PERCEPTUAL AND ACOUSTIC SIMILARITY OF MODERN GREEK AND STANDARD GERMAN VOWELS: DOES CROSS-LINGUISTIC ACOUSTIC SIMILARITY PREDICT PERCEPTION OF NON-NATIVE VOWELS?

Payam Ghaffarvand-Mokari, University of Eastern Finland, Finland
Maria Lialiou, University of Potsdam, Germany

Thirty-four native Modern Greek (MG) listeners with no German language experience completed a perceptual assimilation task in which they categorized tokens of Standard German (SG) vowels in terms of their first language vowel categories. Further, cross-language acoustic similarity of MG and SG vowels was determined by a linear discriminant analysis (LDA) trained on the spectral and temporal information of the vowels. The findings indicate that cross-language acoustic similarity of native and non-native vowels do not always predict non-native perceptual assimilation patterns. Specifically, predictions based on the cross-language LDA were not consistent with the perceptual assimilations patterns for SG front rounded vowels. Finally, MG naïve listeners' discrimination of SG vowel pairs was also tested; all pairs were found to be easily discriminated by the listeners. We discuss these findings in the light of second language speech perception/acquisition models.

Cite as: Ghaffarvand-Mokari, P., & Lialiou, M.. (2022). Perceptual and acoustic similarity of Modern Greek and Standard German vowels: does cross-linguistic acoustic similarity predict perception of non-native vowels?. In J. Levis & A. Guskaroska (eds.). *Proceedings of the 12th Pronunciation in Second Language Learning and Teaching Conference*, held June 2021 virtually at Brock University, St. Catharines, ON. https://doi.org/10.31274/psllt.13270

INTRODUCTION

Over the past few decades, various models have been developed to explain how the native language (L1) phonological system affects non-native and second language (L2) perception and production. Among the influential L2 speech models, Perceptual Assimilation Model (PAM, Best, 1995) and Second-Language Linguistic Perception Model (L2LP; Escudero 2005) postulate that difficulties of non-native sound discrimination can be predicted based on the similarity of native and non-native sounds. PAM/PAM-L2 and L2LP particularly posit that discrimination of the non-native speech sounds and the subsequent L2 sound learning can be predicted based on how non-native sounds are mapped into native sound categories. This study first aims at providing predictions and test for discrimination of Standard German (SG) vowels by naïve Modern Greek (MG) listeners based on perceptual assimilations of L1 and L2 vowels.

According to L2LP, acoustic-phonetic similarity between native and target-language can also be predictive of non-native listeners' perceptual mapping of L2 to L1 sounds (Alispahic et al., 2017; Escudero & Chládková, 2010). L2LP claims a direct link between perception and production of speech sounds by proposing that, at the initial state of learning, an individual's perception of L2 sounds should closely match the acoustic properties of the learner's native language. The model postulates that any acoustic variation in native and non-native vowel production can influence speech perception. Therefore, listeners' initial perception of L2 vowels should closely match the acoustic properties of vowels as produced in the listener's L1. It is assumed that L2 perceptual categorization patterns and discrimination difficulties can be predicted through a detailed comparison of the acoustic features of the sounds of L1 and L2 (Elvin, 2016; Elvin et al., 2021; Williams & Escudero, 2014).

