Perceptual Assimilation of Brazilian Portuguese Mid Vowels into Catalan: A Multimodal Study of Bilinguals with Spanish or Catalan Dominance

Asimilación perceptiva de las vocales medias del portugués brasileño en el catalán: un estudio multimodal en bilingües con dominancia del español o del catalán

Ubiratã Kickhöfel Alves
Universidade Federal do Rio Grande do Sul
ukalves@gmail.com
Larissa Cristina Berti
Universidade Estadual Paulista "Júlio de Mesquita Filho"
larissa.berti@unesp.br

Abstract

This study examines how language dominance and stimulus modality influence perceptual assimilation across languages. Brazilian Portuguese vowels were used in a perceptual classification task with early Catalan-Spanish bilinguals. The task focused on the mid-vowel pairs /e/-/ɛ/ and /o/-/ɔ/, which occur in Brazilian Portuguese and Catalan but not in Spanish. Stimuli were presented in four randomized conditions: auditory-only (AO), visual-only (VO), congruent audiovisual (AV+), and incongruent audiovisual (AV–). Participants identified the vowels as /e/, /ɛ/, /o/, or /ɔ/ in Catalan. The study included 30 self-reported bilinguals with no prior exposure to Brazilian Portuguese: 13 Spanish-dominant and 17 Catalan-dominant. Results showed a significant effect of language dominance: Catalan-dominant participants achieved higher accuracy in AO, VO, and AV+ conditions. Across groups, auditory cues were consistently prioritized, and congruent audiovisual stimuli did not yield higher accuracy than auditory-only stimuli. For incongruent stimuli, both groups favored auditory information, particularly among Catalan-dominant participants. These findings underscore the impact of bilingual language dominance on non-native vowel perception and inform ongoing discussions about the role of auditory and visual cues in bilingual speech processing.

Keywords: bilingualism; perceptual assimilation; language dominance; stimulus modality; audiovisual information.

Resumen

Este estudio investiga el efecto de la dominancia lingüística y de la modalidad del estímulo en la asimilación perceptiva entre lenguas. Se utilizaron vocales del portugués brasileño en una tarea de clasificación perceptiva con bilingües catalán-español. La tarea se centró en los pares de vocales medias /e/-/ɛ/ y /o/-/ɔ/, contrastivos en portugués brasileño y catalán, pero no en español. Los estímulos se presentaron en cuatro condiciones: solo auditiva (AO), solo visual (VO), audiovisual congruente (AV+) y audiovisual incongruente (AV–). Los participantes identificaron las vocales catalanas /e/, /ɛ/, /o/ y /ɔ/. Participaron treinta bilingües sin exposición previa al portugués brasileño: 13 con dominancia del español y 17 con dominancia del catalán. Los resultados mostraron un efecto significativo de la dominancia lingüística, con mayor precisión en los participantes con dominancia del catalán en AO, VO y AV+. En ambas poblaciones se

priorizó la información auditiva, sin ventaja para los estímulos audiovisuales congruentes. En los estímulos incongruentes, ambos grupos favorecieron la información auditiva, especialmente el grupo del catalán. Estos resultados resaltan el impacto de la dominancia lingüística en la percepción de vocales no nativas y contribuyen al debate sobre las pistas auditivas y visuales en el procesamiento bilingüe del habla.

Palabras clave: bilingüismo; asimilación perceptiva; dominancia lingüística; modalidad del estímulo; información audiovisual.

1. INTRODUCTION

Research on the perception of nonnative sounds distinguishes between second-language (L2) and cross-language speech perception (Bohn, 2018). Although these two strands share common interests (most notably the "mapping issue" (Bohn, 2018), which concerns how nonnative sounds are assimilated into a listener's L1 system), they differ in scope. Second-language studies address the gradual development of a new phonological system, whereas cross-language studies focus on "naïve" listeners (Best, 1995) who encounter the sounds of an unfamiliar language for the first time in an experimental setting.

Second-language speech perception studies are abundant, drawing on data from learners at different proficiency levels acquiring various L2s (including multilingual systems). This paper, however, focuses on cross-language speech perception. Although fewer in number than second-language studies, numerous cross-language investigations have been conducted over the past three decades (Best *et al*., 2001; Tyler *et al*., 2014; Faris *et al*., 2018; Li *et al*., 2023, among others). These studies examine the assimilation, or mapping, of L2 sounds into an individual's L1 phonological system. Analyzing such assimilation patterns in cross-language research is also highly relevant to L2 speech studies, because the earliest stage of L2 development involves assimilating L2 sounds into the preexisting L1 system.

The growing number of cross-language studies over the past three decades stems from the research agenda initiated by Best's (1995) Perceptual Assimilation Model (henceforth PAM). In its original formulation, the model focused exclusively on the role of L1 attunement in naïve monolingual adult listeners' perception. PAM is grounded in a direct-realist view of speech perception (Fowler, 1986, 1996) and assumes phonological gestures (Browman & Goldstein, 1986, 1992) as its basic units. To account for this attunement, the model predicts both categorization and discrimination patterns for pairs of nonnative phonemes mapped into the L1 system[1]. The model was later extended to L2 learner contexts (PAM-L2; cf. Best & Tyler, 2007), but its central tenet remains the same across both versions: naïve listeners and beginning L2 learners are attuned to higher-order invariants (namely, the articulatory properties common to all variants of a given phonological category) of their native language.

Despite the growing body of cross-language perception research, the vast majority has focused on monolingual "naïve" listeners. This is not surprising, given that the original version of PAM was designed specifically for this listener type. Much remains to be

[1] As this paper does not address the general predictions of the model, readers are referred to Best (1995) for further information.

explored about the cross-linguistic perception of early bilinguals (that is, individuals who began learning two languages at an early age). As Tyler and Best (2024: 164) observe, "studies on bilingual speech perception have only begun to scratch the surface of how early bilingual language experience shapes perception."

Characterizing early bilinguals is challenging, as Tyler and Best (2024) explain, because key questions remain about whether the two languages were acquired simultaneously, whether one language is dominant, and whether the language of testing influences the results. This complexity helps explain why relatively few cross-linguistic studies (e.g., Antoniou *et al*., 2013; Melguy, 2018) have been conducted with early bilinguals. In this context, Tyler and Best (2024) discuss how the principles of PAM and PAM-L2 can be extended to account for cross-linguistic studies with early bilinguals. They argue that the basic assumptions of PAM are applicable to this population, proposing a single phonological system for both languages in which language-specific and shared categories coexist. Within this unified system, nonnative phonological units are expected to be assimilated into existing categories, and, as in PAM, listeners are sensitive to higher-order invariants that allow them to distinguish between nonnative categories. However, when two nonnative categories are assimilated into a single category in the common system, bilinguals shift attention to lower-order articulatory detail. According to Tyler and Best (2024), this shift occurs in a unilingual mode, most likely in the dominant language, unless task demands favor activation of the nondominant language—which typically requires greater cognitive effort. This extension of PAM to early bilinguals opens promising avenues for future research in cross-language perception, paving the way for more studies focusing on early bilingual individuals.

