP [ɐ] than those who did not have previous experience with German.

One question arises: Why was there an effect of knowledge of German in the discrimination task but not in the identification task? These results may be due to the activation of different language modes (Grosjean 2001). Evidence of language modes in L2 was found in Yazawa et al. (2020). In this study, Japanese native speakers who were learners of English L2 were tested in the discrimination of the contrast /i:/-/ɪ/. The results showed that participants relied more on duration (a key feature in their L1) when they were told they would hear Japanese sounds, and on spectral information (as in the L2) when instructed that they would hear English sounds. However, the stimulus set was identical in the two sessions. In our study, in the identification task, participants were not only presented with instructions in their L1, but they also had to categorize the EP vowels in real words from the L1. Thus, it is logical to assume that they activated their L1, and consequently, they resorted to the L1 inventory in that task. As for the discrimination task, although the instructions were also given in Hungarian, in the main trials, the participants' L1 was absent from the stimuli. It is possible, then, that knowledge of other languages, namely, their knowledge of German vowels, was more readily available. In other words, the activation of the L1 or L2, or the L1 and L2, may be related to the nature of the task: A task using L1 references may induce the activation of the L1, and a discrimination task may promote the use of previously learned languages other than the L1. An example for the latter is the study by Luo et al. (2020). When conducting an AXB task with native Mandarin speakers who had English as the L2, the authors found a combined L1/L2 transfer in the perception of L3 Cantonese vowel length contrasts.

Another interesting point in our results is the fact that the advantages displayed by Hungarian participants with knowledge of German were only observed for the EP vowel [ɐ], which, as we reported, also exists in the German inventory. The fact that those participants did not perform better than the others in the perception of [i] contradicts Onishi's (2016) assertion that the knowledge of an L2 entails a general advantage in perceiving novel L3 contrasts.

In summary, the results of the present study suggest that, although perceptual overlap affects L3 perception, the amount of perceptual overlap may not explain, per se, the differences among L3 contrast discrimination, and other factors may also play a role. First, the robustness of categorization can affect perception, even when the vowels exist in the L1. For example, [i], [e], and [ɛ] were categorized by the Hungarians into the correspondent L1 categories in 98.1%, 81.9%, and 58.8% of the trials, respectively. The results from the discrimination task showed that Hungarians were able to discriminate the [e]-[i] contrast significantly better than [ɛ]-[e]. Second, non-nativeness, that is, the absence of the vowels in the L1 system, also affects perception, hindering discrimination. In our study, participants displayed significantly more difficulties in discriminating contrasts with the EP vowels [ɐ] and [i], which are absent from their L1, than other contrasts. Lastly, knowledge of other previously acquired languages should also be accounted for in L3 discrimination. In the discrimination task, Hungarian speakers with knowledge of German, a language that has the vowel [ɐ] in its system, performed significantly better in trials with this vowel.

A possible limitation of the present study is that we did not test and quantify participants' knowledge of additional spoken languages. For the present sample, administering

a language proficiency test for each reported language would have been impossible, as the participants reported too many languages. Moreover, a general proficiency test might not even capture the relevant phonological functioning of the L2s. To assess the possible effects of the representation of L2 phonology, one should test perception (and production) in these additional languages as well (Cabrelli 2013). However, similar to testing language proficiency, this was not feasible in the present experiment and could serve as a possible line of inquiry for future research.

A second remark regarding the design of our study is related to the auditory stimuli. Although we introduced speaker variability in the stimuli, vowels were produced in one context, [gV]. It would be valuable to collect perception data from EP vowels produced in other consonantal contexts, since this can be a significant factor in non-native vowel perception (Bohn and Steinlen 2003).

Third, previous studies have pointed out the importance of accounting for individual differences. In our study, we attempted to include intersubject variability by establishing a measure for perceptual overlap based on individual performances. However, this measure was not a better predictor for discrimination difficulties than the measure based on group means. This may result from the number of trials each participant completed for the categorization of each EP vowel, which was only three. A higher number of repetitions may contribute to more robust results at the individual level, hence leading to more reliable predictions for discrimination based on these results. Further investigations considering the limitations mentioned above would greatly contribute to our understanding of perception processes in the L3 context.

Lastly, different groups of participants were used for Experiment 1 and Experiment 2. Experiment 1 was designed as a preliminary test to identify problematic contrasts that were then tested in Experiment 2. Therefore, there was a time interval between the completion of the two experiments, and due to the recruiting method and anonymization of data, it was not possible to contact and ask the participants who completed the identification task to complete a second experiment. However, we believe that this option did not compromise the results, due to the robust number of participants in each group. Moreover, the use of different groups prevented a potential learning effect.

