This is the pre-peer reviewed version of the following article: Kapnoula, E. C., & Samuel, A. G. (2024). Sensitivity to Subphonemic Differences in First Language Predicts Vocabulary Size in a Foreign Language. *Language Learning*, which has been published in final form at https://onlinelibrary.wiley.com/doi/full/10.1111/lang.12650.

Manuscript Running Header: L1 Speech Gradiency Helps Foreign Language Learning

Article Type: EMPIRICAL STUDY

Manuscript Title: Sensitivity to L1 Subphonemic Differences Predicts Vocabulary Size in a Foreign Language

Authors: Efthymia C. Kapnoula[a,b], Arthur G. Samuel[a,c,b]

Author Affiliations: [a]Basque Center on Cognition, Brain and Language, [b]Ikerbasque, [c]Stony Brook University

Author notes / acknowledgements

The authors thank Alexandra Kypta-Vivanco for help with data collection and Candice Frances for providing the script for the Corsi task. Support for this project was provided by the Spanish Ministry of Science and Innovation through Grant # PID2020-113348GB-I00, awarded to AGS and ECK. This work was supported by the Basque Government through the BERC 2018-2021 and BERC 2022-2025 programs, and by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation SEV-2015-0490 and CEX2020-001010-S. This project has received funding from the European Union's Horizon 2020 research and innovation program, under the Marie Skłodowska-Curie grant agreement No 793919, awarded to ECK.

Correspondence concerning this article should be addressed to:

Efthymia C. Kapnoula
Paseo Mikeletegi 69
Basque Center on Cognition, Brain and Language (BCBL)
20009 San Sebastián – Donostia, Gipuzkoa
Spain
kapnoula@gmail.com

Abstract

Some listeners exhibit higher sensitivity to subphonemic acoustic differences (i.e., higher speech gradiency). Here, we asked whether higher L1 gradiency facilitates foreign language learning and explored the possible sources of individual differences in L1 gradiency. To address these questions, we tested 164 native Spanish speakers with different linguistic profiles. Speech gradiency was assessed via a Visual Analogue Scale task and foreign language proficiency was assessed via an English vocabulary test. Possible sources of gradiency included domain-general auditory acuity, overall exposure to spoken language (indexed by age), and exposure to phonological diversity. Control measures were collected to account for factors such as phoneme categorization consistency, working memory, and musical training. The results revealed a positive link between L1 speech gradiency and vocabulary acquisition in a foreign language over and above all other factors. L1 speech gradiency itself was predicted by domain-general auditory acuity and overall exposure to spoken language.

Keywords: speech perception, individual differences, foreign language learning, auditory processing, second language learning

Introduction

Becoming proficient in a new language requires learning its phonology. This task is not always easy; individuals often struggle to discriminate non-native speech sounds, even after years of experience with a language (Sebastián-Gallés et al., 2005; Sebastián-Gallés & Baus, 2005). Mastering an unfamiliar phonological system can be challenging because of the perceptual tuning to one's native (L1) phonology. As early as 6-8 months of age, infants learn to discriminate between different phonemes of their native language (between-category differences) and ignore differences within phoneme categories (Kuhl et al., 1992). This tuning reflects a perceptual distortion (i.e., perceptual warping) of speech segments to better match L1 categories. Focusing on meaningful phonetic differences is beneficial for learning one's native language, but it may hinder the learning of other languages that use subphonemic differences that one learned to ignore (Best & Tyler, 2007; Kuhl et al., 2008).

Even though speech perception is shaped by one's native phonology, there is evidence that listeners are sensitive to within-category differences (Andruski et al., 1994; McMurray et al., 2002; Miller, 1997; Samuel, 1977, 1982; Sarrett et al., 2020; Toscano et al., 2010). That is, tuning to L1 phonology creates some category-driven perceptual warping, but listeners still track acoustic differences within a category – this sensitivity to within-category differences is what we refer to as *gradiency*. Critically for our study, the degree of L1-driven perceptual warping seems to be idiosyncratic to each listener (Kapnoula et al., 2017; Kapnoula & McMurray, 2021), which may affect learning a new phonology and, eventually, a new language. As perceptual warping may impede the perception of subphonemic differences, listeners with more warping may experience greater difficulty learning non-native phonemic contrasts that rely on these subphonemic differences. Here, we tested whether gradiency in L1 speech perception predicts an individual's success in learning a foreign language and explored several possible sources of individual differences in gradiency.

The role of L1 in non-native speech perception

It is widely agreed that the perception of non-native speech sounds is affected by one's native phonology. According to the Native Language Magnet Model (Kuhl, 1993; Kuhl et al., 2008), L1 phonological categories act as "magnets": unfamiliar speech sounds are perceptually "drawn to", and eventually categorized as, similar-sounding L1 categories. Similarly, in the Perceptual Assimilation Model (Best et al., 2001; Best & Tyler, 2007), non-native speech sounds are perceptually assimilated to similar-sounding L1 categories, with consequences for perceiving and acquiring non-native contrasts. That is, if two non-native speech sounds assimilate to the same L1 category, discriminating between them should be difficult. In contrast, if each new sound assimilates to a different L1 category, discrimination should be easier. Lastly, according to the Speech Learning Model (Flege, 1995), the perception of non-native speech sounds depends on their similarity to native categories; new categories are established only when new sounds clearly differ from familiar categories.

The common idea among these models is that speech input is often perceptually distorted in the sense that speech segments are warped to map onto existing L1 categories. This warping is thought to facilitate recognition of L1 categories. Support for this view comes mainly from work on individuals with dyslexia, who appear to have higher sensitivity to subphonemic (allophonic) differences (Serniclaes et al., 2001, 2004). However, warping may also conceal subphonemic differences that are meaningful in a foreign language. Given that L1-driven warping may impede perceiving and consequently learning non-native speech contrasts, stronger warping could lead to greater difficulty in learning an L2.

Individual differences in gradiency and their role in L1 and L2 speech perception

It is broadly accepted that listeners differ in how they process L1 speech, including how sensitive they are to subphonemic information (for a review, see Yu & Zellou, 2019). Examining how these differences are linked to other linguistic and non-linguistic processes can shed light on the functional role of gradiency[1] in learning and processing spoken language.

One of the first studies to find robust individual differences in listeners' speech gradiency was by Kong and Edwards (2016). They used a visual analogue scaling (VAS) task, which allows listeners to rate speech sounds on a continuous scale (see also Massaro & Cohen, 1983, and Munson et al., 2010). VAS participants hear speech sounds from a continuum (e.g., *ba* to *pa*) and click on a line to rate each sound (e.g., how *ba*-like versus *pa*-like it is). Kong and Edwards found that listeners vary in how they perform the VAS; some used the entire scale for their ratings, whereas others used mainly the endpoints, following a step-like response pattern. This demonstration of individual differences in L1 speech perception provides a foundation to explore the source of these differences and their consequences for comprehension and L2 learning.

Kapnoula et al. (2017) used a VAS task to ask how gradiency relates to other aspects of speech perception and to non-linguistic cognitive processes. Their results revealed a positive relationship between gradiency and secondary cue use (see also Kim et al., 2020; Kong & Edwards, 2016). Specifically, individuals with higher gradiency exhibited greater use of a secondary voicing cue, F_0, when rating stimuli in terms of voicing (e.g., from /b/ to /p/). Even though this correlation does not reveal the causal nature of the relationship between gradiency and secondary cue use, it suggests a functional role of gradiency in speech perception. This interpretation was supported by a follow-up study; Kapnoula et al. (2021) measured listeners' gradiency, along with their ability to recover from misinterpretations of the speech input (lexical garden-paths) using a task modeled after McMurray et al. (2009).. Higher gradiency predicted higher likelihood of recovering from errors, especially when there was high acoustic discrepancy between the stimulus and the intended target word (i.e., when mispronunciations were larger). This pattern suggests that gradient information about the speech signal can be useful in L1 speech processing.

Having established a functional consequence of gradiency, Kapnoula and McMurray (2021) asked whether differences in gradiency stem from differences in acoustic cue encoding. They focused on voice onset time (VOT), the delay between the release of the articulators and the onset of voicing. VOT is the primary cue distinguishing voiced and unvoiced sounds (e.g., /b/ and /p/) in many languages, including English and Spanish. To quantify differences in VOT encoding, they used an ERP paradigm introduced by Toscano et al. (2010). In this paradigm, participants listen to stimuli along a range of VOT values (e.g., *beach*-to-*peach*). The critical finding in this paradigm is that VOT is linearly reflected in the amplitude of the auditory N1 (a negative ERP component that appears ~100 ms post stimulus onset; see Sharma et al., 2000; Sharma & Dorman, 1999): lower VOT values trigger larger N1 responses (see also Sarrett et al., 2020; see Getz & Toscano, 2021, for a review). Kapnoula and McMurray (2021) found that individual differences in speech gradiency are linked to differences in acoustic cue encoding. Specifically, for participants with lower gradiency (i.e., stronger category-driven perceptual warping), the link between VOT and N1 was best explained by a model combining both a linear *and* a step-function. Given that the step-function was centered at each listener's category boundary, this finding points to robust category-driven warping for those listeners. In contrast, for high-gradiency listeners, adding a step-function did not improve the model fit, suggesting a minimal effect of categorical information in cue encoding. Thus, this study showed that the degree to which L1 categories affect speech perception varies across listeners and that these differences depend on the early encoding of acoustic cues.