A common feature of cross-linguistic studies is their exclusive focus on auditory input. This reflects the methodological challenges of incorporating audiovisual information in laboratory settings and the fact that PAM itself is centered on auditory stimuli. By contrast, research on L2 acquisition has begun to incorporate audiovisual input into experimental designs. While many studies report an advantage for audiovisual over auditory-only input (Reisberg *et al*., 1987; Hardison, 2005; Navarra & Soto-Faraco, 2007; Hirata & Kelly, 2010; Springer, 2021), others have found no such advantage (Hazan *et al*., 2002, 2006; Reyes & Hazan, 2021). Outcomes clearly depend on several variables that influence L2 speech research more broadly, including the phoneme under investigation, learners' proficiency levels, and task demands. This variability underscores the need for further multimodal L2 studies. Identifying how visual input contributes at different proficiency levels may also inform pedagogical practices aimed at enhancing L2 perception. In line with this, studies on L2 perceptual training (Iba *et al*., 2004; Hardison, 2003, 2005; Hazan *et al*., 2005; Hazan & Sennema, 2007; Li, 2015) demonstrate the benefits of audiovisual input in training sessions, further reinforcing the importance of expanding research on visual information.

As in L2 perceptual studies, this paper advocates conducting multimodal cross-linguistic research. Our argument rests on both empirical and theoretical grounds. Empirically, audiovisual information has been shown to contribute to L2 speech perception (Reisberg *et al*., 1987; Hardison, 2005; Navarra & Soto-Faraco, 2007; Hirata & Kelly, 2010; Springer, 2021). Given that beginning L2 learners often display strong L1 attunement (Best & Tyler, 2007), it is plausible that perceptual assimilation is modality-dependent (e.g., auditory-only, visual-only, or audiovisual). In other words, examining how multimodal input shapes perceptual assimilation may reveal distinct assimilation patterns

and, in the context of L2 acquisition, may help identify which modalities support learners in forming L2 categories.

Theoretically, PAM should be able to accommodate audiovisual information in an epistemologically coherent manner. PAM posits articulatory gestures as its primitive units, which constitute the source of both higher- and lower-order information. As argued in previous work (Alves *et al*., 2025), adopting articulatory gestures within a direct-realist framework of speech perception provides a basis for bridging PAM-L2 and multimodal models. Articulatory gestures function as the common currency between perception and production, as well as between acoustic and visual information (Goldstein & Fowler, 2003). As Fowler (1996: 1733) notes, "[…] listeners perceive gestures, and some gestures are specified optically as well as acoustically." Tyler and Best (2024: 149) further explain that "the multimodal output (acoustic signal, facial movements, airflow) of these temporally coordinated vocal tract constriction events provides evidence of the originating articulatory patterns that perceivers detect in speech." In line with Alves *et al*. (2025), the principle of a common gestural unit aligns with Rosenblum's (2005, 2008) account of multimodal perception. Although it is beyond the scope of this paper to examine in detail the potential adaptations required to fully integrate audiovisual information into PAM, we maintain that such a theoretical–methodological extension is epistemologically sound and worth pursuing.

Given the discussion so far, it is clear that cross-modal studies have focused primarily (if not exclusively) on monolinguals exposed to auditory-only input. Proposing cross-language research with early bilinguals that incorporates different input modalities (e.g., auditory-only, visual-only, and audiovisual) could generate a range of new research questions. For instance, one question is whether cross-modal perception outcomes are influenced by stimulus modality in the same way across both languages. Another important question is whether bilinguals are differentially affected by stimulus modality when assimilating sounds into their more- or less-dominant language.

The present study addresses the latter question by examining how Brazilian Portuguese (BP) vowels are assimilated by two groups of early bilinguals born and raised in Catalonia, Spain: Spanish-dominant and Catalan-dominant bilinguals. Our analysis focuses on the BP mid-vowels /e/, /ɛ/, /o/, and /ɔ/. These vowels occur in both the BP and Catalan phonological inventories but are absent from Spanish, whose system comprises five vowels (/i/, /e/, /a/, /o/, /u/). Despite their apparent similarities, BP and Catalan vowels exhibit phonetic differences. Escudero et al. (2009), for instance, examined 10 male speakers from São Paulo[2], while Recaséns and Espinosa (2006) and Recaséns (2014) analyzed data from 5 male speakers from eastern Catalonia (the Barcelona region)[3]. Although based on different methodologies, these studies converge on the suggestion that BP vowel assimilation may pose challenges for Catalan-dominant bilinguals, who may need to rely on both higher- and lower-order invariants to perform the assimilation task.

[2] Escudero *et al*. (2009) report the following formant values for BP vowels (in Hz): F1 – /i/ 283, /e/ 356, /ɛ/ 512, /a/ 677, /ɔ/ 538, /o/ 373, /u/ 307; F2 – /i/ 2204, /e/ 2026, /ɛ/ 1834, /a/ 1340, /ɔ/ 942, /o/ 831, /u/ 802.

[3] Recaséns and Espinosa (2006) and Recaséns (2014) report the following formant values for Catalan vowels (in Hz): F1 – /i/ 334, /e/ 450, /ɛ/ 581, /a/ 730, /ɔ/ 608, /o/ 489, /u/ 304; F2 – /i/ 2078, /e/ 1839, /ɛ/ 1700, /a/ 1358, /ɔ/ 1125, /o/ 1047, /u/ 960.

Similarly, the formant values of Spanish /e/ and /o/ differ phonetically from those of BP and Catalan, as shown by Ayuso and Olivé (2016)[4], who examined two Spanish-dominant bilinguals from Catalonia. Thus, assimilating BP into Catalan, their non-dominant language, may also require selective attention to both higher- and lower-order articulatory invariants — particularly given the possibility that their Catalan vowel categories diverge from those of Catalan-dominant speakers, as they may be shaped by Spanish. This scenario may give rise to shared or hybrid categories for /e/ and /o/, and to /ɛ/ and /ɔ/ categories that, due to their weaker dominance, do not fully resemble those of Catalan-dominant speakers.

These predictions align with studies on the auditory perception of Catalan /e/ and /ɛ/ among Spanish–Catalan bilinguals. Pallier et al. (1997), for example, reported that even bilinguals who acquired Catalan at a very early age did not fully master this contrast. In contrast, Amengual (2016) found that participants achieved high accuracy in both discrimination and assimilation tasks. Taken together, these mixed results underscore the need for further research on the perception of mid vowels among early Catalan–Spanish bilinguals.

Our study differs from those of Pallier et al. (1997) and Amengual (2016) in two main respects: (i) it is designed as a cross-language study, and (ii) it examines the effect of multimodal stimulus presentation. We addressed the first aspect in this introduction, where we emphasized the need for additional cross-linguistic studies with bilinguals. Regarding the second aspect, Navarra and Soto-Faraco (2007) investigated how three speech modalities (AV, AO, and VO) affect the perception of Catalan /e/ and /ɛ/ among Catalan-dominant and Spanish-dominant bilinguals. Their results show that, unlike Catalan-dominant participants, Spanish-dominant bilinguals did not exhibit sensitivity to the /e/–/ɛ/ contrast in AO stimuli, though both groups succeeded when audiovisual stimuli were presented. On this basis, the authors concluded that audiovisual input can significantly facilitate perception in the nondominant language.

Following Navarra and Soto-Faraco (2007) and Barros-Loscertales *et al*. (2013), our cross-language assimilation experiment examined the effects of Auditory Only (AO), Visual Only (VO), and Audiovisual (AV+) stimuli. The study also included test items with incongruent audio–visual input (AV–). Research on this modality appears in a few L2 speech studies (Sekiyama & Burnham, 2008; Wang *et al*., 2008; Chen & Hazan, 2009; Hazan *et al*., 2010; Barros-Loscertales *et al*., 2013; Marian *et al*., 2018, among others). These studies typically test the McGurk effect (McGurk & MacDonald, 1976) in monolinguals and bilinguals, showing that bilinguals tend to experience the effect more strongly than monolinguals (Marian *et al*., 2018).