5. Conclusions

The present study aimed at investigating whether, similarly to what has been observed in L2 speech learning, perceptual overlap affects vowel discrimination in L3. We conducted two experiments: a categorization task and an oddity discrimination task. For each experiment, we recruited Hungarian native speakers, who were tested in EP oral vowels and vowel contrasts. The results confirmed that perceptual overlap increased the difficulty of discrimination of EP vowel contrasts by the Hungarian listeners. However, when comparing the amount of perceptual overlap with the discrimination results, we did not find a straightforward relationship. Specifically, although Hungarian listeners displayed greater discrimination difficulties in the contrasts with a higher amount of perceptual overlap, the hierarchy of difficulties in discrimination was not consistent with the hierarchy of perceptual overlap scores. Importantly, the “non-nativeness” of the target L3 vowels as well as knowledge of other languages significantly influenced discrimination of EP vowel contrasts. We observed that knowledge of languages other than L1 can positively affect L3 perception, although this effect seems to be selective.

In closing, this study is original in its observation of L3 perception in two languages that have been remarkably understudied: European Portuguese and Hungarian. Furthermore, the importance of examining L3 perception in naïve listeners, such as in the present study, lies in its potential to inform L3 teachers about the challenges their learners may encounter with L3 phonology from the onset of the learning process. This information can consequently help in the development of empirically grounded didactic materials tailored for this early stage of L3 learning.

Author Contributions: Conceptualization and methodology, G.T., S.C. and A.D.; investigation, G.T.; formal analysis, G.T.; resources, G.T., S.C. and A.D.; data curation, G.T.; writing—original draft preparation, G.T.; writing—review and editing, G.T., S.C. and A.D.; visualization, G.T. and S.C.; supervision, S.C. and A.D.; project administration, G.T. and S.C.; funding acquisition, G.T., S.C. and A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the FCT—the Portuguese Foundation for Science and Technology—as part of the project with CLUNL, School of Social Sciences and Humanities, NOVA University Lisbon, 1069-061 Lisboa, Portugal (UIDB/LIN/03213/2020 and UIDP/LIN/03213/2020); an individual PhD grant from the FCT (2020.05740.BD); the Bolyai János Research Scholarship of the Hungarian Academy of Sciences; the ÚNKP-20-5 and ÚNKP-21-5 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund; and the National Research, Development and Innovation Office within the framework of the Thematic Excellence Program “Community building: family and nation, tradition and innovation” with TKA-DAAD grant No. 177375 and NKFIH grant No. FK128814.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all participants involved in the study for both experiments.

Data Availability Statement: The following supporting information can be downloaded at www.osf.io/w8kyp: stimuli, acoustic data for the stimuli, results, and statistical analyses.

Acknowledgments: We would like to express our gratitude to the Hungarian and Portuguese volunteers who recorded the stimuli and piloted the experiments, as well as the participants. We also thank LABCC—the Communication Sciences Laboratory of FCSH NOVA University of Lisbon—as well as MTA—ELTE “Momentum” Lingual Articulation Research Group, Hungary, and the Department of Applied Linguistics and Phonetics, ELTE Eötvös Loránd University, for providing the recording equipment. Additionally, we would like to extend our appreciation to the anonymous reviewers who significantly contributed to the improvement of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Real words (orthographic stimuli) used in Experiment 1:

	Target L1 Vowel	Real Word for Main Trials	Real Word for Familiarization Trials
Hungarian real words	[ɒ]	GARÁZS, [ˈgɒrɒːʒ], 'garage'	TANÁR, [ˈtɒnɒːr], 'teacher'
	[aː]	GÁBOR, [ˈgɑːbɒr], 'proper name'	TÁBOR, [ˈtɑːbɒr], 'camp'
	[ɛ]	GERELY, [ˈgɛrɛj], 'javelin'	TEREM, [ˈtɛrɛm], 'room'
	[eː]	GÉPÉSZ, [ˈgɛːpɛːs], 'machinist'	TÉTEL, [ˈtɛːtɛl], 'item'
	[i]	GITÁR, [ˈgitaːr], 'guitar'	TITOK, [ˈtitok], 'secret'
	[o]	GONOSZ, [ˈgonos], 'wicked'	TOJÁS, [ˈtojaːʃ], 'egg'
	[ø]	GÖRÉNY, [ˈgøreːɲ], 'ferret'	TÖMEG, [ˈtømɛg], 'mass'
	[u]	GULYÁS, [ˈgujaːʃ], 'goulash'	TUDÁS, [ˈtudaːʃ], 'knowledge'
	[y]	GÜGYÖG, [ˈgyʝøg], 'to babble'	TÜKÖR, [ˈtykør], 'mirror'
Portuguese real words	[a]	GATO, [ˈgatu], 'cat'	TAPA, [ˈtapɐ], 'cover'
	[ɐ]	GAVETA, [gɐˈvetɐ], 'drawer'	TAREFA, [tɐˈɾɛfɐ], 'task'
	[ɛ]	GUERRA, [ˈgɛrɐ], 'war'	TELA, [ˈtɛlɐ], 'canvas'
	[e]	GUÊ, [ˈgɛ], 'gee'	TEMA, [ˈtemɐ], 'theme'
	[i]	GUERREAR, [gɨˈriːar], 'to fight'	TERROR, [tiˈror], 'terror'
	[i]	GUIO, [ˈgɨ], 'money'	TINA, [ˈtɨnɐ], 'bowl'
	[ɔ]	GOLA, [ˈgɔlɐ], 'collar'	TOMA, [ˈtømɐ], 'take'
	[o]	GOTA, [ˈgotɐ], 'drop'	TODA, [ˈtodɐ], 'every'
	[u]	GULA, [ˈgulɐ], 'gluttony'	TUDO, [ˈtudu], 'everything'