The findings by Kapnoula and McMurray (2021) identify the early encoding of acoustic cues as a source of gradiency, with some listeners experiencing stronger L1-driven perceptual warping around the category boundary than others. Expanding on this idea, differences in perceptual warping may not only affect L1 speech perception, but also perception of non-native contrasts. Specifically, higher gradiency (i.e., weaker category-driven perceptual warping) in L1 may allow listeners to better encode meaningful subphonemic differences in a new language. Thus, *gradiency in L1 speech perception may facilitate learning a new phonology, and consequently a foreign language*. This was the first hypothesis we aimed to test.

Sources of individual differences in gradiency

Although individual differences in L1 speech perception are well documented, it is unclear how these differences come to be. Here, we explored three possible factors: (a) *domain-general auditory acuity*, (b) *overall experience with language*, and (c) *exposure to high phonological diversity*.

The idea that domain-general auditory acuity is important in speech perception is not only intuitive, it also has empirical support. Studies have shown that sensitivity to fine-grained acoustic information that is relevant for L1 speech processing is developed in early infancy (Cutler & Butterfield, 1992; Joanisse & Seidenberg, 1998; Kuhl, 2000). Importantly for our focus, there are substantial individual differences in auditory processing acuity, and these differences have been linked to L1 learning difficulty (Goswami et al., 2011), and to listeners' ability to learn a foreign language (Kachlicka et al., 2019; Kempe et al., 2014; Saito et al., 2022).

Thus, it would not be surprising if speech perception gradiency is linked to general auditory processing acuity.

Another possible predictor of speech gradiency is overall experience with spoken language. As mentioned above, sensitivity to acoustic information that is relevant for L1 speech processing develops in early infancy. However, there is also evidence that this kind of sensitivity is still being fine-tuned years later. Taking the example of VOT, adult listeners can track acoustic differences as small as 5ms when recognizing spoken words (McMurray et al., 2002). Crucially, sensitivity to VOT differences changes through childhood and into late adolescence. McMurray et al. (2018) tested children from three age groups (7-8, 12-13, and 17-18 years) using the Visual World Paradigm (VWP; Allopenna et al., 1998). Participants heard speech continua corresponding to minimal pairs (e.g., *beach*/*peach*), while seeing pictures of those items, and were asked to click on the correct referent. Explicit responses were used to assess category sharpness, whereas eye-movements to the non-clicked picture (i.e., competitor) were used as an index of gradiency. The results showed that gradiency slowly increased over development. At the same time, sharpening of speech categories was seen during the same period, suggesting that increasing sensitivity to acoustic detail (higher gradiency) is linked to better (more discrete) categorization. It is difficult to specify the exact nature of this link; higher gradiency could lead to higher category sharpness, the two could be driven by a third factor, or there could be parallel maturation of two independent processes. The important point here is that speech gradiency continues to increase during adulthood as a natural consequence of higher exposure to spoken language.

A third potential contributing factor to gradiency is systematic exposure to different distributions of speech sounds over long time periods. This idea is consistent with the Speech

Learning Model (SLM-r; Flege & Bohn, 2021), according to which the learning of phonetic categories is based on the distributions of speech sounds listeners are exposed to across their lifespan. This idea also aligns with research on second language acquisition and bilingualism that highlights the role of exposure to different phonologies. For example, Ramon-Casas et al. (2009) showed that Catalan-Spanish infants' ability to discriminate Catalan vowel contrasts was linked to the ratio of Catalan–to–Spanish exposure they received at home. In sum, linguistic exposure seems to shape the way listeners process speech in a fundamental way (see also Samuel & Larraza, 2015; Sebastián-Gallés et al., 2005; Sebastián-Gallés & Bosch, 2009). Following this rationale, here we tested the hypothesis that exposure to broad distributions of speech sounds (i.e., diverse phonologies) leads to higher gradiency.

Present study

Our first goal was to test the idea that *higher gradiency in L1 speech perception facilitates the learning of non-native phonemic contrasts and ultimately foreign language learning*. Specifically, we were interested in the accumulated, long-term effects of L1 gradiency on an individual's success in learning a foreign language. We operationalized this as the participants' English proficiency at the time of testing, quantified in terms of vocabulary size.

Our second goal was to explore three possible predictors of gradiency. The first factor was domain-general auditory acuity. Our measure of speech gradiency was based on a temporally-defined contrast (VOT); thus, we focused on temporal rather than spectral acuity. Specifically, we measured each listener's threshold when judging the temporal order of two tones presented in rapid succession. Our prediction was that *better auditory acuity would predict higher speech gradiency*. Second, previous work shows that the fine-tuning of gradiency continues through late adolescence (McMurray et al., 2018). Building on this work, we hypothesized that accumulated exposure to spoken language (indexed by listeners' age) may predict gradiency. Therefore, we expected that *relatively older adults would exhibit higher speech gradiency compared to younger adults* (because our participants were between 18-40, even our oldest participants were much younger than people often referred to as "older adults"). Lastly, we examined the link between speech gradiency and exposure to diverse phonologies, while controlling for the effect of exposure to multiple languages that are not phonologically diverse. We took advantage of the unique phonological landscape of the Basque Country, where inhabitants are exposed to Spanish, Basque, and English. Importantly, Spanish and Basque have very similar phonological systems (Hualde, 2015), whereas Spanish and English are phonologically quite distinct (Velázquez López, 2015). Thus, the diverse-phonologies prediction was that *native Spanish speakers with higher exposure to English (i.e., highly dissimilar phonology) should exhibit higher speech gradiency in their L1*, whereas exposure to Basque should not have the same effect.

To complement our primary questions, we collected additional measures to validate our measure of gradiency, as well as measures of participants' linguistic history and other characteristics of interest such as working memory and musical education.

Method

Participants

One hundred sixty-four (114 females; mean age = 25.6 years) native Spanish speakers participated in this study. Because participants resided in the Basque Country, they were also familiar with Basque, which was taken into account in the experimental design, participant recruitment, and data analyses. All but four participants reported Spanish as their dominant language – one did not report dominance and three reported Basque as their dominant language.

Despite not self-identifying as Spanish-dominant, none of these four participants reported learning another language before Spanish and all achieved 100% in our primary measure of Spanish proficiency. These participants were thus included in the analyses. Participants' characteristics are reported in Table 1.

Table 1

Participant characteristics

	Source	Min	Max	Mean	SD
Age	BCBL database	18	40	25.6	5.9
AoA Basque	BCBL database	0	29	4.6	6.1
AoA English	BCBL database	2	14	6.3	2.5
Auditory exposure to English	Questionnaire	0%	97%	42%	25%
English picture naming [BEST]	BCBL database	23%	100%	70%	18%
Musical training (in # of months)	Questionnaire	0	276	28	49

All participants self-reported normal/corrected-to-normal vision, no known hearing or neurological impairments, and all successfully completed a hearing screening at four octave-spaced frequencies for each ear. Participants underwent informed consent and were remunerated for their participation. All experimental procedures were approved by the BCBL ethics committee.

Overview of hypotheses and tasks

Hypothesis 1*: *Higher subphonemic sensitivity facilitates foreign language learning. To test this hypothesis, we examined the link between gradiency and proficiency in a foreign language (English) taking into account several control variables. Gradiency was assessed via a VAS task (see below). English proficiency was indexed by participants' performance on a vocabulary test routinely administered at the BCBL for this purpose (see below). In examining this link, one can use different measures of L2 learning such as second language speech perception, or conduct a training experiment. The goal of our study, however, was to assess longer-term and/or broader effects of L1 gradiency on L2 language learning. Hence, we use L2 vocabulary as a more generic (but also objective) measure of L2 proficiency.

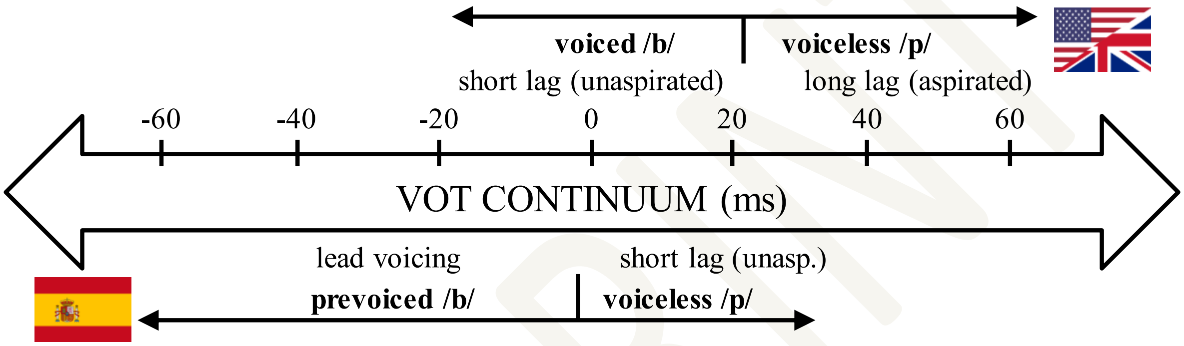
Hypothesis 2a: Auditory acuity predicts higher gradiency. To test this, we measured participants' temporal auditory acuity, which we linked to their gradiency score. We used a staircase procedure to assess listeners' perceptual thresholds when judging the temporal order of brief tones (see below).