In our experiment, however, L1 assimilation was examined using a forced-choice identification task in which AV– stimuli had to be classified as either a high-mid vowel (/e/, /o/) or a low-mid vowel (/ɛ/, /ɔ/). This design allowed us to assess the relative reliance on auditory versus visual information. To our knowledge, this is the first study to examine such reliance in AV–assimilation of an unfamiliar language into the more- or less-dominant language of early bilinguals.

[4] Ayuso and Olivé (2016) report the following formant values for Catalan vowels (in Hz): F1 – /i/ 321, /e/ 514, /a/ 575, /ɛ/ 529, /u/ 329.

From a direct-realist perspective, and given the theoretical grounding of PAM-L2, choosing between mid-low and mid-high vowels under AV– exposure does not require the computational integration of acoustic and visual cues, as posited by models of audiovisual integration such as the Fuzzy Logical Model of Perception (FLMP; Massaro, 1998, 2004) or Bayesian frameworks (Knill & Pouget, 2004; Ernst & Bülthoff, 2004). Rather, auditory and visual inputs are integral to articulatory gestures. This view aligns with Rosenblum's (2005, 2008) account of speech perception, in which gestures are amodal units that directly provide both sources of information. In this approach, "information from different sensory modalities does not require integration within the perceiver's brain but instead takes on a common higher-order informational form from the outset" (Alves *et al*., 2025: 88).

Within a PAM framework, we propose that selecting between auditory and visual input in incongruent stimuli reflects shifts in attention, much like shifting focus between higher- and lower-order articulatory information. These attentional shifts are not random; rather, they are shaped by task demands and the need for attunement in cross-language contexts. Research has shown that less proficient L2 learners rely more heavily on visual input (Wang *et al*., 2008). By contrast, in the early bilinguals examined here, well-established higher-order categories are likely to be processed primarily through auditory input, whereas assimilation difficulties may prompt a shift toward reliance on visual cues—particularly lower-order articulatory information when higher-order invariants are insufficient.

This study addresses three research questions:

- **RQ1** – Does language dominance (Catalan more vs. less dominant relative to Spanish) affect the assimilation of Brazilian Portuguese vowels into Catalan categories?
- RQ2 – Do assimilation rates differ across stimulus modalities, and do these differences vary with participants' dominant language (Catalan vs. Spanish)?
- **RQ3** – For incongruent stimuli (AV–), does reliance on auditory vs. visual information vary as a function of language dominance (Catalan/Spanish)?

The first two questions concern assimilation patterns across AO, VO, and AV+ stimuli, while the third focuses on resolving cross-modal conflict in speech. When asked to assimilate BP vowels into Catalan, Spanish-dominant bilinguals are expected to be more likely to shift attention to visual information under AV– conditions. For AV+, AO, and VO stimuli, we assume that gestures convey both auditory and visual information, though AO and VO are expected to be more challenging than AV+. Specifically, AO should remain sufficiently informative for listeners to recover articulatory information, whereas VO should be less effective, providing only partial access to auditory information. This distinction is especially relevant for Spanish-dominant participants, whose categories for /e/ and /o/ show strong Spanish influence and whose /ɛ/ and /ɔ/ categories are not fully established. Accordingly, Spanish-dominant listeners are expected to achieve higher accuracy in AV+ than in AO or VO conditions. Finally, given our assumption that listeners attend more readily to auditory than to visual information, AO is predicted to yield higher accuracy than VO.

2. METHOD

2.1. Participants

Thirty participants, all born and living in the city of Barcelona or its metropolitan area, took part in the study. All were university-aged (18 to 28 years old) Catalan–Spanish bilinguals and were divided into two groups based on their degree of language dominance: (i) Catalan-dominant bilinguals and (ii) Spanish-dominant bilinguals.

To determine dominance in one of the two languages, the following recruitment criteria were applied: participants had to (i) report using only one of the two languages to communicate with both parents in childhood and in adulthood; (ii) consider that language the one in which they felt most comfortable communicating; (iii) have used that language in at least 80% of interactions during childhood; and (iv) use that language in at least 80% of their daily interactions (cf. Ayuso & Olivé, 2016).

Based on these criteria, the Catalan-dominant group included 17 participants, and the Spanish-dominant group included 13. As noted earlier, they had no prior exposure to Brazilian Portuguese. All participants had intermediate or advanced proficiency in English but not in any other Romance language.

2.2. Stimulus Preparation

The perceptual assimilation task was developed using oral productions from four Brazilian speakers (two male and two female, aged 22–26), all born and living in a city in the state of São Paulo, Brazil. Audio and video recordings of the stimuli were made in a studio in São Paulo using a Logitech MX Brio 705 4K webcam and a Sony ECM-CS3 omnidirectional microphone. Recordings were conducted in front of a green-screen background; all speakers wore plain black clothing and no accessories, such as piercings, necklaces, or earrings.

Each speaker was asked to produce five isolated repetitions of the Brazilian Portuguese mid vowels /e/, /ɛ/, /o/, /ɔ/, as well as the vowel /a/, which served as a catch trial to ensure participants' attention to the task (accuracy rates below 80% for /a/ in the AV and AO modalities resulted in exclusion from the study, as they suggested lack of attention or task commitment). After recording five repetitions of each vowel, all tokens were analyzed acoustically by the first author for F1, F2, and duration. The token closest to the reference values reported in the literature was selected as the experimental stimulus.

Once the target productions were selected, stimuli were manually edited in *Filmora* to produce two-second clips. Video clips were cropped to show the lower face (from the shoulders to the nose), excluding the eyes. Using the same software and the original audio–video recordings, further manipulations were performed to create stimuli in the AO, VO, and incongruent AV modalities.

2.3. Perceptual assimilation task and data collection procedures

The perceptual assimilation task was administered via the online platform *SurveyMonkey*. The task included stimuli in the AO, VO, AV+, and AV– modalities, presented in random order by the platform. In total, 76 stimuli were used: 20 stimuli for each of the AO, VO, and AV+ conditions (4 speakers × 5 vowels), and 16 stimuli for the AV– condition (4

speakers for each of the incongruent pairs: "/e/ audio – /ɛ/ video," "/ɛ/ audio – /e/ video," "/o/ audio – /ɔ/ video," "/ɔ/ audio – /o/ video").

In this task, bilingual participants were informed that they would hear sounds from an unfamiliar language and instructed to classify them into the Catalan vowel categories /a/, /e/, /ɛ/, /o/, and /ɔ/. For each stimulus, five response options were provided: (i) *a*, (ii) *é*, (iii) *è*, (iv) *ó*, and (v) *ò*. These options were based on Catalan grapho-phonological conventions and were presented with analogies to Catalan words, such that *é* and *ó* represented the mid-high vowels [e, o], while *è* and *ò* represented the mid-low vowels [ɛ, ɔ].