Several studies have evaluated the predictive power of acoustic similarities in non-native and L2 categorization/discrimination patterns (e.g., Alispahic et al., 2017; Elvin et al., 2021; Escudero & Vasiliev, 2011; Escudero, Sisinni & Grimaldi, 2014; Elvin, Escudero & Vasiliev, 2014; Gilichinskaya & Strange, 2010; Nishi et al., 2008). Alispahic et al. (2017) found that L1–L2 sounds' acoustic proximity predicts listeners' initial perception and discrimination patterns. The authors also found that acoustic similarities between L1 and L2 vowels were more important predictors of discrimination difficulties than vowel inventory size. Studies have reported that results of cross-language discriminant analyses (LDAs) trained on the acoustic properties of native vowels and tested on those of non-native vowels follow trends seen in listeners' perceptual assimilation patterns (Escudero & Vasiliev, 2011; Gilichinskaya & Strange, 2010). Gilichinskaya and Strange (2010) reported that Russian listeners' perceptual assimilation patterns for American English monophthongs mostly matched the patterns of cross-language acoustic similarity obtained from LDAs. However, cross-language acoustic comparisons have not always been successful in predicting listeners' perceptual assimilation behavior for all vowels. Strange et al. (2004) investigated the cross-language spectral similarity of North German and American English vowels and found that acoustic similarity determined by using LDAs did not always predict American English listeners' perceptual similarity patterns, especially for North German front rounded vowels. Regarding these contradictory findings in the literature, the main aim of the present study is to shed more light on the predictive power of cross-linguistic acoustic comparisons in the perception of non-native vowels by evaluating the degree of consistency between acoustic and perceptual similarity of SG and MG vowels.

Research Questions

The present study aims to answer the following questions.

> **RQ1** How do naïve MG listeners perceptually assimilate Standard German (SG) vowels to Modern Greek (MG) vowels?
> **RQ2** To what extent acoustic similarities between native and non-native vowels can predict non-native vowel perception patterns in the early stage of the SG acquisition by MG learners?
> **RQ3** Do perceptual similarities predict naïve listeners' discrimination between MG and SG vowels?

METHODS

Participants

Thirty-four native Greek listeners (17 females; 17 males) with a mean age of 23.9 years (range:18–34) participated in this study. All participants were native speakers of Standard Modern Greek with no prior experience with German (except for one female speaker whose data was excluded from all analyses). All participants reported normal or corrected-to-normal vision and no hearing or language impairments. Participants were all volunteers, they signed an informed consent form before entering the study, and were tested individually in a quiet room in Thessaloniki, Greece.

Cross-linguistic Perceptual Assimilation Task: Materials and Procedure

The stimuli were produced by four native Standard German speakers (two male and two female) with a mean age of 29.5 years (range: 26–33). Speakers were recorded in an anechoic chamber at the University of Potsdam in Germany using a unidirectional microphone (Audiotechnica 4028a) connected to an M-Audio Delta Audiophile sound card via a Phonic MM 1705A mixer. The recordings were saved in a single channel sampled at 44.1 kHz with 16 bits in wave-file format. Fifteen SG monophthongs /i, ɪ, yː, ʏ, uː, ʊ, eː, ɛː, ɛ, oː, øː, œ, ɔ, a, aː/ and three diphthongs /aʊ, aɪ, ɔʏ/ were recorded embedded in /bVp/, /dVk/, and /gVk/ non-words in the carrier phrase "Ich habe [__] gesagt" ("I have said [__]"). The participants were asked to speak naturally with a normal speaking rate, loudness, and intonation. To prevent incorrect pronunciations, the speakers were instructed to read aloud the real words provided in a word-list before uttering the non-word instances. For instance, they were asked to first read the word 'bieten' ("to offer something") before reading the carrier phrase "Ich habe [Biet] gesagt". A total of 648 stimuli were recorded (18 vowels × 3 contexts × 3 repetitions × 4 speakers). From the recorded data, for each vowel across contexts, the best token was chosen to serve as the stimuli in the tasks. The "best" token was defined as the most naturally produced one with a normal duration and minimal variation in the intensity and intonation. The final stimulus corpus consisted of 216 utterances (18 vowels × 3 contexts × 4 speakers). Speakers were students at the University of Potsdam, and they received course credits or payment for their participation.

For the perceptual assimilation task, participants completed a multiple forced-choice identification task in the computer software Praat (Boersma & Weenink, 2019). Randomized trials were presented in two blocks so that participants could have a short break between them. A familiarization block of 6 trials preceded the beginning of the task. The trials were presented at a comfortable loudness level over a wired Marshall Major III headphone. Participants were instructed to listen to each presented SG non-word containing the target vowel and choose the most similar MG vowel to the one presented by clicking one of the five response boxes on the screen. The labels of the five vowels /i/, /e/, /a/, /o/, and /u/ response boxes were given in Greek orthography. After selecting an MG vowel category, participants had to judge its perceived goodness on a five-point scale (1 = unlike to 5 = identical). Results from the three SG diphthongs were not included in further cross-linguistic analyses since no phonemic diphthongs exist in the MG vowel system and thus no comparison to the SG diphthongs could be made.