Hypothesis 2b: Overall exposure to spoken language leads to higher gradiency. Participants' age at the time of testing was used as an index of their overall exposure to spoken language. We acknowledge that age encompasses a wide variety of life experiences, not all of which are related to language exposure. We return to this point in the Discussion.

Hypothesis 2c: ***Phonological diversity leads to higher gradiency.*** We tested this hypothesis while controlling for overall exposure to different languages by testing trilinguals with two languages (Spanish-Basque) that are phonologically similar to each other and a third language (English) that is phonologically different from the first two. We extracted a gradiency measure using the contrast between voiced and unvoiced bilabial stops (/b/ and /p/). These two sounds have indistinguishable VOT values across Spanish and Basque (/b/ ≈ -65 ms; /p/ ≈ 20 ms; Souganidis, 2023), whereas in English, both stops differ significantly from the Spanish/Basque values (/b/ ≈ 0 ms; /p/ ≈ 60 ms; Lisker & Abramson, 1964; see Fig. 1). The diverse-phonologies hypothesis predicts that higher exposure to English VOTs would lead to higher sensitivity to within-category differences in the perception of Spanish (L1) bilabial stops, whereas differences in exposure to Basque should not have the same effect. Information about auditory exposure to English and Basque was collected via a questionnaire (see **Questionnaire**).

Figure 1

The relationship between English and Spanish bilabial stops (adapted from Llama et al., 2010)



Secondary/control measures. In addition to the main measures listed above, we collected several secondary and control measures. First, we used VAS responses to extract a measure of response consistency (i.e., the degree to which listeners give similar ratings to the same stimuli). This allowed us to validate our measure of gradiency by ensuring it does in fact reflect subphonemic sensitivity rather than response inconsistency. Previous work has shown that response consistency is largely independent of gradiency (Kapnoula et al., 2017), however the measures could in principle be related: A shallow slope could also reflect highly inconsistent VAS ratings (for a review of measures that can be extracted from the VAS task, see (Apfelbaum et al., 2022). We disentangled the effects of subphonemic sensitivity and response consistency a) by making sure the two measures are not correlated and b) by accounting for the variance in response consistency before inserting subphonemic sensitivity in the regression analysis. Lastly, Fuhrmeister et al. (2023) showed that higher consistency predicts better learning of non-native contrasts which we directly tested by including consistency as a predictor in the regression.

Second, we extracted a measure of listeners' use of F_0 as a secondary voicing cue. In English there is a well-documented relationship between gradiency and F_0 use (Kapnoula et al., 2017; Kim et al., 2020; Kong & Edwards, 2016). In Spanish, F_0 co-varies with VOT (Dmitrieva et al., 2015), but it does not seem to play an equally important role in listeners' judgements of voicing (Llanos et al., 2013). Therefore, it was unclear whether we would observe this link here.

Third, participants performed two additional VAS tasks – one visual and one non-speech auditory VAS task. These control tasks allow us to measure the degree to which differences in the Speech-VAS task are due to response-style differences in using a visual scale and account for that, if necessary.

Fourth, there is some evidence that higher working memory may be linked to gradient VAS responses, possibly by allowing listeners to maintain gradient activation of speech sounds long enough to affect responses (Kapnoula et al., 2017). However, the task used in that work was partly linguistic (NBack), making it unclear whether the (weak) effect was driven by this linguistic component, or modality-independent working memory. We therefore used the visual Corsi task (Berch et al., 1998; Corsi, 1972) here to assess participants' working memory.

Lastly, it has been suggested that musical training may be linked to speech processing (e.g., Kempe et al., 2014). We therefore collected a measure of musical experience (see Appendix S4).

Participants came to the lab for a 1-hr session that included a hearing screening followed by the experimental tasks. The order of administration and primary purpose of each task are presented in Table 2.

Table 2

Order and description of tasks

Order	Task	Measure(s)	Duration (mins)
1	Speech-VAS	a. Phoneme categorization gradiency b. Response consistency c. Secondary cue (F_0) use	10
2	2AFC	a. Response consistency b. Secondary cue (F_0) use	6
3	Visual-VAS	Visual categorization gradiency	10
4	Audio-VAS	Auditory categorization gradiency	4
5	Corsi	Working memory	3
6	TOJ	Auditory acuity/ Temporal integration threshold	4
7	Questionnaire	Linguistic history, linguistic exposure, and musical training	5

Visual analogue scaling tasks (VAS): Gradiency, response consistency, and F_0 use

Speech-VAS. A VAS task with speech stimuli assessed participants' gradiency, response consistency, and secondary cue use. Participants rated tokens from a *ba*-to-*pa* continuum using a continuous scale. Stimuli were based on natural speech recordings spoken by a female native Spanish speaker, which were used to construct five 9-step VOT continua that differed in their F_0 values.

We first extracted the pitch contour of the voiced endpoint (*ba*) and used that to construct two contours of identical shape that were shifted downwards and upwards so that the mean onset pitch would be 179 Hz and 193 Hz[2] respectively. These two contours were used as endpoints to create three intermediate F_0 steps, resulting in five steps in total. The resulting pitch contours were approximately 3.34 Hz apart. We then replaced the original contours of /ba/ and /pa/ using the pitch-synchronous overlap-add (PSOLA) algorithm in Praat (Boersma & Weenink, 2016).

Next, we constructed five voicing continua using the progressive cross-splicing method described by Andruski et al. (1994) and McMurray et al. (2008), adapted to incorporate prevoicing. For the prevoiced steps we started at -55 ms VOT and cut progressively longer portions of prevoicing until reaching -5 ms VOT. Then, we stripped all prevoicing from the voiced sound and replaced progressively longer portions of its onset with equivalent durations taken from the short-lag sound, up to 25 ms VOT. VOT steps varied in principle from -55 ms to 25[3] ms and were 10 ms apart; however, small adjustments were made to respect zero crossings (actual VOT values are presented in Table 3). Each stimulus was presented three times, resulting in 135 trials (9 VOTs × 5 F_0s × 3 reps).

Table 3

VOT values of stimuli used in the Speech-VAS task

Step	Original VOT values	VOT value adjusted for zero-crossings
1	-55	-58
2	-45	-44
3	-35	-37
4	-25	-24
5	-15	-17
6	-5	-5
7	5	5
8	15	15
9	25	25

Visual-VAS. Visual gradiency was assessed with a two-dimensional *apple*-to-*pear* continuum spanning color and shape in a 7×5 matrix using the stimuli from Kapnoula et al., (2021). Two pictures (an apple and a pear) were edited to intensify prototypical characteristics and manipulated using Fantamorph (ver. 5) software to create a 7-step shape continuum. These were recolored in a five step continuum from yellow-ish (prototypical *pear*) to red (prototypical *apple*). Each picture was presented five times, resulting in 175 trials (7 shapes × 5 colors × 5 repetitions).

Audio-VAS. Non-linguistic auditory gradiency was assessed using a *trumpet*-to-*piano* continuum. Stimuli were a subset of those used in Pitt (1995). In the original study, an 11-step trumpet-piano continuum was created by digitally mixing a trumpet and a piano tone in different proportions. For this study, we used the six odd-numbered continuum steps. Each step was presented eight times, resulting in 48 trials. In contrast to the other VAS tasks, there was only one dimension. However, the rotated logistic (described below) needs two-dimensional data. We therefore entered the data in the form of a 6×3 structure, in which the values of the first dimension were repeated for each of the values of the second dimension.

Procedure. Participants saw a line with a label/picture at each endpoint (see Appendix S1). They listened to or saw each stimulus and clicked on the line to indicate the corresponding position of the stimulus. When they clicked, a rectangular bar appeared at that location; they could change their response (by clicking elsewhere) or press the space bar to verify it.

Quantifying gradiency, response consistency, and secondary cue use. The VAS task allows us to extract multiple measures, each tapping different aspects of speech perception and categorization processes. Here, we used it to quantify each listener's gradiency, response consistency, and secondary cue use.

As in previous work (Kapnoula, 2016; Kapnoula et al., 2017, 2021; Kapnoula & McMurray, 2021), we used the rotated logistic function (Eq. 1) to fit participants' VAS responses in all VAS tasks. This function is particularly useful for multidimensional categorization as it provides orthogonal measures of gradiency and secondary cue use.

$$p(resp) = b_1 + \frac{(b_2 - b_1)}{1 + e^{\left(\frac{-4 \cdot s \cdot 2 \cdot \upsilon(\theta)}{(b_2 - b_1)}\right) \cdot \left(\frac{\tan(\theta) \cdot (x_0 - VOT) - F_0)}{\sqrt{1 + \tan(\theta)^2}}\right)}} \tag{1}$$

Here, *b₁* and *b₂* are the lower and upper asymptotes. For the category boundary, the rotated logistic assumes a diagonal boundary in a two-dimensional space that is described as a line with a crossover point along the primary cue and an angle, θ; a θ of 90° indicates exclusive use of the primary cue (the x axis) and a θ of 45° reflects relatively equal use of both cues. Once the boundary vector is identified, this equation rotates the coordinate space to be orthogonal to this boundary –the *tan*(θ) term– and the slope (*s*) of the function is thus perpendicular to the diagonal boundary. Lastly, υ(θ) switches the direction of the function, if θ is less than 90, to keep the function continuous. Unlike the standard logistic, this function allows for asymptotes other than 0/1, it avoids conflating the boundary along each dimension and the slope, and, finally, it allows a single estimate of slope that pools across both dimensions.