Once participants met the eligibility criteria, confirming dominance in either Catalan or Spanish, they received an email with a link to complete the task on the *SurveyMonkey* platform. Participants were instructed to complete the task in a quiet environment using a computer. Task instructions were always provided in participants' dominant language. Specifically, the Spanish-dominant group received all instructions in Spanish, while the Catalan-dominant group received them in Catalan. This ensured that participants could fully understand the task, given their strong dominance in one of the two languages.

All participants completed the 76-item assimilation task. On average, they required 25 to 45 minutes to complete all tasks.

2.4.Statistical analysis procedures

Data were extracted from the *SurveyMonkey* platform and organized into Excel spreadsheets for subsequent statistical analysis in *RStudio* (R Core Team, 2025).

To address the first two research questions, we conducted mixed-effects logistic regressions with "accuracy" as the response variable, "Group" (Spanish-dominant vs. Catalan-dominant) and "Modality" (AO, VO, AV+) as fixed effects, and random intercepts for "Participant" and "Test Item".

To address the third research question regarding incongruent stimuli, we conducted a mixed-effects logistic regression with "preference for auditory vs. visual information" as the response variable, "Group" (Spanish-dominant vs. Catalan-dominant) as the fixed effect, and random intercepts for Participant and Item.

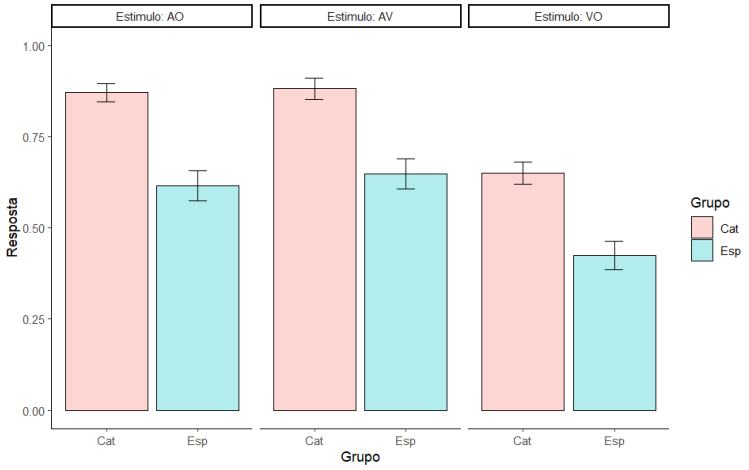
3. RESULTS

We first report results addressing the first two research questions, which examine participants' responses to the AO, VO, and AV+ stimuli. We then turn to the third research question, which focuses on responses to incongruent AV stimuli. For clarity, we present these two sets of results separately.

3.1. AO, VO, and AV+ stimuli

To address RQ1 and RQ2, Figure X presents the mean percentage accuracy[5] for assimilating mid vowels across the AO, VO, and AV+ conditions.

Figure 1. Mean accuracy rates in the assimilation of mid vowels (AO, VO, and AV+ stimuli).



In the Catalan-dominant group, both the AO and AV modalities yielded high accuracy rates (M = 87.13%, SD = 14.26; and M = 88.24%, SD = 17.12, respectively). However, the VO modality showed lower accuracy (M = 65.07%) and the highest standard deviation among the three modalities (SD = 18.15). These results suggest greater difficulty and greater variability in responses to this modality, consistent with previous findings that VO stimuli are particularly challenging (Navarra & Soto-Faraco, 2007; Barros-Loscertales *et al*., 2013).

In the Spanish-dominant group, AO and AV accuracy rates were comparable (61.54% and 64.90%, respectively) but notably lower than those of the Catalan-dominant group, as shown in Figure 1. The standard deviations for AO and AV in this group were 21.19% and 20.92%, respectively, indicating greater variability. This suggests that, despite lower overall accuracy, Spanish-dominant participants exhibited greater uncertainty in their responses. For the VO modality, accuracy was again lower than for the other two modalities (M = 42.5%, SD = 19.09%).

Overall, the descriptive data indicate that the Catalan-dominant group achieved higher accuracy than the Spanish-dominant group across all three modalities. However, in both groups, the VO condition yielded the lowest accuracy, highlighting the greater difficulty of assimilating stimuli without auditory information.

[5] By *accuracy*, we refer to cases in which BP vowels were assimilated to the corresponding Catalan categories (namely, when the BP vowels /e/, /ɛ/, /o/, and /ɔ/ were identified as /e/, /ɛ/, /o/, and /ɔ/, respectively).

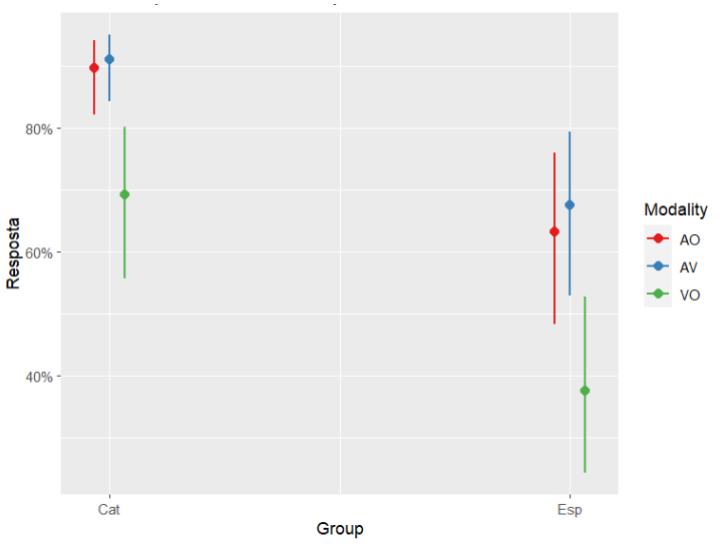
Table 1 presents the results of a mixed-effects logistic regression with "Accuracy" as the response variable, "Group" and "Modality" as fixed effects, and random intercepts for "Participant" and "Test Item". In this model, we assess the effects of the predictor variables on accuracy rates.

Table 1. Mixed-Effects Logistic Regression Model (AO, VO, and AV+ conditions) considering participants' accuracy rates in vowel assimilation.

Predictors	*Log-Odds*	*CI*	*p*
(Intercept)	2.15	1.52 – 2.78	**<0.001**
Group [Esp]	-1.61	-2.26 – -0.97	**<0.001**
Modality [AV]	0.17	-0.63 – 0.96	0.679
Modality [VO]	-1.34	-2.10 – -0.59	**<0.001**
Group [Esp] × Modality [AV]	0.02	-0.66 – 0.70	0.948
Group [Esp] × Modality [VO]	0.29	-0.35 – 0.93	0.376
Random Effects			
σ^2	3.29		
$\tau_{00\ Question}$	0.73		
$\tau_{00\ Participant}$	0.37		
ICC	0.25		
$N_{\ Participant}$	30		
$N_{\ Question}$	48		
Observations	1440		
Marginal R^2 / Conditional R^2	0.178 / 0.383		

The model's predicted probabilities are shown in Figure 2.

Figure 2. Predicted probabilities of accuracy rates from the mixed-effects logistic regression model (AO, VO, and AV+ conditions).