Cross-linguistic Acoustic Similarity of Vowels

For SG vowels, recordings from the two native female speakers of the perception test stimuli were used. A total of 270 tokens were used (2 speakers × 15 monophthongs × 3 contexts × 3 repetitions). For MG vowels, five female participants from the perceptual assimilation task were randomly selected to complete a production task. For MG vowels' recordings, we used the same procedure as for SG vowels. The five MG vowels /a, e, i, o, u/ were embedded in /bVp/, /gVk/, /dVk/ non-word contexts (as in perception test) and recorded in the carrier phrase "Θα πω [___] ξανά" ("I will say [___] again"). Recordings took place individually in a quiet room in Thessaloniki, Greece. To prevent incorrect pronunciations, the speakers were instructed to read aloud real words provided in the word-list before uttering the non-word instances. Recordings were made using an M-audio USB Producer microphone connected to a laptop computer and were saved in a single channel sampled at 44.1 kHz with 16 bits in wave-file format. A total of 225 stimuli were recorded (5 participants × 5 vowels × 3 repetitions × 3 contexts). The recordings were segmented and labeled manually using Praat.

Linear Discriminant Analyses

Following Strange et al. (2004), Escudero and Vasiliev (2011), and Elvin et al. (2021), a series of discriminant analyses were conducted as a quantitative measurement of acoustic similarity between the vowels of the SG and MG and later the results were compred to the results from the perceptual assimilation task. Linear discriminant analyses were established based on syllabic midpoint formant frequencies (F1, F2, F3 in Barks) and duration values of SG and MG vowels. Classification rules were specified by linear combinations of formant values and duration as input. For cross-language LDA, an MG vowel corpus served as a training set to establish centroids for each of the five MG categories which were used to classify SG vowel tokens. In other words, the test corpus (SG vowels) was categorized with respect to linear combination of the acoustic and temporal values of the input corpus (MG vowels).

Perceptual Discrimination Task: Material and Procedure

Prior to the perceptual assimilation task, participants completed an AXB auditory two-alternative forced-choice (2AFC) discrimination task (designed and run in Praat) in which listeners hear three sounds and have to decide if the middle token (X) sounds more like the first A or the third B. The task was used to test MG listeners' ability in discrimination of six SG vowel contrasts (/œ-ʏ/, /u:-y:/, /ʊ-ʏ/, /o:-ø:/, /ɛ-ɪ/, /i:-ɪ/). The contrasts were chosen so that each pair included at least a vowel that does not have any counterpart in the MG vowel system. The stimuli consisted of recorded vowels in /bVp/, /gVk/, and /dVk/ contexts produced by two female speakers (see subsection for the perceptual assimilation task). Each vowel contrast was examined in four different conditions; AAB, BBA, ABB, and BAA with the target vowel consistently presented as the second (X) sound. There were 144 trials in the task (6 vowel contrasts × 3 non-words × 4 conditions × 2 repetitions). The trials were presented in random order with 1.2s inter-stimulus interval, in two blocks. The participants were instructed to identify if the second sound (target) was more similar to the first or the third sound they heard by clicking on the relevant box on the screen. In each trial, A and B were tokens from different vowels, while the middle (target) token was always physically different from A and B tokens so that listeners could not make an acoustic identity judgment.