We used this function to quantify: 1) *phoneme categorization gradiency*, reflected by the slope parameter, with steeper slopes indicating more step-like responses, and 2) *secondary cue use,* reflected by the θ angle. The equation was fit to each participant's VAS responses using a constrained gradient descent method implemented in Matlab that minimized the least squared error (software available at McMurray, 2017). Fits were good (R^2 = .98, R^2 = .97, and R^2 = .74 for the Speech-, Visual, and Audio-VAS task respectively[4]).

Kapnoula et al. (2021) showed that individual differences in speech gradiency are due to differences in the degree of category-driven perceptual warping around the boundary. This suggests that individual differences in gradiency are likely to be more prominent close to the boundary, i.e., listeners should differ mainly in how they perceive *ambiguous* speech stimuli: Gradient listeners should be able to maintain the ambiguity, giving ratings close to the scale midpoint, while categorical listeners should be more affected by category-driven perceptual warping, giving ratings far from the midpoint (i.e., stronger category-driven warping should pull ratings towards the endpoints). We therefore computed a measure of the strength of this warping specifically for the most ambiguous VOT for a given participant (i.e., closest to their b/p boundary). The location of each participant's VOT boundary was determined based on their VAS responses; it was the VOT step for which the equation-predicted response was closest to the midpoint of the scale. Using only trials in which the VOT of the stimulus corresponded to the participant's boundary, we quantified warping as the absolute distance between each VAS rating (ranging from 0 to 100) and the scale midpoint (50). We call this measure *boundary warping*, with higher scores reflecting higher perceptual warping (i.e., lower gradiency).

Finally, we used the VAS ratings to extract a measure of *response consistency* using the same procedure as in Kapnoula et al. (2017). We first computed the difference between each VAS rating and the predicted value based on that participant's fitted curve and then computed the standard deviation of these residuals for each participant. The reverse of this measure (i.e., -1 × [raw_value]) was our main measure of response consistency.

Two-alternative forced choice task (2AFC): Response consistency and F_0 use

In addition to the response consistency measure extracted from the VAS ratings, we extracted a measure of response consistency from a 2AFC speech categorization task. Previous work in English has shown that (1) the steepness of the slope in this task reflects mainly response consistency and (2) this measure is not correlated with gradiency as measured by the VAS slope (Apfelbaum et al., 2022; Kapnoula et al., 2017). Thus, if Spanish-speaking listeners perform the Speech-VAS similarly to English-speaking listeners, the two slopes should be uncorrelated.

Participants performed the 2AFC task immediately after the Speech-VAS task. A subset of the VAS stimuli was used in the 2AFC task: all nine VOT steps, but only the two extreme F_0 values. Each of the 18 stimuli was presented 10 times (180 trials).

Procedure. On each trial participants saw a square on each side of the screen (see Appendix S2). Each square contained one of two printed syllables (*ba*/*pa*), with *ba* always appearing on the left. Participants were asked to click in the box with the syllable that best matched what they heard.

Data Pre-processing. To assess 2AFC categorization slopes, we fit each participant's response curve using a four parameter logistic function (see McMurray et al., 2010; see Eq. 2).

$$p(resp) = b_1 + \frac{b_2 - b_1}{1 + e^{\left(\frac{-4 \cdot s}{(b_2 - b_1)}(x - co)\right)}} \qquad (2)$$

In this equation, b_1 is the lower asymptote, b_2 is the upper asymptote, s is the slope, and co is the x-intercept. This function was fit to each participant's responses separately for each F_0 and the average slope across F_0 values was used in the analyses. In addition, the difference between the two intercepts provided a measure of F_0 use. Curves were fit using a constrained gradient descent method implemented with FMINCON in Matlab. Fits were good ($R^2 > .99$).

Corsi task: Working memory

The backwards version of the Corsi block-tapping task was used to measure working memory (Corsi, 1972). On each trial, participants saw nine blue squares randomly placed on the screen (see Appendix S3). A sequence of squares would turn green for one second, one at a time, and then the participant had to click on each of them in the reverse order. Participants started with two practice trials with a sequence of two squares in each trial to make sure they understood the task. After that, the number of squares that changed color in each trial increased by one on each round. Participants had to successfully complete at least one of the two trials at each difficulty level to continue to the next round, with a maximum of nine squares. The highest difficulty level in which a participant clicked on all squares accurately in at least one of the two trials was used as a measure of working memory capacity.

Temporal order judgment (TOJ) task: Auditory acuity/temporal integration threshold

This task provided a measure of participants' auditory acuity. We focused on temporal acuity because time is particularly relevant in the perception of voicing; accurate perception of voicing largely depends on distinguishing between VOT segments that differ by a few milliseconds. To extract a measure of temporal auditory acuity, we used the temporal order judgment task (TOJ) following the same procedure as Simon and Winkler (2018; who used a paradigm derived from Fink et al., 2005; Fostick & Babkoff, 2013).

Stimuli and procedure. Stimuli were pairs of 10-ms long 800-Hz pure tones delivered at 70 dB. Between the two tones there was a variable silent interval (inter-stimulus interval; ISI) between 5 and 150 ms. Each tone in a pair was delivered to a different ear. The participants' task was to press the left arrow on a standard keyboard if the first sound was played in their left ear and the right arrow if the first sound was played in their right ear. The inter-trial interval was between 600 and 900 ms (jittered). Each participant's temporal threshold was measured with a three-down one-up staircase procedure: The initial ISI was 150 ms and after three correct responses the ISI was shortened by a step size. After an incorrect response, the ISI increased by the same step size. The step size was initially 20 ms; it was halved after each downward step until reaching 5 ms. The task ended if the participant gave eight incorrect responses in total, or if they gave three incorrect responses in a row. A participant's threshold was the average ISI across the last six incorrect trials.

Questionnaire: Exposure to English and musical training

At the end of the session, participants filled out a short questionnaire to assess their auditory exposure to English in a variety of settings (e.g., watching movies/videos, listening to music, playing videogames; see Appendix S4). Exposure to English phonology was quantified as proportion of auditory exposure to English over exposure to all languages.

Participants also reported their music background, quantified as the number of months of musical training.

Information in the BCBL database: English proficiency and other data

Upon registration in the BCBL database, participants perform a series of tasks and fill out a set of questionnaires. Thus, in addition to the data collected during the experimental session, we had access to information in the database.

Participants' English proficiency is assessed via the *Basque, English, and Spanish Tests* (BEST) battery (de Bruin et al., 2017). In this battery, participants see 65 pictures of common entities (e.g., animals, tools, body parts, taken from the *MultiPic* database, Duñabeitia et al., 2017), and name each in the three languages. The number of correctly named pictures is used as the participant's proficiency score in the corresponding language. The database provided participants' age (used as an index of exposure to spoken language) and age of acquisition (AoA) for English and Basque. English AoA indexed exposure to diverse phonologies (together with the self-reported measure of auditory exposure to English collected from the questionnaire), and Basque AoA was a control measure.

Results

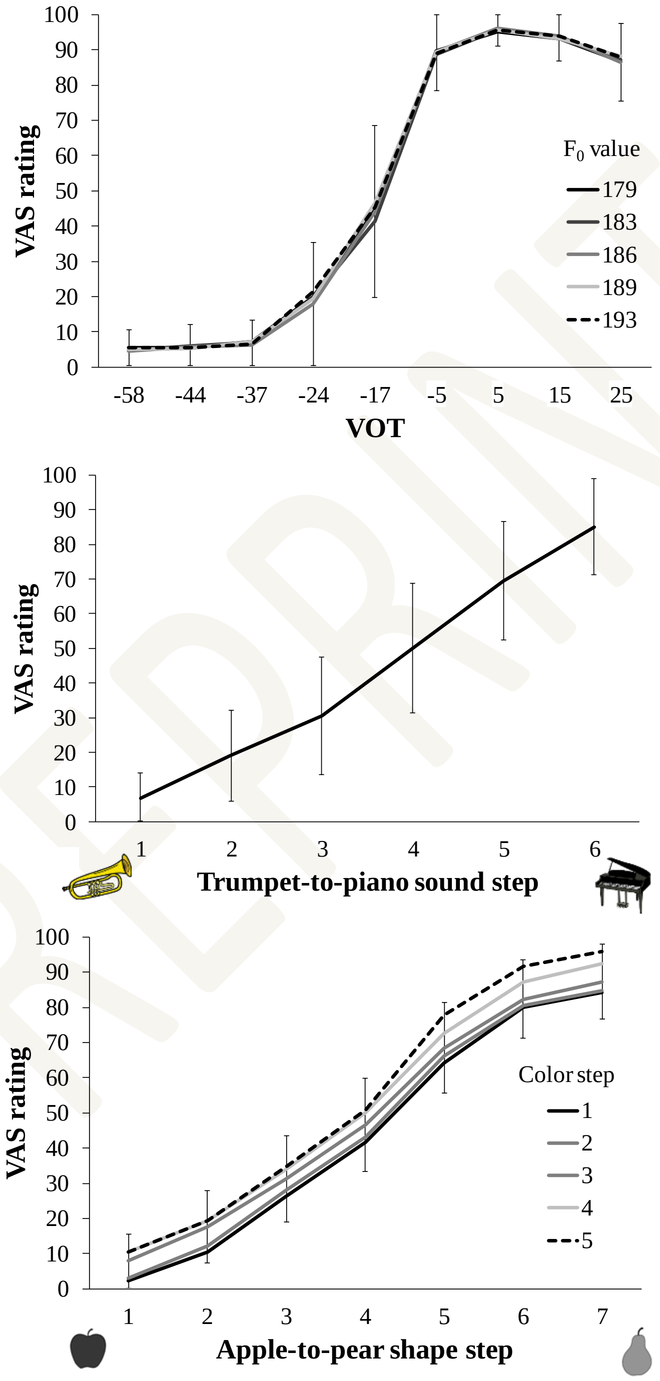
We start with a descriptive overview of the data, proceed to the validation/selection of our main gradiency measure, and finally present the results related to our theoretical questions.