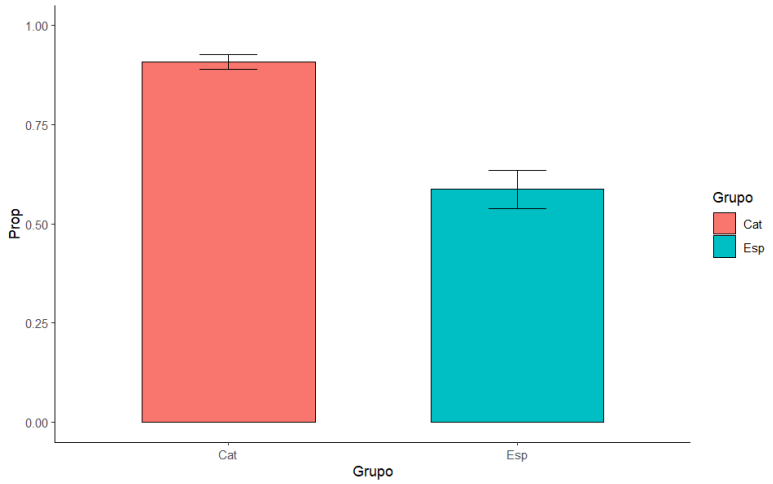


The model indicates that: (i) Spanish-dominant participants were significantly less likely to correctly identify the vowels than Catalan-dominant participants; (ii) the VO condition was associated with a marked decrease in the likelihood of correct responses; (iii) no significant differences emerged between the AO and AV+ conditions; and (iv) no significant interactions were found between Group and Modality. This suggests that the main effects account for most of the observed variation, as both groups — albeit at different rates — achieved higher accuracy in the AO and AV+ conditions. A noteworthy finding is the absence of an advantage for audiovisual stimuli over audio-only stimuli.

3.2. AV– Stimuli

The third research question examined whether the two groups differed in their preference for responding to incongruent (AV–) stimuli. Figure 3 presents the preference rates for the auditory signal in both groups.

Figure 3. Preference for auditory information in incongruent (AV–) stimuli.



The figure shows a marked difference in preference for auditory information when processing incongruent stimuli. In the Catalan-dominant group, the preference for the auditory signal was 91.17% (SD = 8.86). In contrast, the Spanish-dominant group showed a more balanced pattern, with 58.16% (slightly above chance) favoring the auditory signal. The standard deviation for auditory preference in the Spanish-dominant group was also higher (SD = 21.12), indicating greater variability and uncertainty in their choices between auditory and visual information.

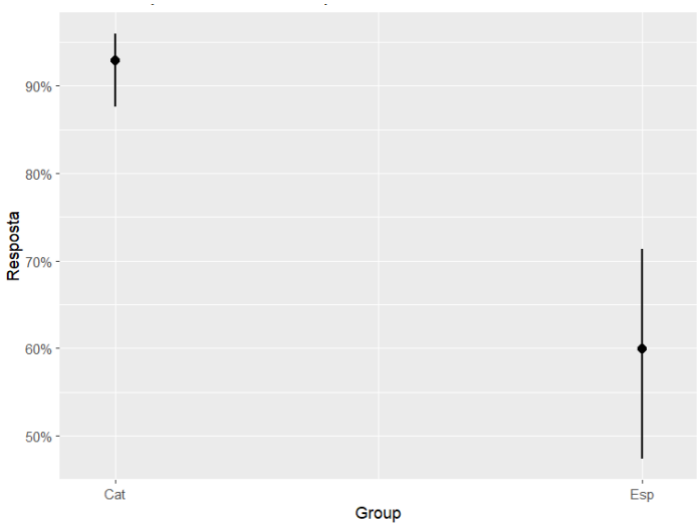
To assess the inferential validity of these results, we fitted a mixed-effects logistic regression model with "Preference for auditory vs. visual information" as the response variable, "Group" as the fixed effect, and random intercepts for "Participant' and "Test Item".

Table 2. Mixed-Effects Logistic Regression Model (AV condition) considering the preference for audio information.

Predictors	*Log-Odds*	*CI*	*P*
(Intercept)	2.56	1.95 – 3.17	**<0.001**
Group [Esp]	-2.16	-2.92 – -1.39	**<0.001**
Random Effects			
σ^2	3.29		
$\tau_{00\ Participant}$	0.53		
$\tau_{00\ Question}$	0.05		
ICC	0.15		
$N_{\ Participant}$	30		
$N_{\ Question}$	16		
Observations	480		
Marginal R^2 / Conditional R^2	0.228 / 0.343		

The prediction plot obtained from the model is presented in Figure 4.

Figure 4. Predicted probabilities of choosing the audio stimuli obtained from the mixed-effects logistic regression mode (AV- condition).



The model in Table 2 estimates preferences for auditory information. For the Catalan-dominant group, the significant intercept indicates that the probability of choosing the auditory signal in incongruent stimuli was above chance. By contrast, Spanish-dominant participants had significantly lower probabilities of choosing the auditory alternative. In short, the group whose dominant language matched the non-native language showed a clear preference for auditory information.

4. DISCUSSION

Based on the statistical analyses, this section addresses the three research questions posed in the study.

4.1. RQ1 – Language dominance

The first question asked whether language dominance (Catalan vs. Spanish) affects the assimilation of Brazilian Portuguese (BP) vowels into Catalan categories. The results confirm this prediction: Spanish-dominant participants were significantly less accurate than Catalan-dominant participants. These results provide additional evidence for Tyler and Best's (2024) assumption that the principles of PAM and PAM-L2 can be extended to cross-linguistic studies of early bilinguals. Moreover, our findings are consistent with previous research on the perception of Catalan vowels, which used both auditory stimuli (Pallier *et al*., 1997) and audiovisual or other modalities (Navarra & Soto-Faraco, 2007) and showed that Spanish-dominant bilinguals struggle to distinguish Catalan vowel contrasts.

This pattern can be explained by the vowel systems of the two groups. Catalan-dominant bilinguals are expected to have well-established categories for both mid-high (/e/, /o/) and mid-low (/ɛ/, /ɔ/) vowels. Even when categories are shared between Catalan and Spanish, Catalan dominance likely shapes them toward Catalan-like acoustic properties, facilitating assimilation of BP vowels into their Catalan equivalents. By contrast, Spanish-dominant bilinguals lack stable categories for the mid-low vowels, since these are absent from Spanish. Even for the shared categories /e/ and /o/, their representations

are closer to Spanish than to Catalan, which accounts for their greater difficulty with assimilation.

However, Spanish-dominant participants still achieved accuracy above 60% in the AO and AV+ conditions. This indicates that bilinguals can assimilate unfamiliar sounds into their non-dominant language system. From a developmental perspective, this suggests that Spanish-dominant bilinguals might later acquire BP vowels more successfully than Spanish monolinguals, who initially assimilate both mid-high and mid-low vowels into the Spanish /e/ and /o/ categories. Future studies comparing Spanish monolinguals and Spanish-dominant bilinguals will be essential to test this assumption. More broadly, the findings highlight the need in L2 acquisition research to consider the entire linguistic repertoire of bilingual and multilingual learners, including less-dominant languages that may nonetheless shape assimilation patterns.

4.2. RQ2 – Stimulus modality

The second question examined whether assimilation rates differed across modalities (AO, VO, AV+) and whether these differences varied with language dominance. It was expected that Spanish-dominant participants, given their weaker Catalan categories, would benefit from adding visual cues (AV+). Contrary to this prediction, no significant differences emerged between AO and AV+, and no interactions were observed between Group and Modality.