Appendix B

The best-fitting model for the effect of perceptual overlap (presence/absence) and L1 (Hungarian/Portuguese) for A' scores collected in Experiment 2:

	Estimate	Std. Error	df	t Value	p Value
(Intercept)	0.68396	0.01499	150.29919	45.628	$<2.0 \times 10^{-16}$
contrast.typestressed.unstressed	0.14457	0.01800	498.00001	8.0330	6.94×10^{-15}
L1Portuguese	0.22245	0.03788	150.29919	5.8730	2.64×10^{-8}
contrast.typestressed.unstressed: L1Portuguese	-0.12256	0.04548	498.00001	-2.6950	0.00727

Number of observations = 581; participants = 83; AIC = -165.1; BIC = -138.9; log-likelihood = 88.6

R syntax: `lmer(a.prime ~ contrast.type*L1 + (1 | participant.ID), REML = FALSE, data = vowels)`

The best-fitting model for the effect of specific contrasts ([e]-[ε], [ε]-[e], [e]-[i], [e]-[i]) and L1 (Hungarian/Portuguese) for A' scores collected in Experiment 2:

	Estimate	Std. Error	df	t Value	p Value
(Intercept)	0.68125	0.026963	304.540895	25.266	$<2.0 \times 10^{-16}$
vowel.contrastgaveta_guerrear	-0.13036	0.035239	243	-3.699	0.000267
vowel.contrastgue_guito	0.13696	0.035239	243	3.887	0.000131
vowel.contrastguerra_gue	0.00423	0.035239	243	0.12	0.904638
L1Portuguese	0.25144	0.068129	304.540895	3.691	0.000265
vowel.contrastgaveta_guerrear: L1Portuguese	0.15920	0.089041	243	1.788	0.075025
vowel.contrastgue_guito: L1Portuguese	-0.20139	0.089041	243	-2.262	0.024597
vowel.contrastguerra_gue: L1Portuguese	-0.07378	0.089041	243	-0.829	0.408154

Number of observations = 332; participants = 83; AIC = -43.8; BIC = -5.7; log-likelihood = 31.9
R syntax: `lmer(a.prime ~ vowel.contrast*L1 + (1 | participant.ID), REML = FALSE, data = vowels)`

The best-fitting model for the effect of non-nativeness (contrasts with vowels absent from the L1/contrasts with vowels existing in the L1) for A' scores collected for the Hungarian participants in Experiment 2:

	Estimate	Std. Error	df	t Value	p Value
(Intercept)	0.61607	0.02191	168.65311	28.12	$<2.0 \times 10^{-16}$
non.nativenessno	0.13577	0.02765	210	4.91	1.83×10^{-6}

Number of observations = 280; participants = 70; AIC = 11.8; BIC = 26.3; log-likelihood = -1.9
R syntax: `lmer(a.prime ~ non.nativeness + (1 | participant.ID), REML = FALSE, data = vowels)`

The best-fitting model for the effect of L3 German knowledge (with/without) and non-nativeness (contrasts with vowels absent from the L1/contrasts with vowels existing in the L1) for A' scores collected for the Hungarian participants in Experiment 2:

	Estimate	Std. Error	df	t Value	p Value
(Intercept)	0.67685	0.04224	175.76197	16.023	$<2.0 \times 10^{-16}$
L3_Germanno	-0.08182	0.04901	175.76197	-1.669	0.09682
non.nativenessno	0.14525	0.05453	210	2.664	0.00832
L3_Germanno:non.nativenessno	-0.01276	0.06326	210	-0.202	0.84032