Descriptive results

VAS and 2AFC tasks. Participants performed the VAS tasks as instructed. As expected, participants rated stimuli with higher VOT as more /p/-like (Fig. 2). In contrast to previous results from English (Kapnoula et al., 2017; Kim et al., 2020; Kong & Edwards, 2016), participants did not appear to use F_0 as a secondary cue for voicing judgments (see Llanos et al., 2013, for related Spanish results).

Figure 2

VAS ratings in the Speech-VAS (top), Audio-VAS (middle), and Visual-VAS (bottom)

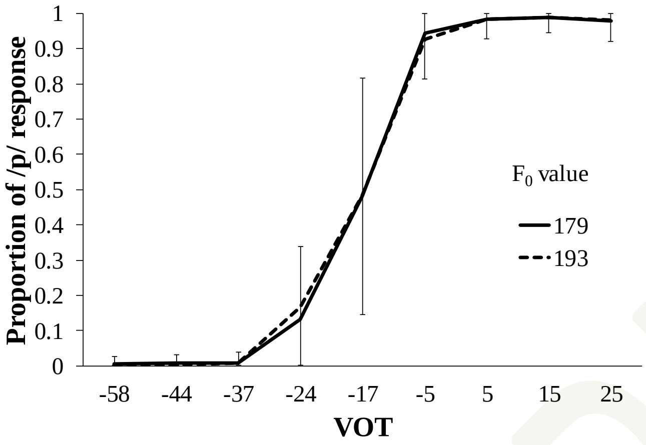


Note. Error bars show standard deviations (SDs). In the top and bottom plots, error bars correspond to the average SDs across secondary cue values and are centered at the middle value.

The 2AFC results also show that participants used only VOT (and not F_0) to categorize stimuli as /ba/ versus /pa/ (see Fig. 3).

Figure 3

Proportion of /p/ responses in the 2AFC task as a function of VOT



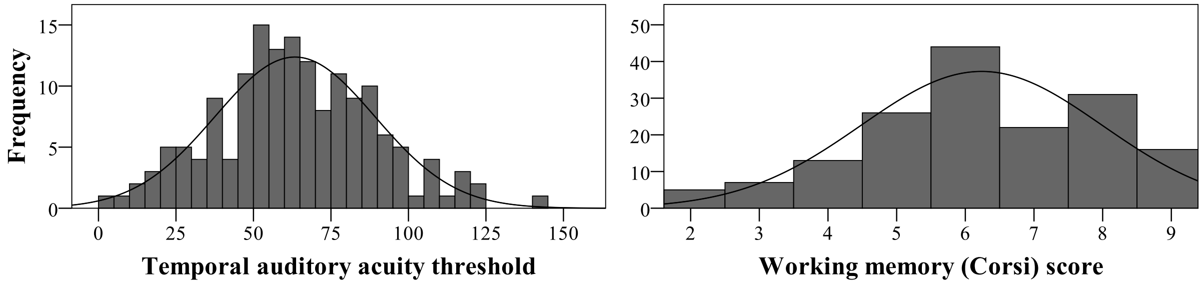
Note. Error bars show average SDs per VOT step and are centered at F_0 value 179.

We fitted participants' 2AFC responses using Eq.2 and extracted the function intercepts separately for each F_0. To statistically assess the F_0 effect we compared the difference between the two intercepts against zero using a one-sample t-test. The difference (.029) was not significantly different from zero, t(164)=.872, p=.385. Given the absence of an effect of F_0, we excluded this factor from further analyses.

Auditory acuity and working memory tasks. Participants performed the Auditory acuity and Working memory (Corsi) tasks without problems (see Fig.4), except four participants whose Auditory acuity data were excluded from the analyses due to a technical error.

Figure 4

Histograms of responses in Auditory acuity and Working memory (Corsi) tasks



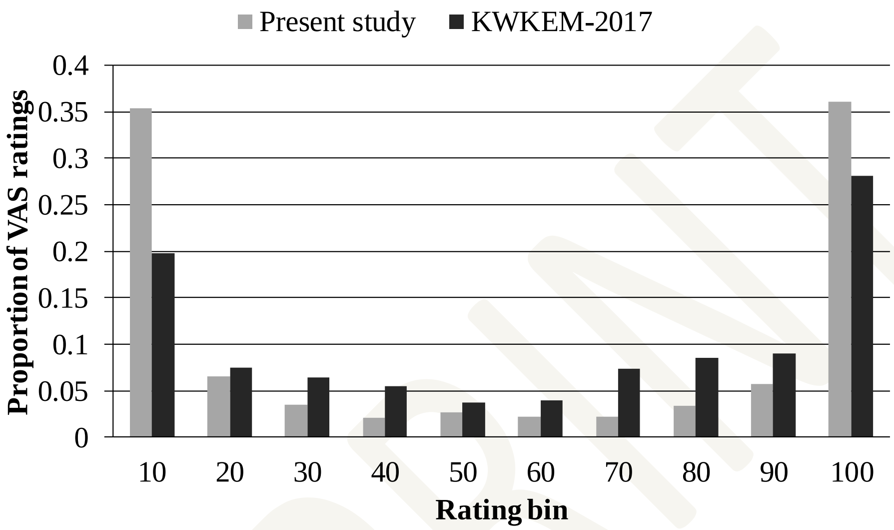
Validation of gradiency measures extracted from the VAS task

VAS slope. The VAS slope has been validated as a measure of speech gradiency in English (Kapnoula et al., 2017, 2021; Kapnoula & McMurray, 2021). However, because this is the first use of the VAS task with Spanish stimuli and Spanish listeners, we begin by examining the Spanish VAS data.

We started by directly comparing the Spanish VAS slopes from the rotated logistic function (Eq. 1) against English VAS slopes extracted using the same procedures in three previous experiments (Kapnoula et al., 2017, 2021; Kapnoula & McMurray, 2021). We conducted a one-way ANOVA with (log-transformed) VAS slope as the dependent measure and Experiment as the independent factor with four levels (Spanish; KWKEM-2017; KEM-2021; KM-2021). The average VAS slope values were lower in the KWKEM-2017 (1.75), KEM-2021 (1.75), and KM-2021 (1.78) studies compared to the current study (2.00), yielding a significant effect of Experiment, F(3,411) =12.160, p<.001. Bonferroni-corrected comparisons showed that the Spanish experiment VAS slopes were significantly higher than the VAS slopes in all three English experiments (all p<.001), whereas no significant differences were found among the English experiments (all p>.999). This pattern suggests that Spanish listeners performed the task differently than English listeners, with the former exhibiting a stronger preference towards the scale endpoints (see Fig. 5), leading to steeper VAS slopes.

Figure 5

Proportion of responses per rating bin in present study versus KWKEM-2017



Spanish-speaking participants might use the VAS endpoints more than English-speaking participants because the difference between Spanish /b/ and /p/ is more qualitative – it is based on the presence/absence of pre-voicing. In contrast, in English both endpoints typically have a positive VOT, making the corresponding contrast quantitative. Also, the VOT step size was larger here (~10ms) compared to the experiments in English (~7ms). Larger step sizes near the boundary make it more difficult to detect individual differences in gradiency. In sum, the Spanish participants may have treated the VAS task more like a 2AFC task.

To test this possibility, we correlated the speech VAS slopes with the 2AFC categorization slopes and the two non-linguistic (visual and auditory) VAS slopes. Previous work in English found that VAS and 2AFC slopes were not correlated (Kapnoula et al., 2017), suggesting that the two measures tap different processes (for a discussion on what the two measures may tap, see

Apfelbaum et al., 2022). Previous work has also shown that VAS slopes for speech stimuli were not correlated with VAS slopes for a visual continuum (Kapnoula, 2016).

As in the English studies, Spanish speech VAS slopes were not correlated with either the visual, r(159)=.081, p=.307, or the auditory, r(155)=.058, p=.474, VAS slopes. These results indicate that the VAS patterns are not simply a function of general response styles. However, in contrast to the English studies, the Spanish speech VAS slopes were significantly correlated with 2AFC slopes, r(159)=.434, p<.001, further confirming the divergence of the Spanish speech VAS pattern from the English pattern. This unexpected finding suggests that, unlike English natives, Spanish participants performed the VAS similarly to a 2AFC task. If so, the VAS slope may reflect response consistency rather than speech gradiency, leading to a correlation between VAS slope and our main measure of response consistency. Indeed, in contrast to Kapnoula et al. (2017), the two measures were significantly correlated, r(159)=.400, p<.001, suggesting that in the present study shallow VAS slopes reflect response inconsistency. This surprising result may reflect a discrepancy between Spanish and English native listeners either in terms of L1 gradiency, or in how they perform the VAS task. As mentioned above, this discrepancy could stem from differences in the specific contrast used (e.g., the qualitative difference between Spanish /b/ and /p/, based on the presence/absence of pre-voicing), or it could reflect differences in the VOT step size used.