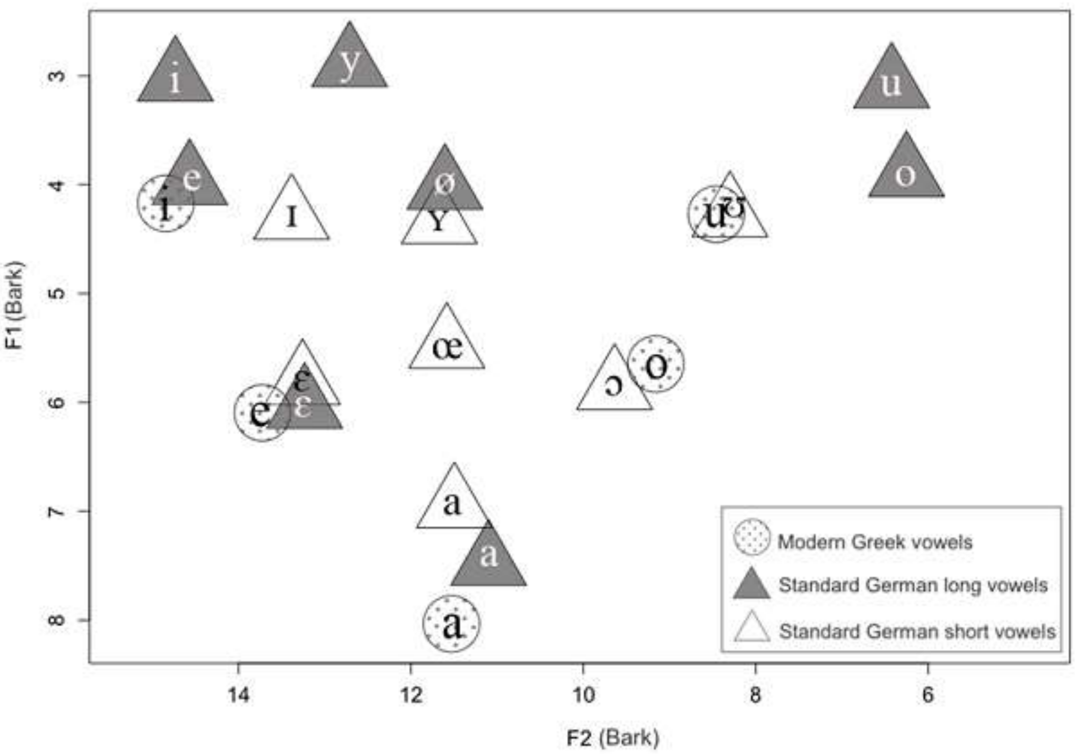
RESULTS

Figure 1 displays the Standard German (SG) and Modern Greek (MG) monophthongs in an F1 × F2 Bark scale based on averaged formant data from two native female SG and five native female MG speakers produced in three consonantal contexts. The Standard German vowels (triangles) are produced by two native female speakers, and Modern Greek vowels (circles) produced by five native female speakers. The grey and unfilled triangles show Standard German long and short vowels, respectively (Figure 1). SG and MG vowel systems are very different. SG has a wider vowel space than MG. Yet, some SG and MG vowels are very close to each other in the vowel space, like SG /e/

– MG /i/, SG /ɛ/ – MG /e/, and SG /ʊ/ – MG /u/. The larger vowel system of SG is expected to cause multiple non-native vowels to get perceptually (and acoustically) assimilated to a single MG vowel.

Figure 1

Average F1 and F2 values (in Bark scale)



Cross-linguistic perceptual and acoustic assimilation

As for RQ1, frequencies of selected MG vowel categories in the perceptual assimilation task were computed for each SG vowel (tallied across all listeners) and percentages were calculated. Goodness ratings assigned to each response category were computed for median response. Table 1 presents the modal MG response category (column 2) for each of SG vowels (column 1), percentage of the modal response and median rating for the goodness of fit for the modal response (column 3) across listeners. As it can be seen in Table 1, most of the SG vowels were consistently categorized as one of the MG categories with high median ratings. SG /a, aː, ɛ, ɛː, eː, iː, ɔ, ʏ, uː/ were categorized as one MG category with more than 90% consistency and median goodness rating of 5. This indicates that these vowels were judged to be very similar to some MG vowels. Less consistently, SG /ʊ/, /o/, /ɪ/ and /œ/ were categorized as MG vowels /u/ (83%), /u/ (72%), and /i/ (68%), and /e/ (45%), respectively.