Overall, both groups showed the same pattern: comparable accuracy between AO and AV+, but reduced performance in VO. Regarding VO stimuli, this modality provides less information to aid individuals in recovering phonemic information (Bernstein *et al*., 2000). Additionally, from the perspective of PAM-L2, which treats articulatory gestures as multimodal phonological primitives, this result is not surprising. In turn, visual-only information proved less informative. This may be because VO input is less ecological in everyday life, where auditory-only input is common, but purely visual access arises only in exceptional circumstances (e.g., background noise, muted devices). Because such situations are rare, VO input is expected to be less robust, a finding that also supports our claim that listeners tend to prioritize auditory information, shifting attention to visual cues only as a secondary resource.

The lack of significant differences between AO and AV+ in the Spanish-dominant group warrants further discussion. First, although previous studies on the perception of Catalan vowels by Spanish-dominant early bilinguals suggest an advantage of AV+ over AO (cf. Navarra & Soto-Faraco, 2007), they did not address non-native language assimilation, which may have influenced the results. One possible explanation is that, even for Spanish-dominant participants, daily exposure to Catalan may have been sufficient to support accuracy rates above 50% in these conditions. These results raise the possibility that, for listeners with considerable experience or exposure to the target language, audio input may be informative enough for participants to assimilate BP into their more- or less-dominant language. It is also possible that AV+ stimuli could be more informative when assimilation rates are lower (e.g., for participants with less exposure to Catalan), as the possibility of an accuracy threshold beyond which AV+ stimuli are not significantly more informative than AO cannot be ruled out.

Regarding the similar accuracy rates found in the AO and AV+ conditions by the Spanish-dominant group in audiovisual speech processing, Campbell (2008) distinguishes between two modes: the correlated mode and the complementary mode. In the former, visual information duplicates information about dynamic articulatory patterning. In the latter, visual information provides additional, novel information for aspects of the speech event that are difficult to hear with auditory information alone (such as the distinction between /m/ and /n/ or /θ/ and /s/, cf. Summerfield, 1987). As this study focuses on vowels, which have well-defined acoustic regions (informative enough for participants to extract articulatory information) and clear distinctions in their visual articulatory information, we consider this a case of correlated-mode processing. The redundant status of video information not only accounts for the equivalent rates between the AO and AV conditions but may also lead to a "relative (over-) reliance on acoustic parameters of the speech stream" (Campbell, 2008: 1004), which accounts for the lower results found in the VO condition.

4.3. RQ3 – Incongruent stimuli

The third question examined whether preferences for auditory versus visual information in incongruent (AV–) stimuli varied by language dominance. Results showed that Catalan-dominant participants almost exclusively preferred auditory cues, whereas Spanish-dominant participants relied more evenly on both modalities, with a preference for audio stimuli.

The strong preference for audio stimuli in the Catalan-dominant group, which has more established vowel categories in both Spanish and Catalan, suggests that more proficient users of a language tend to favor auditory information when it is sufficiently informative (i.e., when noise or other factors do not distort perception). This supports Campbell's (2008) claim discussed earlier: in this context, visual information serves only a redundant role. From a direct-realist account of speech perception, as outlined in PAM, we can infer that auditory information is sufficiently informative for participants to recover articulatory details from the acoustics.

For the Spanish-dominant group, their more balanced outcome may align with our prediction that listeners rely on visual cues when auditory information is insufficient. For Spanish-dominant participants, weaker or less-stable Catalan categories (/e/ and /o/ resembling Spanish, and incomplete /ɛ/ and /ɔ/) led to an attentional shift toward visual articulatory information. In Campbell's (2008) terms, their processing cannot be considered an exemplar of the correlated mode, since the auditory and visual streams were not redundant. Although the information in the video stimuli was not actually complementary to that in the audio, participants in the Spanish-dominant group treated the visual information as complementary, given the audio's low informativeness relative to their not-yet-developed categories. In other words, while an advantage of natural AV+ stimuli over AO was not found in natural stimuli, when confronted with incongruent, manipulated, or "unnatural" information, Spanish-dominant speakers were forced to choose between two different phonological categories.

Still, for the Spanish-dominant group, the above explanation holds only if we assume the assimilation rates are accurate (i.e., participants in this group correctly identify the BP vowel's height when assimilating it to their Catalan counterpart). However, an alternative explanation is also possible: these results may indicate that their Catalan vowel categories

are not well established, which would make it difficult to identify vowel height. In other words, we should ask whether these results truly reflect a preference for visual stimuli in the Spanish-dominant group or simply errors due to an inability to identify vowel distinctions accurately. This latter possibility warrants consideration, as the audio-stimuli preference rate in this group (58.16%) in the AV condition is close to the accuracy rate for the AO modality (61.54%). Given these similar rates, one possible explanation is that participants prioritize auditory over visual information, and their supposed preference for visual stimuli may instead reflect inaccurate answers when they attend to auditory information. In this alternative scenario, auditory stimuli remain the primary source of information when responding to incongruent stimuli, as both groups would respond similarly.

The two explanatory scenarios outlined above suggest a preference for auditory stimuli in both groups (and, in the second scenario, very similar behavior between the groups). This preference for one type of information (audio) over the other (video) can be explained by the direct-realist approach. Unlike enrichment models such as FLMP or Bayesian frameworks, which posit that perceivers construct representations or inferences beyond the signal (Massaro, 1998, 2004; Knill & Pouget, 2004), the direct-realist approach underlying PAM assumes that perception involves attunement to the information available across energy arrays (E. J. Gibson & Pick, 2000; Alves *et al*., 2025). Under this account, incongruent stimuli in a forced-choice task require listeners to select between two distinct categories, and the stability of their phonological system determines the extent to which they rely on visual versus auditory cues.

4.4.Broader implications

Together, these findings extend cross-linguistic perception research by showing that language dominance shapes assimilation patterns not only in auditory and audiovisual conditions but also in how listeners resolve conflicting cues. They also highlight the explanatory power of PAM and PAM-L2 in accounting for modality effects and bilingual perception, while pointing to promising avenues for future research on multimodal speech processing and bilingual category formation.

5. CONCLUSION

In this study, we investigated how Spanish–Catalan bilinguals assimilated Brazilian Portuguese mid vowels into Catalan categories, with participants divided into two groups based on language dominance: Catalan-dominant and Spanish-dominant bilinguals. Because low-mid vowel categories were expected to be less established among Spanish-dominant participants, we predicted higher accuracy for the Catalan-dominant group across AO, VO, and AV+ stimuli. We also hypothesized that Spanish-dominant participants would show higher accuracy in AV+ than in AO, reflecting a stronger reliance on visual cues, particularly under incongruent conditions.

The results confirmed higher accuracy for the Catalan-dominant group, supporting the view that their Catalan vowel categories are more firmly established. However, we observed no significant differences by modality between AO and AV+, even among Spanish-dominant participants. This suggests that auditory information alone was sufficient for vowel assimilation, consistent with the PAM assumption that listeners recover articulatory information from acoustic cues. In contrast, the VO condition yielded

lower accuracy for both groups, indicating that visual input alone is less informative, regardless of the stability of Catalan categories.

A different pattern emerged with incongruent (AV–) stimuli. Catalan-dominant participants strongly favored auditory information, whereas Spanish-dominant participants, while still preferring auditory information, showed a more balanced reliance on auditory and visual cues. This finding may align with our assumption that listeners shift attention to visual input when auditory information is insufficient — a tendency more likely among Spanish-dominant participants, whose categories for /e/ and /o/ resemble Spanish more closely and whose categories for /ɛ/ and /ɔ/ are less well defined. In such cases, listeners may fall back on lower-order articulatory information provided by visual cues. Alternatively, both groups might respond to incongruent stimuli in the same way if participants are primarily focused on auditory stimuli and the supposed preference for visual information reflects error rates when attending to audio stimuli. Regardless of the scenario, it is clear that both groups prioritize auditory information as the primary input source.