Number of observations = 280; participants = 70; AIC = 10.4; BIC = 32.2; log-likelihood = 0.8
R syntax: lmer(a.prime ~ L3_German*non.nativeness + (1 | participant.ID))

The best-fitting model for the effect of specific contrasts ([e]-[ε], [ε]-[e], [e]-[i], [e]-[i]) for accuracy rates in change trials for the Hungarian participants in Experiment 2:

	Estimate	Std. Error	z Value	p Value
(Intercept)	-0.1145	0.098	-1.168	0.2427
vowel.contrastgaveta_guerrear	0.6105	0.1404	4.348	0.0000137
vowel.contrastgue_guito	0.9063	0.1439	6.298	3.02×10^{-10}
vowel.contrastguerra_gue	-0.2714	0.1395	-1.946	0.0516

Number of observations = 1680; participants = 70; AIC = 2236.3; BIC = 2263.4;
log-likelihood = -1113.2
R syntax: glmer(correct.answers ~ vowel.contrast + (1 | participant.ID), data = vowelsHU, family = binomial)

The best-fitting model for the effect of specific contrasts ([e], [ε], [e], [i], [i]) for accuracy rates in catch trials for the Hungarian participants in Experiment 2:

	Estimate	Std. Error	z Value	p Value
(Intercept)	0.5878	0.2494	2.356	0.018454
vowel.contrastguerra	1.9772	0.5269	3.753	0.000175
vowel.contrastgue	0.4731	0.3701	1.278	0.201214
vowel.contrastguerrear	-1.3024	0.3563	-3.655	0.000257
vowel.contrastguito	1.6094	0.4701	3.424	0.000617

Number of observations = 350; participants = 70; AIC = 353.2; BIC = 376.4; log-likelihood = -170.6
R syntax: glmer(correct.answers ~ vowel.contrast + (1 | participant.ID), data = vowelsHU, family = binomial)

Notes

- 1 In previous studies cited in the present article, different terms are used to refer to tasks in which the participants are asked to match an auditory stimulus into a native category, namely, categorization (Elvin et al. 2021), classification (Flege and Mackay 2004), identification (Guion et al. 2000), or perceptual assimilation (Faris et al. 2018). In describing our studies, we opted to use the term “categorization” on the grounds that we are observing the phenomenon of categorical perception, that is, the labeling of speech sounds varying along a continuum into abruptly distinct categories (different vowels). Therefore, for consistency, we will use the label “categorization task” throughout the article.
- 2 In Flege and Mackay (2004), the authors present an overlap score of 72%, not 73%. However, since we only have access to rounded values for identification, our calculation is of 73%.
- 3 The standard dialect of European Portuguese is spoken in the region of Lisbon and Coimbra, and it is “the most commonly heard on radio and television,” and also “accepted as a reference for teaching Portuguese as a second language” (Mateus and d’Andrade 2000, p. 4).
- 4 The formant dispersion parameter is an indicator of body size, and it is calculated as the “average distance between each adjacent pair of formants” (Fitch 1997, p. 1216).
- 5 Although in Guion et al. (2000) the authors calculate the fit index only for categorizations above 30%, we considered all identification rates above chance level (11.1%).
- 6 The formula for the A’ score calculation, retrieved from Snodgrass et al. (1985, p. 451), was adapted to the following formula for R, in which “hit” represents the proportion of trials when the odd item was correctly selected and “fa” the false alarms—that is, the proportion of catch trials when the odd item was incorrectly selected: $database\$a.prime = ifelse(database\$hit > database\$fa, 0.5 +$

$$\frac{((\text{database}\$hit - \text{database}\$fa) \times (1 + \text{database}\$hit - \text{database}\$fa))}{(4 \times \text{database}\$hit \times (1 - \text{database}\$fa))}, \text{ifelse}(\text{database}\$hit == \text{database}\$fa, 0.5, 0.5 - \frac{((\text{database}\$fa - \text{database}\$hit) \times (1 + \text{database}\$fa - \text{database}\$hit))}{(4 \times \text{database}\$fa \times (1 - \text{database}\$hit))}).$$