Boundary warping. The boundary warping measure indexed the degree to which participants' ratings diverged from the scale midpoint when rating ambiguous stimuli. We again directly compared the Spanish results to the results of the three English experiments (Kapnoula et al., 2017, 2021; Kapnoula & McMurray, 2021). We did not expect to find a difference in warping between languages and indeed there was none; the lowest degree of warping was observed in KM-2021 (*M*=27), followed by KWKEM-2017 (*M*=28), the present study (*M*=29), and lastly KEM-2021 (*M*=33). In a one-way ANOVA with warping measure as the dependent measure and Experiment as the independent factor with four levels (Spanish; KWKEM-2017; KEM-2021; KM-2021), there was a significant effect of Experiment, F(3,416) =6.905, p<.001. Bonferroni-corrected comparisons revealed a significant difference between KM-2021 and KEM-2021 (p=.002), KM-2021 and Spanish (p=.020), and between KWKEM-2017 and KEM-2021 (p=.003). Thus, any differences between experiments were not driven by language differences. Importantly, unlike the VAS slope, boundary warping was not significantly correlated with response consistency, r(164)=.008, p=.917. Thus, for this dataset, boundary warping appears to be a better measure of gradiency than VAS slope. Therefore, we use boundary warping as our main measure of gradiency.

Does higher gradiency facilitate foreign language learning?

To address this question, we conducted a step-wise regression with English vocabulary as the dependent variable. In the first step, we accounted for the variance explained by a set of base variables that were not the focus of our study and were not likely to be affected by factors entered in subsequent steps: age, auditory acuity, working memory (Corsi) scores, and music training. As shown in Table 4, age and working memory were significant predictors of English vocabulary; older participants (or, as we interpret this, ones with longer language experience) and individuals with higher working memory had higher English vocabulary scores.

Table 4

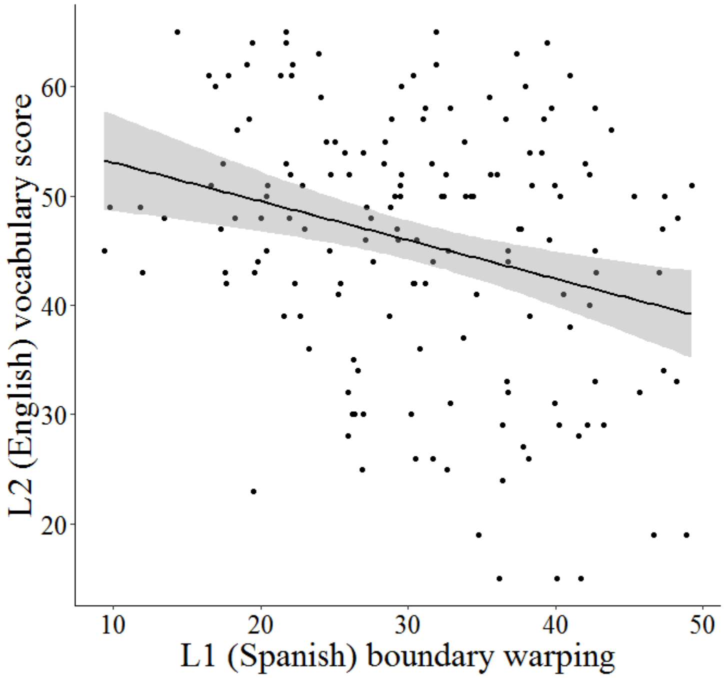
Hierarchical regression analysis predicting English vocabulary score

Step	Variable	*β*	ΔR^2	*p*
1	Age	.226		.005
	Auditory acuity	−.071		.389
	Working memory (Corsi)	.167		.037
	Music training	.043		.601
			.090	.007
2	AoA English	.082		.311
	AoA Basque	−.017		.824
	Exposure English	.469		< .001
			.223	< .001
3	VAS response consistency	−.034	.001	.702
4	Boundary warping	−.220	.042	.003

In the second step, we added a set of factors that were likely to be significant predictors, but were still not the focus of our study: AoA for English and Basque, and auditory exposure to English. As expected, auditory exposure to English was a significant predictor of English vocabulary. In the third step, we added VAS response consistency, which was our first predictor of interest. In contrast to previous work (Fuhrmeister et al., 2023), response consistency was not a significant predictor of English vocabulary. In the fourth and final step, we added our main measure of interest: boundary warping. As expected, boundary warping was a significant predictor of English vocabulary: Individuals with lower warping (higher gradiency) had higher English vocabulary scores (see Fig. 6)

Figure 6

Scatterplot showing English proficiency as a function of boundary warping



What predicts higher gradiency?

To address this question, we conducted a hierarchical regression with boundary warping (i.e., the reverse of gradiency) as the dependent variable. In the first step, we again accounted for the variance explained by a set of base variables that were not the focus of our study: working memory (Corsi) scores, music training, VAS response consistency, and AoA for Basque. As shown in Table 5, none of these factors were significant predictors of gradiency.

Table 5

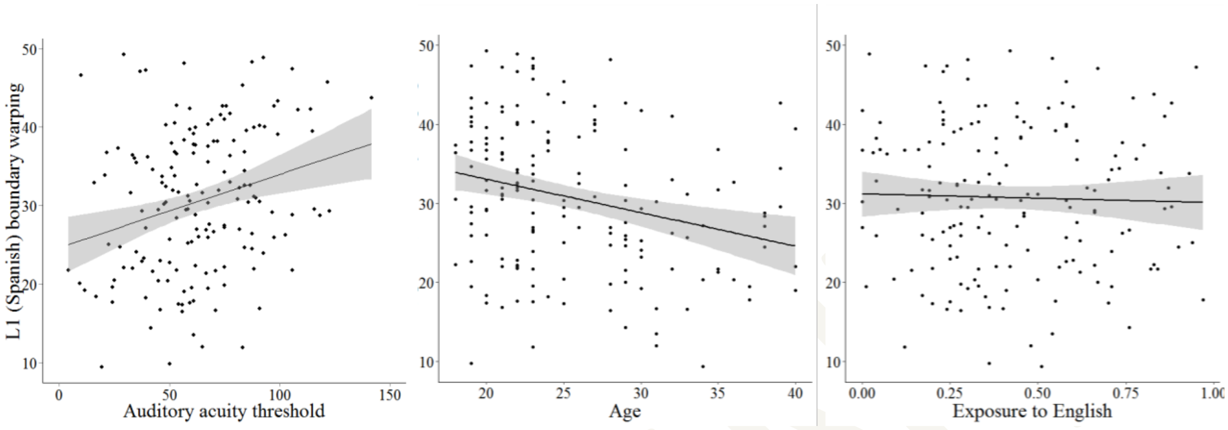
Hierarchical regression analysis predicting boundary warping (reverse of gradiency)

Step	Variable	*β*	*ΔR²*	*p*
1	Working memory (Corsi)	−.071		.378
	Music training	−.125		.124
	VAS response consistency	−.045		.576
	AoA Basque	−.001		.991
			.024	.460
2	Auditory acuity	.187		.025
	Age (exposure to spoken language)	−.266		.004
	AoA English	.001		.989
	Exposure English	.008		.921
			.104	.002

In the second step, we added our main measures of interest: auditory acuity, overall experience with spoken language (indexed by age), and our two measures of exposure to phonological diversity: AoA for English and auditory exposure to English. In line with the outlined hypotheses, age and auditory acuity were significant predictors of boundary warping; older participants (i.e., with more accumulated exposure to spoken language) and individuals with higher auditory acuity showed less warping (i.e., higher gradiency; see Fig.7). However, in contrast to the diverse-phonologies hypothesis, neither the AoA of English nor its current level of exposure was a significant predictor of boundary warping.

Figure 7

Scatterplots showing boundary warping as a function of Acoustic acuity (left), Exposure to spoken language/Age (middle), and Exposure to English (right)



Discussion

Our main goal was to test the hypothesis that gradiency in L1 speech perception helps listeners learn non-native phonemic contrasts – leading to better foreign language learning. Indeed, we found evidence that it does. After accounting for potentially important factors such as general auditory acuity, working memory, and exposure to the target language, an individual's sensitivity to subphonemic differences (i.e., speech perception gradiency) accounts for a significant portion of the remaining variance in foreign language proficiency (indexed by English vocabulary scores).