To answer RQ2 – to what extent acoustic comparisons of the vowels systems can predict the perception patterns – a cross-language LDA was conducted to establish the similarity of SG and MG vowels. This classification provides a quantitative way of determining the acoustically closest native vowel for each non-native vowel. The LDA classification results (percentage classifications) show

how well SG vowels fall into the centers of gravity of the MG corpus tokens (Table 1; columns 4 and 5). The LDA results corresponded with the modal responses obtained in the perceptual assimilation task for SG /a, a:, ɛ, ɛ:, ɪ, e:, i:, ɔ, o:, ʊ, u:/. However, the modal MG responses for SG front rounded vowels /ʏ, ø, y:, œ/ were different across perceptual assimilations and LDAs. All SG front rounded vowels were perceptually assimilated to the MG back rounded /u/, while according to LDAs they were all categorized as MG front vowels /i/ and /e/.

Table 1

Perceptual assimilation and acoustic similarity results.

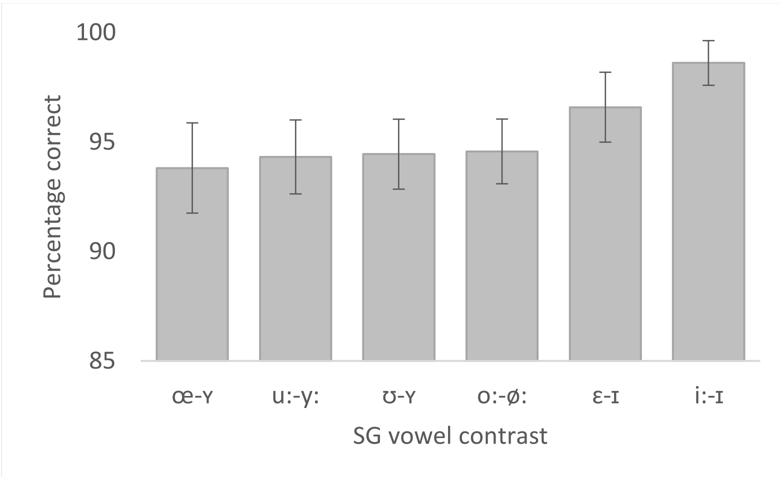
SG vowel	**Modal MG category**	**Percent modal Responses**	**Acoustic assimilation to L1 vowel**	**Percent modal LDA classification**
A	a	98 (5)	a	100
a:	a	100 (5)	a	100
ɛ	e	93 (5)	e	100
ɛ:	e	99 (5)	e	100
ɪ	i	69 (5)	i	81
e:	i	98 (5)	i	100
i:	i	100 (5)	i	100
ɔ	o	94 (5)	o	94
o:	u	72 (5)	u	81
ʊ	u	83 (5)	u	100
u:	u	100 (5)	u	100
ʏ	u	94 (5)	e	47
ø:	u	95 (4)	i	61
y:	u	98 (4)	i	100
œ	u	45 (4)	e	81

Note. Shaded rows indicate selected modal Modern Greek (MG) category for the Standard German (SG) vowel is not the same across perceptual and acoustic assimilation results

Finally, RQ3 asked how perceptual assimilation predicts discrimination of different SG vowel contrasts by MG listeners. Here we present the discrimination task results and we compare them to PAM’s predictions in the discussion section. Discrimination of six SG vowel contrasts by MG listeners was tested using an AXB paradigm. As shown in Figure 2, all six pairs were discriminated with high accuracy (> 90%). The SG contrast /i:-ɪ/ was discriminated with very high accuracy (98.6 %), followed by /ɛ-ɪ/ (96,6 %), /o:-ø:/ (94.6 %), /ʊ-ʏ/ (94.4 %), and /u:-y:/ (94.3 %), and /œ-ʏ/ (93.8 %).

Figure 2

Overall discrimination accuracy for six German vowel contrasts averaged across participants.