References

- Andrade, Amália. 2020. Vocalismo. In *Gramática Do Português. Volume III*. Edited by Eduardo B. P. Raposo, Maria Fernanda B. do Nascimento, Maria Antónia C. da Mota, Luísa Segura, Amália Mendes and Amália Andrade. Lisboa: Fundação Calouste Gulbenkian, pp. 3241–30.
- Anwyl-Irvine, Alexander, Jessica Massonnié, Adam Flitton, Natasha Kirkham, and Jo Evershed. 2020. Gorilla in Our Midst: An Online Behavioral Experiment Builder. *Behavior Research Methods* 52: 388–407. [CrossRef]
- Aoyama, Katsura, James Emil Flege, Susan G. Guion, Reiko Akahane-Yamada, and Tsuneo Yamada. 2004. Perceived Phonetic Dissimilarity and L2 Speech Learning: The Case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics* 32: 233–50. [CrossRef]
- Audacity Team. 2020. *Audacity(R): Free Audio Editor and Recorder* [Computer Application]. Version 2.4.2. Available online: <https://www.audacityteam.org/> (accessed on 30 June 2024).
- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using Lme4. *Journal of Statistical Software* 67: 1–48. [CrossRef]
- Best, Catherine T. 1995. A Direct Realist View of Cross-Language Speech Perception. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Speech Research*. Edited by Winifred Strange. Timonium: York Press, pp. 171–204.
- Best, Catherine T., and Michael D. Tyler. 2007. Nonnative and Second-Language Speech Perception. Commonalities and Complementaries. In *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. Edited by Murray J. Munro and Ocke-Schwen Bohn. Amsterdam: John Benjamins, pp. 13–34.
- Best, Catherine T., Gerald W. McRoberts, and Elizabeth Goodell. 2001. Discrimination of Non-Native Consonant Contrasts Varying in Perceptual Assimilation to the Listener's Native Phonological System. *Journal of Acoustical Society of America* 109: 775–94. [CrossRef] [PubMed]
- Boersma, Paul, and David Weenink. 2020. Praat: Doing Phonetics by Computer [Computer Program]. Version 6.1.16. Available online: <https://www.fon.hum.uva.nl/praat/> (accessed on 30 June 2024).
- Bohn, Ocke-Schwen, and Aanja Steinlen. 2003. Consonantal context affects cross-language perception of vowels. Paper Presented at the 15th International Congress of Phonetic Sciences, Barcelona, Spain, 3–9 August 2003; Edited by Maria-Josep Solé and Joaquin Romero. ICPhS Archive. pp. 2289–92. Available online: <http://www.internationalphoneticassociation.org/icphs/icphs2003> (accessed on 30 June 2024).
- Bohn, Ocke-Schwen, and Catherine T. Best. 2012. Native-Language Phonetic and Phonological Influences on Perception of American English Approximants by Danish and German Listeners. *Journal of Phonetics* 40: 109–28. [CrossRef]
- Cabrelli, Jennifer. 2012. L3 Phonology. An Understudied Domain. In *Third Language Acquisition in Adulthood*. Edited by Jennifer Cabrelli Amaro, Suzanne Flynn and Jason Rothman. Amsterdam: John Benjamins, pp. 33–60. [CrossRef]
- Cabrelli, Jennifer. 2013. Methodological Issues in L3 Phonology. *Studies in Hispanic and Lusophone Linguistics* 6: 101–17. [CrossRef]
- Carr, Philip. 2019. *English Phonetics and Phonology: An Introduction*, 3rd ed. Hoboken: Wiley-Blackwell.
- Castelo, Adelina, and Rita Nazaré Santos. 2017. As Vogais Do Português Entre Aprendentes Chineses e Suas Implicações No Desenvolvimento de Um Programa de Português. In *Português Como Língua Estrangeira, de Herança e Materna: Abordagens, Contextos e Práticas*. Roosevelt: Boavista Press, pp. 123–36.
- Chen, Sally, and Janice Fon. 2007. The Effect of Phonetic Similarity on L2 Speakers in Discerning L1 and L2 Consonants. *The Journal of the Acoustical Society of America* 121: 3073–74. [CrossRef]
- Correia, Susana, Joseph Butler, Marina Vigário, and Sónia Frota. 2015. A Stress 'Deafness' Effect in European Portuguese. *Language and Speech* 58: 48–67. [CrossRef]
- Delgado-Martins, Maria Raquel. 1986. *Sept Etudes Sur La Perception*. Lisboa: Instituto Nacional de Investigação Científica.
- Deme, Andrea. 2014. Intelligibility of Sung Vowels: The Effect of Consonantal Context and the Onset of Voicing. *Journal of Voice* 28: 523.e19–523.e25. [CrossRef]
- Elvin, Jaydene, Daniel Williams, Jason A. Shaw, Catherine T. Best, and Paola Escudero. 2021. The Role of Acoustic Similarity and Non-Native Categorisation in Predicting Non-Native Discrimination: Brazilian Portuguese Vowels by English vs. Spanish Listeners. *Languages* 6: 44. [CrossRef]
- Elvin, Jaydene, Paola Escudero, and Polina Vasiliev. 2014. Spanish Is Better than English for Discriminating Portuguese Vowels: Acoustic Similarity versus Vowel Inventory. *Frontiers in Psychology* 5: 1188. [CrossRef] [PubMed]
- Escudero, Paola. 2005. Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization (Publication No. 113). Doctoral dissertation, LOT, Dissertation Series Utrecht University, Utrecht, The Netherlands.
- Escudero, Paola, and Karin Wanrooij. 2010. The Effect of L1 Orthography on Non-Native Vowel Perception. *Language and Speech* 53: 343–65. [CrossRef] [PubMed]
- Escudero, Paola, and Paul Boersma. 2004. Bridging the Gap Between L2 Speech Perception Research and Phonological Theory. *Studies in Second Language Acquisition* 26: 551–85. [CrossRef]