Methodological considerations also warrant discussion. First, task instructions were provided in participants' dominant language. While this ensured clarity, for Spanish-dominant participants it may also have activated Spanish more strongly, consistent with research on bilingual language mode (Antoniou *et al*., 2012, 2013). Since bilinguals' languages are always co-activated (Kroll, 2008; Kroll *et al*., 2022), this may have influenced attention to higher- and lower-order invariants, potentially involving Spanish categories. Future studies should test this possibility by ensuring instructions are equally clear in both languages while minimizing unintended activation effects.

Second, the stimuli, which focused on isolated vowels, may also have affected the results. Given that lexical development supports attunement to phonological invariants (Best *et al*., 2009; Best, 2015; Bundgaard-Nielsen *et al*., 2011, 2012), it remains an open question whether AO input would have been as informative as AV+ if lexical items or pseudowords had been used. Future research should examine how lexicality influences modality effects and extend investigations into naturalistic linguistic, pragmatic, and social contexts (Matyjkek *et al*., 2024).

The study also faced limitations in participant recruitment. Strict criteria to ensure clear dominance in one language limited the sample size and may have reduced statistical power. This challenge is particularly acute in Catalonia, where exposure to both Spanish and Catalan is widespread and nearly unavoidable, making it difficult to identify participants with strict dominance, especially among younger populations. Nevertheless, the study identified robust group effects in accuracy rates and auditory preference under incongruent conditions, as well as a clear advantage of AO and AV+ over VO.

Future research should further test the assimilation conditions predicted by PAM, following Tyler and Best's (2024) recommendations. This includes analyzing assimilation in both the dominant and non-dominant languages of bilinguals, as well as monolingual control groups. Such work would help clarify how category formation unfolds within a shared phonological space.

Despite its limitations, the present study contributes to ongoing discussions of bilingual perception and L2 acquisition. A promising line of research compares Spanish

monolinguals and Spanish-dominant bilinguals in the L2 acquisition of BP vowels. Based on the assimilation rates observed here, these groups are expected to differ in accuracy even at comparable proficiency levels. This highlights the importance of mapping all languages in a bilingual/multilingual learner's repertoire, since each may affect vowel assimilation and the formation of new categories (Flege & Bohn, 2021).

Although no AV+ advantage emerged in the bilingual groups examined, the study also opens avenues for investigating the role of audiovisual input in additional language learning. For instance, comparing L1 Spanish learners of BP at basic and advanced proficiency levels may reveal whether AV+ stimuli facilitate vowel assimilation more effectively at earlier stages. This issue, which emerged only under incongruent conditions in this paper, has important pedagogical implications for L2 pronunciation instruction and warrants further exploration. Future studies should also examine exposure in acoustically noisy contexts, as such conditions may favor the AV+ pattern over AO.

Taken together, the results of this study help bridge the gap between cross-linguistic and multimodal research and illuminate how early bilinguals assimilate unfamiliar sounds. By integrating perspectives on cross-linguistic perception, multimodal processing, and bilingualism, this work advances our understanding of the complex processes underlying multimodal speech perception across languages.

ACKNOWLEDGEMENTS

We thank Prof. Dr. Salvador Soto-Faraco (Universitat Pompeu Fabra, Spain) for hosting our visit and facilitating data collection at the institution. The first author also acknowledges financial support from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil) for funding this project in Barcelona (grant no. 402710/2024-0).

References

Alves, Ubiratã Kickhöfel, Larissa Cristina Berti & Felipe Santos dos Reis. 2025. L2 speech perception models: A theoretical discussion on the incorporation of visual information. *Gradus – Revista Brasileira de Fonologia de Laboratório* 10(1). 75–102.

Amengual, Mark. 2016. The perception of language-specific phonetic categories does not guarantee accurate phonological representations in the lexicon of early bilinguals. *Applied Psycholinguistics* 37(5). 395–401.

Antoniou, Mark, Catherine T. Best & Michael D. Tyler. 2013. Focusing the lens of language experience: Perception of Ma'di stops by Greek and English bilinguals and monolinguals. *Journal of the Acoustical Society of America* 133(4). 2397–2411.

Antoniou, Mark, Michael D. Tyler & Catherine T. Best. 2012. Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics* 40(4). 582–594.

Ayuso, María Machuco & Dolors Poch Olivé. 2016. Dinámica de las vocales del español en contacto con el catalán. *Oralia* 19. 153–175.

Barrós-Loscertales, Alfonso, Noelia Ventura-Campos, Maya Visser, Agnès Alsius, Christophe Pallier, César Ávila Rivera & Salvador Soto-Faraco. 2013. Neural correlates of audiovisual speech processing in a second language. *Brain & Language* 126. 253–262.

Bernstein, Lynne E., Paula E. Tucker & Marilyn E. Demorest. 2000. Speech perception without hearing. *Perception & Psychophysics* 62(2). 233–252.

Best, Catherine T. 1995. A direct-realist view of cross-language speech perception. In Winifred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. 171–204. Baltimore: York Press.

Best, Catherine T. 2015. Devil or angel in the details? Perceiving phonetic variation as information about phonological structure. In Joaquín Romero & María Riera (eds.), *Phonetics–phonology interface: Representations and methodologies*. 3–31. Amsterdam: John Benjamins.

Best, Catherine T., Gerald W. McRoberts & Elizabeth Goodell. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America* 109(2). 775–794.

Best, Catherine T. & Michael D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In Ocke-Schwen Bohn & Murray J. Munro (eds.), *Language experience in second language speech learning: In honor of James Emil Flege*. 13–34. Amsterdam: John Benjamins.

Best, Catherine T., Michael D. Tyler, Tiffany N. Gooding, Corey B. Orlando & Chelsea A. Quann. 2009. Development of phonological constancy: Toddlers' perception of native- and Jamaican-accented words. *Psychological Science* 20(5). 539–542.

Bohn, Ocke-Schwen. 2018. Cross-language and second language speech perception. In Eva M. Fernández & Helen Smith Cairns (eds.), *The handbook of psycholinguistics*. 213–239. Hoboken, NJ: Wiley.

Browman, Catherine & Louis Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3. 219–252.

Browman, Catherine & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49. 155–180.

Bundgaard-Nielsen, Rikke L., Catherine T. Best, Christian Kroos & Michael D. Tyler. 2012. Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility. *Applied Psycholinguistics* 33(3). 643–664.

Bundgaard-Nielsen, Rikke L., Catherine T. Best & Michael D. Tyler. 2011. Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition* 33(3). 433–461.

Campbell, Ruth. 2008. The processing of audiovisual speech: Empirical and neural bases. *Philosophical Transactions of the Royal Society B* 363. 1001–1010.

Chen, Yuchun & Valerie Hazan. 2009. Developmental factors and the non-native speaker effect in audiovisual speech perception. *Journal of the Acoustical Society of America* 126(2). 858–865.

Ernst, Marc O. & Heinrich H. Bülthoff. 2004. Merging the senses into a robust percept. *Trends in Cognitive Sciences* 8(4). 162–169.