We propose that higher gradiency in L1 allows listeners to better perceive fine subphonemic differences that are meaningful in a foreign language. There are at least two ways in which this sensitivity could help foreign language learning. First, gradiency in L1 speech perception may facilitate the formation of new categories for non-native contrasts, e.g., by creating a new category boundary within the perceptual space of a native phoneme category. Second, L1 gradiency may allow listeners to maintain their sensitivity to within-category differences in real time, thus facilitating L2 spoken word recognition. Such a positive link between speech perception gradiency and spoken word recognition would be expected particularly in L2 speech perception, because L2 speech input does not always neatly map onto L1 phonological categories, which may cause misperceptions. Maintaining sensitivity to subphonemic information –instead of strongly committing to one phoneme category– should facilitate recovery from misperceptions (Kapnoula et al., 2021). In contrast, in L1 speech perception, stronger top-down/lexical reliance effects have been linked to *weaker* reliance on low-level acoustic-phonetic information (see Colby et al., 2018; Giovannone & Theodore, 2023). Regardless of the specific mechanism, gradiency can be broadly thought of as a “shield” against the assimilation of non-native sounds by native categories, either at a structural level, or in real-time processing. This interpretation is consistent with established accounts of non-native speech perception, such as the Native Language Magnet Model (Kuhl, 1993; Kuhl et al., 2008), and the Perceptual Assimilation Model (Best et al., 2001; Best & Tyler, 2007). Importantly, our findings expand these models by showing that there are individual differences in the degree of perceptual assimilation, driven by differences in L1 speech perception.

Our second goal was to examine the potential role of three factors as possible sources of gradiency: general auditory acuity, overall exposure to language (as indexed by age), and exposure to phonological diversity. We found evidence that two of these factors are linked to gradiency. First, we found a positive relationship between auditory acuity and gradiency. It is reasonable to expect that auditory acuity would help a listener perceive fine acoustic differences between speech sounds. However, the theoretical value of this finding goes beyond this assumption. Recent work has provided substantial evidence that domain-general auditory processing is linked to second language learning outcomes (Kachlicka et al., 2019; Kempe et al., 2014; Saito et al., 2021, 2022), but the exact mechanism underlying this relationship has been unclear. Our findings provide preliminary evidence that the missing link between auditory acuity and foreign language learning may be speech perception gradiency.

We also found evidence that gradiency increases with age. At first sight, this result might seem surprising, given that aging typically comes with hearing loss. However, we only tested adults up to 40 years old and our hearing screening assured that all participants had normal hearing. This finding extends the results reported by McMurray et al. (2018), who found that sensitivity to within-category differences slowly increases during childhood and adolescence. Together, these results show that speech perception improves with age; listeners get better at perceiving small VOT differences, and at the same time they get better at ignoring these differences when they are irrelevant for the task. Of course, we acknowledge that age encompasses many different experiences beyond exposure to language. Thus, the idea that it is specifically exposure to spoken language that drives this increase in L1 gradiency should be tested in future studies.

More broadly, the positive link between age and gradiency suggests that experience shapes the speech perception system so that it slowly becomes increasingly sensitive to fine-grained acoustics. An open question is whether the driving force behind this change is exposure to language in general, or exposure to a *specific kind* of linguistic input. Here, we tested one possible kind of input that may be particularly beneficial in increasing gradiency: phonologically diverse input. Evidence that phonological diversity affects gradiency would be consistent with distributional accounts of phonological learning (e.g., SLM-r; Flege & Bohn, 2021), as well as with empirical work showing that the speech perception system is shaped by linguistic exposure (Kutlu et al., 2022; Ramon-Casas et al., 2009; Samuel & Larraza, 2015; Sebastián-Gallés et al., 2005; Sebastián-Gallés & Bosch, 2009). Despite this, our results did not support this hypothesis. It is possible that there might not be enough variability among our participants in terms of their exposure to phonological diversity. All participants reside in a multicultural/multilingual environment (Spanish/Basque as official languages with some exposure to English and to French, given the nearby Spanish-French border). Thus, all participants might be close to ceiling in terms of phonological diversity. Perhaps future work can better test this hypothesis by including participants from a larger range of phonological diversity exposure (e.g., see preliminary evidence reported by Kutlu et al., 2022).

In sum, our results show that (1) speech perception gradiency is a significant predictor of foreign language proficiency and (2) gradiency increases with age and is linked to higher domain-general auditory acuity. At a theoretical level, these results provide important insights into the role of gradiency in speech processing and inform our understanding of how L1 speech perception is linked to foreign language learning. In addition, there are potentially broader implications, e.g., for foreign language educators, as our results can be used as a starting point for future research on the role of individual differences in L1 speech processing and the ways that these differences can be leveraged to facilitate second language learning.

Footnotes

[1] Throughout the manuscript, the term "gradiency" is meant to refer to gradiency in L1 speech perception; however, this should not be interpreted as a theoretical claim that L1 gradiency is distinct from L2 gradiency. Indeed, gradiency may very well be an individual trait that applies to different languages an individual knows. Whether gradiency is a crosslinguistic trait is a (very interesting) question that falls outside the scope of this study.

[2] Work by Dmitrieva et al. (2015) showed that in Spanish F0 differs between /b/ and /p/ during the first 53 ms into the vowel, with a typical difference of 1.24 semitones. In our recordings, the mean F0 value across the /ba/ and /pa/ steps during the first 53 ms was 186 Hz, which was the value set as the midpoint point for our pitch continuum. We set the two extreme points to 179 Hz and 193 Hz, a difference of 1.24 semitones.

[3] Pilot work showed that our participants' boundary was around -15 VOT, which is why we used this as the midpoint for our continuum.

[4] Eight fits (4 from the Speech-VAS and 4 from the Audio-VAS) were excluded due to problematic fits. We speculate that the fits of Audio-VAS curves were not as good as those of the other two VAS tasks because the trumpet-piano continuum was not as well centered as the other two continua, making it difficult for the fitter to find good fits.

References

Allopenna, P. D., Magnuson, J., & Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*, *38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, *52*(3), 163–187. https://doi.org/10.1016/0010-0277(94)90042-6

Apfelbaum, K. S., Kutlu, E., McMurray, B., & Kapnoula, E. (2022). Don't force it! Gradient speech categorization calls for continuous categorization tasks. *The Journal of the Acoustical Society of America*, *152*(6), 3728–3745. https://doi.org/10.1121/10.0015201

Berch, D. B., Krikorian, R., & Huha, E. M. (1998). The Corsi block-tapping task: Methodological and theoretical considerations. *Brain and Cognition*, *38*(3), 317–338. https://doi.org/10.1006/brcg.1998.1039

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775. https://doi.org/10.1121/1.1332378

Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception. In *Language* *experience in second language speech learning: In honour of James Emil Flege* (pp. 13–34).

Boersma, P., & Weenink, D. (2016). *Praat: doing phonetics by computer [Computer program]*.

Colby, S., Clayards, M., & Baum, S. (2018). The Role of Lexical Status and Individual Differences for Perceptual Learning in Younger and Older Adults. *Journal of Speech, Language, and Hearing Research*, *61*(8), 1855–1874. https://doi.org/10.1044/2018_JSLHR-S-17-0392

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain.*

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*(2), 218–236. https://doi.org/https://doi.org/10.1016/0749-596X(92)90012-M

de Bruin, A., Carreiras, M., & Duñabeitia, J. A. (2017). The BEST Dataset of Language Proficiency. *Frontiers in Psychology*, *8*(MAR), 522. https://doi.org/10.3389/fpsyg.2017.00522

Dmitrieva, O., Llanos, F., Shultz, A., & Francis, A. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*. https://doi.org/https://doi.org/10.1016/j.wocn.2014.12.005

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2017). MultiPic: A standardized set of 750 drawings with norms for six European languages. *The Quarterly Journal of Experimental Psychology*, 1–24. https://doi.org/10.1080/17470218.2017.1310261

Fink, M., Churan, J., And, M. W.-R. neurology, & 2005, U. (2005). Assessment of auditory temporal-order thresholds–A comparison of different measurement procedures and the influences of age and gender. *Restorative Neurology and Neuroscience*, *23*(5–6), 281–296.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In *Speech perception and linguistic experience: Issues in cross-language research* (Vol. 92, pp. 233–

277).

Flege, J. E., & Bohn, O. (2021). The revised speech learning model (SLM-r). In *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press.

Fostick, L., & Babkoff, H. (2013). Different response patterns between auditory spectral and spatial temporal order judgment (TOJ). *Experimental Psychology*, *60*(6), 432–443. https://doi.org/10.1027/1618-3169/A000216

Fuhrmeister, P., Phillips, M. C., McCoach, D. B., & Myers, E. B. (2023). Relationships between native and non-native speech perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/XLM0001213

Getz, L. M., & Toscano, J. C. (2021). The time-course of speech perception revealed by temporally-sensitive neural measures. In *Wiley Interdisciplinary Reviews: Cognitive Science* (Vol. 12, Issue 2, p. e1541). Wiley-Blackwell. https://doi.org/10.1002/wcs.1541

Giovannone, N., & Theodore, R. M. (2023). Do individual differences in lexical reliance reflect states or traits? *Cognition*, *232*, 105320. https://doi.org/10.1016/J.COGNITION.2022.105320

Goswami, U., Wang, H. L. S., Cruz, A., Fosker, T., Mead, N., & Huss, M. (2011). Language-universal sensory deficits in developmental dyslexia: English, Spanish, and Chinese. *Journal of Cognitive Neuroscience*, *23*(2), 325–337. https://doi.org/10.1162/JOCN.2010.21453

Hualde, J. I. (2015). Basque as an extinct language hualde. In B. Fernández & P. Salaburu (Eds.), *Ibon Sarasola, Gorazarre* (pp. 319–326). University of the Basque Country.