- Faris, Mona M., Catherine T. Best, and Michael D. Tyler. 2016. An Examination of the Different Ways That Non-Native Phones May Be Perceptually Assimilated as Uncategorized. *The Journal of the Acoustical Society of America* 139: EL1–EL5. [CrossRef] [PubMed]
- Faris, Mona M., Catherine T. Best, and Michael D. Tyler. 2018. Discrimination of Uncategorized Non-Native Vowel Contrasts Is Modulated by Perceived Overlap with Native Phonological Categories. *Journal of Phonetics* 70: 1–19. [CrossRef]
- Fitch, W. Tecumseh. 1997. Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *Journal of Acoustical Society of America* 102: 1213–22. [CrossRef]
- Flege, James Emil. 1995. Second Language Speech Learning: Theory, Findings, and Problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Timonium: York Press, pp. 233–76.
- Flege, James Emil. 2003. A Method for Assessing the Perception of Vowels in a Second Language. In *Issues in Clinical Linguistics*. Edited by Elisabetta Fava and Alberto Mioni. Padova: Unipress, pp. 19–43.
- Flege, James Emil, and Ian R. A. Mackay. 2004. Perceiving Vowels in a Second Language. *Studies in Second Language Acquisition* 26: 1–34. [CrossRef]
- Flege, James Emil, and Ocke-Schwen Bohn. 2021. The Revised Speech Learning Model (SLM-R). In *Second Language Speech Learning. Theoretical and Empirical Progress*. Edited by Rtree Wayland. Cambridge: Cambridge University Press, pp. 3–83. [CrossRef]
- Frota, Sónia, Marina Vigário, Fernando Martins, and Marisa Cruz. 2010. FrePOP (Version 1.0). Lisboa: Laboratório de Fonética (CLUL), Faculdade de Letras da Universidade de Lisboa. Available online: <https://frepop.letras.ulisboa.pt/index.php> (accessed on 30 August 2023).
- Georgiou, Georgios P. 2021. Toward a New Model for Speech Perception: The Universal Perceptual Model (UPM) of Second Language. *Cognitive Processing* 22: 277–89. [CrossRef]
- Georgiou, Georgios P., and Dimitra Dimitriou. 2023. Perception of Dutch Vowels by Cypriot Greek Listeners: To What Extent Can Listeners' Patterns Be Predicted by Acoustic and Perceptual Similarity? *Attention, Perception, and Psychophysics* 85: 2459–74. [CrossRef]
- Georgiou, Georgios P., Aretousa Giannakou, and Katarzyna Alexander. 2024. Perception of Second Language Phonetic Contrasts by Monolinguals and Bidialectals: A Comparison of Competencies. *Quarterly Journal of Experimental Psychology*, 1–15. [CrossRef]
- Gósy, Mária. 2004. *Fonetika, a Beszéd Tudományja*. Budapest: Osiris Kiadó.
- Grosjean, Francois. 2001. The Bilingual's Language Modes. In *One Mind, Two Languages: Bilingual Language Processing*. Edited by Janet Nicol. Oxford: Blackwell, pp. 1–22.
- Guion, Susan G., James Emil Flege, Reiko Akahane-Yamada, and Jessica C. Pruitt. 2000. An Investigation of Current Models of Second Language Speech Perception: The Case of Japanese Adults' Perception of English Consonants. *The Journal of the Acoustical Society of America* 107: 2711–24. [CrossRef] [PubMed]