Escudero, Paola, Paul Boersma, Andréia Schurt Rauber & Ricardo A. H. Bion. 2009. A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *Journal of the Acoustical Society of America* 126(3). 1379–1393.

Faris, Mona M., Catherine T. Best & Michael D. Tyler. 2016. An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *Journal of the Acoustical Society of America* 139(1). EL1–EL5.

Flege, James Emil & Ocke-Schwen Bohn. 2021. The revised Speech Learning Model. In Richard Wayland (ed.), *Second language speech learning: Theoretical and empirical progress*. 3–83. Cambridge: Cambridge University Press.

Fowler, Carol A. 1986. An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics* 14. 3–28.

Fowler, Carol A. 1996. Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America* 99(3). 1730–1741.

Gibson, Eleanor J. & Anne D. Pick. 2000. *An ecological approach to perceptual learning and development*. Oxford: Oxford University Press.

Goldstein, Louis & Carol A. Fowler. 2003. Articulatory phonology: A phonology for public language use. In Niels O. Schiller & Antje S. Meyer (eds.), *Phonetics and phonology in language comprehension and production*. 159–207. Berlin: Mouton de Gruyter.

Hardison, Debra M. 2003. Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics* 24(4). 495–522.

Hardison, Debra M. 2005. Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Applied Psycholinguistics* 26(4). 579–596.

Hazan, Valerie, Jeesun Kim & Yuchun Chen. 2010. Audiovisual perception in adverse conditions: Language, speaker and listener effects. *Speech Communication* 52(11–12). 996–1009.

Hazan, Valerie & Anke Sennema. 2007. The effect of visual training on the perception of non-native phonetic contrasts. In *Proceedings of the 16th International Congress of Phonetic Sciences*. 1585–1588.

Hazan, Valerie, Anke Sennema, Midori Iba & Andrew Faulkner. 2005. Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication* 47(3). 360–378.

Hazan, Valerie, Anke Sennema & Andrew Faulkner. 2002. Audiovisual perception in L2 learners. In *Proceedings of the 7th International Conference on Spoken Language Processing*. 1685–1688.

Hazan, Valerie, Anke Sennema, Andrew Faulkner, Marta Ortega-Llebaria, Midori Iba & Hyunsong Chung. 2006. The use of visual cues in the perception of non-native consonant contrasts. *Journal of the Acoustical Society of America* 119(3). 1740–1751.

Hirata, Yukari & Spencer D. Kelly. 2010. Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research* 53(2). 298–310.

Iba, Midori, Anke Sennema, Valerie Hazan & Andrew Faulkner. 2004. Use of visual cues in the perception of a labial/labiodental contrast by Spanish- and Japanese-L1 learners of English. In *Proceedings of the 8th International Conference on Spoken Language Processing*. 2497–2500.

Knill, David C. & Alexandre Pouget. 2004. The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences* 27(12). 713–719.

Kroll, Judith F. 2008. Juggling two languages in one mind. *Psychological Science Agenda* 22(1). 1–6.

Kroll, Judith F., Kinsey Bice, Mona Roxana Botezatu & Megan Zirnstein. 2022. On the dynamics of lexical access in two or more languages. In Anna Papafragou, John C. Trueswell & Lila R. Gleitman (eds.), *The Oxford handbook of the mental lexicon*. 583–597. Oxford: Oxford University Press.

Li, Ying. 2015. *Audio-visual training effect on L2 perception and production of English /θ/-/s/ and /ð/-/z/ by Mandarin speakers*. Doctoral dissertation, Newcastle University.

Li, Yamping, Michael D. Tyler, Denis Burnham & Catherine T. Best. 2023. Categorization and discrimination of Mandarin lexical tones by naïve English listeners. *Journal of the Acoustical Society of America* 154(4). A157.

Marian, Viorica, Sayuri Hayakawa, Tuan Q. Lam & Scott R. Schroeder. 2018. Language experience changes audiovisual perception. *Brain Sciences* 8(85). 1–14.

Massaro, Dominic W. 1998. *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.

Massaro, Dominic W. 2004. From multisensory integration to talking heads and language learning. In Gemma A. Calvert, Charles Spence & Barry E. Stein (eds.), *The handbook of multisensory processes*. 153–176. Cambridge, MA: MIT Press.

McGurk, Harry & John MacDonald. 1976. Hearing lips and seeing voices. *Nature* 264. 746–748.

Mattyjek, Magdalena, Sotaro Kita, Mireia Torralba Cuello & Salvador Soto-Faraco. 2024. Multisensory integration of speech and gestures in a naturalistic paradigm. *Human Brain Mapping* 45(11). e26797.

Melguy, Yevgeniy V. 2018. Exploring the bilingual phonological space: Early bilinguals' discrimination of coronal stop contrasts. *Language and Speech* 61(2). 173–198.

Navarra, Jordi & Salvador Soto-Faraco. 2007. Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychological Research* 71. 4–12.

Pallier, Christophe, Laura Bosch & Núria Sebastián-Gallés. 1997. A limit on behavioral plasticity in speech perception. *Cognition* 64. B9–B17.

R Core Team. 2025. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/

Recasens, Daniel. 2014. *Fonètica i fonologia experimentals del català: Vocals i consonants*. Barcelona: Institut d'Estudis Catalans.

Recasens, Daniel & Aina Espinosa. 2006. Dispersion and variability in Catalan vowels. *Speech Communication* 48. 645–666.

Reisberg, Daniel, John McLean & Anna Goldfield. 1987. Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In Barbara Dodd & Ruth Campbell (eds.), *Hearing by eye: The psychology of lip-reading*. 97–113. London: Erlbaum.

Reyes, Yasna Pereira & Valerie Hazan. 2021. English vowel perception by non-native speakers: Impact of audio and visual training modalities. *Onomázein* 51. 111–136.

Rosenblum, Lawrence D. 2005. The primacy of multimodal speech perception. In David B. Pisoni & Robert E. Remez (eds.), *Handbook of speech perception*. 51–78. Oxford: Blackwell.

Rosenblum, Lawrence D. 2008. Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science* 17(6). 405–409.

Sekiyama, Kaoru & Denis Burnham. 2008. Impact of language on development of audiovisual speech perception. *Developmental Science* 11(2). 306–320.

Springer, Helene. 2021. *L2 listeners benefit from audiovisual information in the processing of Swedish vowels*. MA thesis, Lund University.

Summerfield, Quentin. 1987. Some preliminaries to a theory of audiovisual speech processing. In Barbara Dodd & Ruth Campbell (eds.), *Hearing by eye*. 58–82. London: Erlbaum.

Tyler, Michael D. & Catherine T. Best. 2024. The Perceptual Assimilation Model: Early bilingual adults and developmental foundations. In Mark Amengual (ed.), *The Cambridge handbook of bilingual phonetics and phonology*. 147–172. Cambridge: Cambridge University Press.

Tyler, Michael D., Catherine T. Best, Louis M. Goldstein & Mark Antoniou. 2014. Investigating the role of articulatory organs and perceptual assimilation in infants' discrimination of native and non-native fricative place contrasts. *Developmental Psychology* 56. 210–227.

Wang, Yue, Dawn M. Behne & Haisheng Jiang. 2008. Linguistic experience and audiovisual perception of non-native fricatives. *Journal of the Acoustical Society of America* 124(3). 1716–1726.