DISCUSSION

We first set out to study how Standard German (SG) vowels are perceptually assimilated to Modern Greek (MG) vowels by native listeners to provide predictions based on PAM (Best, 1995) for discrimination of SG vowels by MG learners at their early stages of L2 learning. As summarized above, most SG vowels were consistently perceptually assimilated to one of the MG vowel categories. According to Harnsberger (2001), the modal response choice of the listener group has to be 90 % or more of the listener group's responses to be considered as categorized. Based on this criterion, /œ-ʏ/, /ʊ-ʏ/, /o:-ø:/, /ɛ-ɪ/ and /i:-ɪ/ are considered as cases of uncategorized-categorized cases where each member of the contrast receives a different modal response by the group and one member of the contrast receives 90% or more of the listener group's responses and the other receives less than 90% of the responses. In the mentioned pairs, /ʏ/, /ø:/, /ɛ/, and /i:/ are considered as categorized while /œ/, /ʊ/, /o:/, and /ɪ/ as uncategorized. For uncategorized-categorized cases, discriminability is predicted to be very good. In line with this prediction, results of the discrimination task showed a high degree of accuracy for these pairs. According to PAM, if one of the members of the contrast has better goodness of fit for an L1 vowel category, it could result in relatively good discrimination (category-goodness assimilation). This was the case for /u:-y:/ where SG /u:/ was assimilated to MG /u:/ 100% of the times and with a median goodness rating of 5, while /y:/ was also consistently assimilated to MG /u/ but with a lower median goodness rating (= 4). Also in line with this prediction, /u:-y:/ pair was discriminated with high accuracy. It should be noted that there is no consensus on the best cut-off value for the perceptual assimilation categorizations. Considering different thresholds for the categorizations would result in different sets of predictions. More research is needed to find solutions for the cut-off value problem (see Tyler, 2021, for further discussion on this issue).

The main aim of this study was to test whether acoustic similarities between the native and non-native vowels can predict perceptual assimilation patterns. Linear discriminant analysis (LDAs) results based on acoustic and temporal information of the vowels were used to classify SG non-native vowel tokens in terms of MG native ones. The obtained results were compared to the perceptual assimilation patterns. We found that LDAs were not always successful in predicting perception patterns. The assimilation patterns of eleven out of fifteen SG vowels were consistent across acoustic and perceptual tests. Yet, the modal MG responses for SG front rounded vowels were different across perceptual assimilations and LDAs. All SG front rounded vowels were perceptually assimilated to MG back rounded /u/, while according to LDAs, they were all categorized as MG front vowels /i/ and /e/. These results are largely in line with previous findings from cross-language perceptual and acoustic comparisons of North German – American English vowels (Strange et al., 2004), and Russian – English vowels (Gilichinskaya and Strange, 2010). Strange et al. (2004) and Gilichinskaya and Strange (2010) have reported a partial success for acoustic comparisons in predicting perceptual assimilations of non-native sounds. Comparable to our findings, Strange et al. (2004) reported that North German /y:/ and /ø:/ were acoustically similar to front American English vowels although they were generally perceptually assimilated to American English back rounded vowels.

According to the L2LP model (Escudero, 2005), acoustic-phonetic similarity of native and non-native vowels should closely match the perception of non-native sounds. Our results were not fully in line with L2LP as the predictions based on acoustic comparisons were not successful in predicting perception patterns for all SG vowels. Elvin et al. (2021) also observed some discrepancies between our L2LP acoustic predictions and non-native categorization patterns. A possibility for the observed inconsistencies between acoustic and perceptual assimilations is that the present acoustic comparisons are based on static spectral information from the middle point of the vowels. Future studies may consider inclusion of dynamical aspects of vowels like formant trajectories in the LDAs for more accurate comparisons with perceptual assimilations. It is also worth mentioning that, since L2LP predictions cover L2 learners with different proficiency levels, a desirable way to later test this model's predictions would be a longitudinal study involving naïve participants who gradually gain L2 proficiency over a period. In sum, the perceptual similarity patterns and the acoustic comparisons reported here provide a baseline for future studies of Standard German vowel discrimination by Modern Greek learners.