- Halácsy, Péter, András Kornai, László Németh, András Rung, István Szakadát, and Viktor Trón. 2003. A Szószablya Projekt. In *I. Magyar Számítógépes Nyelvészeti Konferencia Előadásai (Proceedings)*. Edited by Zoltán Alexin and Dóra Csendes. Szeged: Szegedi Tudományegyetem Informatikai Tanszék, p. 298. Available online: <http://acta.bibl.u-szeged.hu/id/eprint/59176> (accessed on 30 June 2024).
- Hammarberg, Björn. 2010. The languages of the multilingual: Some conceptual and terminological issues. *International Review of Applied Linguistics in Language Teaching* 48: 91–104. [CrossRef]
- Harnsberger, James D. 2001. On the Relationship between Identification and Discrimination of Non-Native Nasal Consonants. *The Journal of the Acoustical Society of America* 110: 489–503. [CrossRef] [PubMed]
- Honbolygó, Ferenc, Andrea Kóbor, and Valéria Csépe. 2017. Cognitive Components of Foreign Word Stress Processing Difficulty in Speakers of a Native Language with Non-Contrastive Stress. *International Journal of Bilingualism* 23: 1–15. [CrossRef]
- Ilari, Rudolfo. 2013. O Português no contexto das línguas românicas. In *Gramática do Português. Volume I*. Edited by Eduardo B. P. Raposo, Maria Fernanda B. do Nascimento, Maria Antónia C. da Mota, Luísa Segura and Amália Mendes. Lisboa: Fundação Calouste Gulbenkian, pp. 49–66.
- Ingvallson, Erin M., James L. McClelland, and Lori L. Holt. 2011. Predicting Native English-like Performance by Native Japanese Speakers. *Journal of Phonetics* 39: 571–84. [CrossRef]
- Iverson, Paul, and Bronwen G. Evans. 2009. Learning English Vowels with Different First-Language Vowel Systems II: Auditory Training for Native Spanish and German Speakers. *The Journal of the Acoustical Society of America* 126: 866–77. [CrossRef]
- Kiss, Gábor. 1985. A Magyar Magánhangzók Első Két Formánsának Meghatározása Szintetizált Hangmintákat Felhasználó Percepció Kísérlet Segítségével. *Nyelvtudományi Közlemények* 87: 160–74.
- Larson-Hall, Jenifer. 2004. Predicting Perceptual Success with Segments: A Test of Japanese Speakers of Russian. *Second Language Research* 20: 33–76. [CrossRef]
- Lenth, Russell V. 2024. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R Package Version 1.10.2. Available online: <https://cran.r-project.org/web/packages/emmeans/emmeans.pdf> (accessed on 30 June 2024).
- Liu, Jiaqi, and Jiayan Lin. 2021. A Cross-Linguistic Study of L3 Phonological Acquisition of Stop Contrasts. *SAGE Open* 11: 1–15. [CrossRef]
- Liu, Jiaqi, Ting Zeng, and Xiuchuan Lu. 2019. Challenges in Multi-Language Pronunciation Teaching: A Cross-Linguistic Study of Chinese Students' Perception of Voiced and Voiceless Stops. *Circulo de Linguística Aplicada a La Comunicacion* 79: 99–118. [CrossRef]
- Luo, Jingxin, Vivian Guo Li, and Peggy Pik Ki Mok. 2020. The Perception of Cantonese Vowel Length Contrast by Mandarin Speakers. *Language and Speech* 63: 635–59. [CrossRef] [PubMed]