Joanisse, M. F., & Seidenberg, M. S. (1998). Specific language impairment: a deficit in grammar or processing? *Trends in Cognitive Sciences*, *2*(7), 240–247. https://doi.org/10.1016/S1364-6613(98)01186-3

Kachlicka, M., Saito, K., & Tierney, A. (2019). Successful second language learning is tied to robust domain-general auditory processing and stable neural representation of sound. *Brain* *and Language*, *192*, 15–24. https://doi.org/10.1016/J.BANDL.2019.02.004

Kapnoula, E. (2016). *Individual differences in speech perception: Sources, functions, and consequences of phoneme categorization gradiency*. University of Iowa.

Kapnoula, E., Edwards, J., & McMurray, B. (2021). Gradient activation of speech categories facilitates listeners' recovery from lexical garden paths, but not perception of speech-in-noise. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(4), 578–595. https://doi.org/https://doi.org/10.1037/xhp0000900

Kapnoula, E., & McMurray, B. (2021). Idiosyncratic use of bottom-up and top-down information leads to differences in speech perception flexibility: Converging evidence from ERPs and eye-tracking. *Brain and Language*, *223*, 105031. https://doi.org/10.1016/j.bandl.2021.105031

Kapnoula, E., Winn, M. B., Kong, E. J., Edwards, J., & McMurray, B. (2017). Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(9), 1594–1611. https://doi.org/10.1037/xhp0000410

Kempe, V., Bublitz, D., & Brooks, P. J. (2014). Musical ability and non-native speech-sound processing are linked through sensitivity to pitch and spectral information. *British Journal of Psychology*. https://doi.org/10.1111/bjop.12092

Kim, D., Clayards, M., & Kong, E. J. (2020). Individual differences in perceptual adaptation to unfamiliar phonetic categories. *Journal of Phonetics*, *81*, 100984. https://doi.org/10.1016/j.wocn.2020.100984

Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: cue weighting and executive function. *Journal of Phonetics*, *59*, 40–57. https://doi.org/10.1016/j.wocn.2016.08.006

Kuhl, P. K. (1993). Innate Predispositions and the Effects of Experience in Speech Perception: The Native Language Magnet Theory. *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, 259–274. https://doi.org/10.1007/978-94-015-8234-

6_22

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 11850–11857. https://doi.org/10.1073/PNAS.97.22.11850/ASSET/1393D943-C852-4BA2-A9E5-613B2757951F/ASSETS/GRAPHIC/PQ2003427008.JPEG

Kuhl, P. K., Conboy, B. T. B., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). *Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e)*. *363*(1493), 979. https://doi.org/10.1098/rstb.2007.2154

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science (New York, N.Y.)*, *255*(5044), 606–608. https://doi.org/10.1126/science.1736364

Kutlu, E., Fell, A., Apfelbaum, K., & McMurray, B. (2022). Effects of multilingual and monolingual social networks on speech perception. *The Journal of the Acoustical Society of America*, *152*(4), A236–A236. https://doi.org/10.1121/10.0016126

Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384–422. https://doi.org/https://doi.org/10.1080/00437956.1964.11659830

Llama, R., Cardoso, W., & Collins, L. (2010). The influence of language distance and language status on the acquisition of L3 phonology. *Https://Doi.Org/10.1080/14790710902972255*, *7*(1), 39–57. https://doi.org/10.1080/14790710902972255

Llanos, F., Dmitrieva, O., Shultz, A., & Francis, A. (2013). Auditory enhancement and second language experience in Spanish and English weighting of secondary voicing cues. *Journal of the Acoustical Society of America*, *134*(3), 2213–2224. https://doi.org/10.1121/1.4817845

Massaro, D. W., & Cohen, M. M. (1983). Categorical or continuous speech perception: A new test. *Speech Communication*, *2*(1), 15–35. https://doi.org/10.1016/0167-6393(83)90061-4

McMurray, B. (2017). *Nonlinear Curvefitting for Psycholinguistic (and other) Data*. OSF.

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient Sensitivity to Within-Category Variation in Words and Syllables. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1609–1631. https://doi.org/10.1037/a0011747

McMurray, B., Danelz, A., Rigler, H., & Seedorff, M. (2018). Speech categorization develops slowly through adolescence. *Developmental Psychology*, *54*(8), 1472. https://doi.org/10.1037/dev0000542

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition : Implications for SLI. *Cognitive Psychology*, *60*(1), 1–39. https://doi.org/10.1016/j.cogpsych.2009.06.003

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33–B42. https://doi.org/10.1016/S0010-0277(02)00157-9

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from lexical garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language*, *60*(1), 65–91. https://doi.org/10.1016/j.jml.2008.07.002

Miller, J. L. (1997). Internal Structure of Phonetic Categories. *Language and Cognitive Processes*, *12*(5–6), 865–870. https://doi.org/10.1080/016909697386754

Munson, B., Edwards, J., Schellinger, S. K. S. K., Beckman, M. E., & Meyer, M. K. (2010). Deconstructing phonetic transcription: Covert contrast, perceptual bias, and an extraterrestrial view of Vox Humana. *Clinical Linguistics & Phonetics*, *24*(4–5), 245–260. https://doi.org/10.3109/02699200903532524

Pitt, M. A. (1995). Evidence for a central representation of instrument timbre. *Perception & Psychophysics*, *57*(1), 43–55. https://doi.org/10.3758/BF03211849/METRICS

Ramon-Casas, M., Swingley, D., Sebastián-Gallés, N., & Bosch, L. (2009). Vowel categorization during word recognition in bilingual toddlers. *Cognitive Psychology*, *59*(1), 96–121. https://doi.org/10.1016/j.cogpsych.2009.02.002

Saito, K., Kachlicka, M., Suzukida, Y., Petrova, K., Lee, B. J., & Tierney, A. (2022). Auditory precision hypothesis-L2: Dimension-specific relationships between auditory processing and second language segmental learning. *Cognition*, *229*, 105236. https://doi.org/10.1016/J.COGNITION.2022.105236

Saito, K., Suzukida, Y., Tran, M., & Tierney, A. (2021). Domain-General Auditory Processing Partially Explains Second Language Speech Learning in Classroom Settings: A Review and Generalization Study. *Language Learning*, *71*(3), 669–715. https://doi.org/10.1111/LANG.12447

Samuel, A. G. (1977). The effect of discrimination training on speech perception: Noncategorical perception. *Perception & Psychophysics*, *22*(4), 321–330. https://doi.org/10.3758/BF03199697

Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics*, *31*(4), 307–314. https://doi.org/10.3758/BF03202653

Samuel, A. G., & Larraza, S. (2015). Does listening to non-native speech impair speech perception? *Journal of Memory and Language*, *81*, 51–71. https://doi.org/https://doi.org/10.1016/j.jml.2015.01.003

Sarrett, M. E., McMurray, B., & Kapnoula, E. (2020). Dynamic EEG analysis during language comprehension reveals interactive cascades between perceptual processing and sentential expectations. *Brain and Language*, *211*, 104875. https://doi.org/10.1016/j.bandl.2020.104875

Sebastián-Gallés, N., & Baus, C. (2005). *Twenty-first Century Psycholinguistics: Four Cornerstones* (A. Cutler (ed.); p. 412).

Sebastián-Gallés, N., & Bosch, L. (2009). Developmental shift in the discrimination of vowel contrasts in bilingual infants: is the distributional account all there is to it? *Developmental Science*, *12*(6), 874–887. https://doi.org/10.1111/j.1467-7687.2009.00829.x

Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory*

And, *52*(2), 240–255. https://doi.org/https://doi.org/10.1016/j.jml.2004.11.001

Serniclaes, W., Sprenger-Charolles, L., Carré, R., & Démonet, J. (2001). Perceptual discrimination of speech sounds in developmental dyslexia. *Journal of Speech, Language, and Hearing Research*, *44*(2), 384–399. https://doi.org/10.1044/1092-4388(2001/032)

Serniclaes, W., Van Heghe, S., Mousty, P., Carré, R., & Sprenger-Charolles, L. (2004). Allophonic mode of speech perception in dyslexia. *Journal of Experimental Child Psychology*, *87*(4), 336–361. https://doi.org/10.1016/j.jecp.2004.02.001

Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *The Journal of the Acoustical Society of America*, *106*(2), 1078–1083. https://doi.org/10.1121/1.428048

Sharma, A., Marsh, C. M., & Dorman, M. F. (2000). Relationship between N1 evoked potential morphology and the perception of voicing. *The Journal of the Acoustical Society of America*, *108*(6), 3030–3035. https://doi.org/https://doi.org/10.1121/1.1320474

Simon, J., & Winkler, I. (2018). The role of temporal integration in auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(11), 1683–1693. https://doi.org/10.1037/xhp0000564

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization: electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science*, *21*(10), 1532–1540. https://doi.org/10.1177/0956797610384142

Velázquez López, D. (2015). *A contrastive analysis of the English and Spanish phonetic systems with special emphasis on andalusian accent features*. Universidad de Cádiz.

Yu, A. C. L., & Zellou, G. (2019). Individual Differences in Language Processing: Phonology. *Annual Review of Linguistics*, *5*(1), 131–150. https://doi.org/10.1146/annurev-linguistics-011516-033815