ABOUT THE AUTHORS

Payam Ghaffarvand Mokari is currently a post-doctoral fellow in the School of Humanities at University of Eastern Finland. His broad area of research interests are phonetics and second-language speech acquisition. He is currently involved in research projects, including studies on linguistic and acoustic correlates of L2 speech comprehensibility, acoustic parameters in the classification of fricative sounds, and perceptuomotor compatibility effects in vowels.
e-mail: payam.ghaffarvand.mokari@uef.fi

Maria Lialiou currently is a Ph.D. student in Linguistics at Cologne University. This project was carried out during her M.Sc. studies in experimental linguistics studies at Potsdam University where she also worked as a research assistant at the Laboratory of Phonology and Phonetics. Her research interests revolve around intonation, prosody, experimental phonology, and speech processing. Currently, she is working on the differential processing of pitch rises and falls as well as of the position of a pitch rise in relation to prosodic structure in prominence marking and attention orienting.

e-mail: mlialiou@uni-koeln.de

REFERENCES

Alispahic, S., Mulak, K. E., & Escudero, P. (2017). Acoustic properties predict perception of unfamiliar Dutch vowels by adult Australian English and Peruvian Spanish listeners. *Frontiers in Psychology, 8*, 52.

Best, C. T. (1995). A direct realist perspective on cross–language speech perception. In W. Strange (eds). *Speech Perception and Linguistic Experience: Issues in Cross–Language Research*. Timonium, MD: York Press.

Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.0.54. Retrieved from http://www.praat.org/.

Elvin, J. (2016). *The role of the native language in non-native perception and spoken word recognition: English vs. Spanish learners of Portuguese*, PhD Thesis, Western Sydney University, Australia.

Elvin, J., Escudero, P., & Vasiliev, P. (2014). Spanish is better than English for discriminating Portuguese vowels: acoustic similarity versus vowel inventory size. *Frontiers in Psychology, 5*, 1188.

Elvin, J., Williams, D., Shaw, J. A., Best, C. T., & Escudero, P. (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*(1), 44.

Escudero, P. (2005). *Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization*, PhD thesis, LOT Dissertation Series 113. Utrecht University, Netherlands..

Escudero, P., & Chládková, K. (2010). Spanish listeners' perception of American and Southern British English vowels. *The Journal of the Acoustical Society of America, 128*, EL254–EL259.

Escudero, P., & Vasiliev, P. (2011). Cross-language acoustic similarity predicts perceptual assimilation of Canadian English and Canadian French vowels. *The Journal of the Acoustical Society of America, 120*, EL277–EL283.

Escudero, P., Sisinni, B, & Grimaldi. M. (2014). The effect of vowel inventory and acoustic properties in Salento Italian learners of Southern British English vowels. *The Journal of the Acoustical Society of America, 135*, 1577–84.

Gilichinskaya, Y. D., & Strange, W. (2010). Perceptual assimilation of American English vowels by inexperienced Russian listeners. *The Journal of the Acoustical Society of America, 128*, EL80–EL85.

Harnsberger, J. D. (2001). On the relationship between identification and discrimination of non-native nasal consonants. *The Journal of the Acoustical Society of America*, *110*(1), 489–503.

Nishi, K., Strange, W., Akahane–Yamada, R., Kubo, R., & Trent–Brown, S. A. (2008). Acoustic and perceptual similarity of Japanese and American English vowels. *The Journal of the Acoustical Society of America*, *124*(1), 576–588.

Strange, W., Bohn, O. S., Trent, S. A., & Nishi, K. (2004). Acoustic and perceptual similarity of North German and American English vowels. *The Journal of the Acoustical Society of America*, *115*(4), 1791–1807.

Tyler, M. D. (2021). Phonetic and phonological influences on the discrimination of nonnative phones. In R. Wayland (Ed). *Second language speech learning: Theoretical and empirical progress* (pp. 157–174). Cambridge, UK: Cambridge University Press.

Williams, D., & Escudero, P. (2014). Influences of listeners' native and other dialects on cross–language vowel perception. *Frontiers in Psychology*, *5*, 1065.