- Makowski, Dominique. 2018. Compute Signal Detection Theory Indices with R. *The Psycho Blog*. Available online: <https://neuropsychology.github.io/psycho.R/2018/03/29/SDT.html> (accessed on 16 July 2024).
- Markó, Alexandra. 2017. Hangtan. In *Nyelvtan*. Edited by András Imrényi, Nóra Kugler, Mária Ladányi, Alexandra Markó, Szilárd Tátrai and Gábor Tolcsvai Nagy. Budapest: Osiris Kiadó, pp. 75–203.
- Markó, Alexandra, Márton Bartók, Tamás Gábor Csapó, Andrea Deme, and Tekla Etelka Grácsi. 2019. The Effect of Focal Accent on Vowels in Hungarian: Articulatory and Acoustic Data. Paper presented at the 19th International Congress of Phonetic Sciences, Melbourne, Australia, 5–9 August 2019; Edited by Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren. Canberra: Australasian Speech Science and Technology Association Inc., pp. 2715–19. Available online: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS_2764.pdf (accessed on 30 June 2024).
- Mateus, Maria Helena Mira, and Ernesto d’Andrade. 2000. *The Phonology of Portuguese*. Oxford: Oxford University Press.
- Oliveira, Inês Gomes. 2006. A Aquisição Do Sistema Vocálico Por Falantes de Português Europeu Como Língua Não Materna. Unpublished Master’s dissertation, Universidade de Lisboa, Lisboa, Portugal.
- Onishi, Hiromi. 2016. The Effects of L2 Experience on L3 Perception. *International Journal of Multilingualism* 13: 459–75. [CrossRef]
- Peirce, Jonathan, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. 2019. PsychoPy2: Experiments in Behavior Made Easy. *Behavior Research Methods* 51: 195–203. [CrossRef] [PubMed]
- Peperkamp, Sharon, and Emmanuel Dupoux. 2002. A Typological Study of Stress ‘Deafness’. In *Laboratory Phonology 7*. Edited by Carlos Gussenhoven and Natasha Warner. Berlin and New York: Mouton de Gruyter, pp. 203–40. [CrossRef]
- Peperkamp, Sharon, Inga Vendelin, and Emmanuel Dupoux. 2010. Perception of Predictable Stress: A Cross-Linguistic Investigation. *Journal of Phonetics* 38: 422–30. [CrossRef]
- Pereira, Isabel. 2020. Acento de Palavra. In *Gramática Do Português. Volume III*. Edited by Eduardo B. P. Raposo, Maria Fernanda B. do Nascimento, Maria Antónia C. da Mota, Luísa Segura, Amália Mendes and Amália Andrade. Lisboa: Fundação Calouste Gulbenkian, pp. 3397–425.
- R Core Team. 2021. R: A Language and Environment for Statistical Computing. version 4.4.1. Vienna: R Foundation for Statistical Computing.
- Rakerd, Brad. 1984. Vowels in Consonantal Context Are Perceived More Linguistically than Are Isolated Vowels: Evidence from an Individual Differences Scaling Study. *Perception & Psychophysics* 35: 123–36. [CrossRef]
- Rice, Keren. 2007. Markedness in Phonology. In *The Cambridge Handbook of Phonology*. Edited by Paul de Lacy. New York: Cambridge University Press, pp. 79–98. [CrossRef]
- Siptár, Péter, and Miklós Törkenczy. 2000. *The Phonology of Hungarian*. Oxford: Oxford University Press.
- Snodgrass, Joan Gay, Gail Levy-Berger, and Martin Haydon. 1985. *Human Experimental Psychology*. Oxford: Oxford University Press.
- Stan, Irina. 2024. A Cross-Linguistic Study of L3 Speech Perception of Voiced and Voiceless Stops. *Estudios de Fonética Experimental* 33: 137–52.
- Strange, Winifred, Erika S. Levy, and Franzo F. Law. 2009. Cross-Language Categorization of French and German Vowels by Naïve American Listeners. *The Journal of the Acoustical Society of America* 126: 1461–76. [CrossRef]
- Sturgis, Patrick, Caroline Roberts, and Patten Smith. 2014. Middle Alternatives Revisited: How the Neither/nor Response Acts as a Way of Saying ‘I Don’t Know’? *Sociological Methods and Research* 43: 15–38. [CrossRef]
- Torres-Tamarit, Francesc. 2020. Phonemic contrast and neutralization. In *The Routledge Handbook of Spanish Phonology*. Edited by Sonia Colina and Fernando Martínez-Gil. Abingdon and New York: Taylor & Francis, pp. 3–33.
- Trofimovich, Pavel, Wendy Baker, James Emil Flege, and Molly Mack. 2003. Second-Language Sound Learning in Children and Adults: Learning Sounds, Words, or Both? In *27th Boston University Conference on Language Development*. Edited by Barbara Beachley, Amanda Brown and Frances Conlin. Somerville: Cascadilla Press, pp. 775–86.
- Tyler, Michael D. 2021. Perceived Phonological Overlap in Second-Language Categories: The Acquisition of English /r/ and /l/ by Japanese Native Listeners. *Languages* 6: 4. [CrossRef]
- Tyler, Michael D., Catherine T. Best, Alice Faber, and Andrea G. Levitt. 2014. Perceptual Assimilation and Discrimination of Non-Native Vowel Contrasts. *Phonetica* 71: 4–21. [CrossRef]
- Wang, Di, and Claire Nance. 2023. Third Language Phonological Acquisition: Understanding Sound Structure in a Multilingual World. *Language and Linguistics Compass* 17: e12497. [CrossRef]
- Wiese, Richard. 2000. *The Phonology of German*. Oxford: Oxford University Press.
- Wrembel, Magdalena, Marta Marecka, and Romana Kopečková. 2019. Extending Perceptual Assimilation Model to L3 Phonological Acquisition. *International Journal of Multilingualism* 16: 513–33. [CrossRef]
- Yazawa, Kakeru, James Whang, Mariko Kondo, and Paola Escudero. 2020. Language-Dependent Cue Weighting: An Investigation of Perception Modes in L2 Learning. *Second Language Research* 36: 557–81. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.