DOI: 10.1111/infa.12486

INFANCY THE OFFICIAL JOURNAL OF THE INTERNATIONAL CONGRESS OF INFANT STUDIES WILEY

RESEARCH ARTICLE

Before perceptual narrowing: The emergence of the native sounds of language

Konstantina Zacharaki | **Nuria Sebastian-Galles**

Center for Brain and Cognition (CBC), Universitat Pompeu Fabra, Barcelona, Spain

Correspondence
Konstantina Zacharaki, Center for Brain and Cognition (CBC), Universitat Pompeu Fabra, 08005 Barcelona, Spain.
Email: konstantina.zchrk@gmail.com

Funding information
Institució Catalana de Recerca i Estudis Avançats, Grant/Award Number: Academia 2019 award; Catalan Government, Grant/Award Number: FI2018, SGR 2017-268, FI-9015-456763; Spanish Ministerio de Ciencia, Innovación y Universidades, Grant/Award Number: PGC2018-101831-B-I00

Abstract

The present study investigates the precursors of representations of phonemes in 4.5-month-olds. The emergence of phonemes has been mainly studied within the framework of perceptual narrowing, that is, infants tuning to their native language and losing sensitivity to non-native speech. One of the mechanisms behind this phenomenon is distributional learning. In this article, we tested the preference of 4.5-month-old infants using lists of pseudowords that resemble the vowel distribution of the native or a non-native language. We found that infants prefer listening to the lists mirroring the native language. The results suggest that infants can extract vowel information from novel stimuli, and they can map it on pre-existing knowledge on vowels that leads to a preference for the native lists.

1 | INTRODUCTION

One fundamental milestone in early language acquisition is the establishment of the phoneme categories of the language of exposure. Initially newborns are able to discriminate virtually all phoneme contrasts existing in different languages of the world, even if never exposed to them or their caregivers cannot produce them (Cheour-Luhtanen et al., 1995; Eimas et al., 1971; Trehub, 1976; for reviews see Aslin et al., 1998; Werker & Gervain, 2013). This initial capacity starts to decline around 6 months of age, when infants' perception of the native categories improves and perception of the non-native ones declines: the so-called perceptual narrowing. Perceptual narrowing has been extensively reported, taking place first for lexical stress and tone (Yeung et al., 2013), then for vowels (Kuhl et al., 1992; Polka & Werker, 1994; for meta-analysis see Tsuji & Cristia, 2013) and later for consonants (Werker & Tees, 1984). Perceptual narrowing of vowels has been reported for Spanish-learning infants between 4.5 and 8 months of age. Bosch and Sebastián-Gallés (2003) showed that 4.5 Spanish-learning infants can discriminate the Catalan-specific /e/ – /ɛ/ vowel contrast, that Spanish native adults, as well as 8-month-old Spanish learning infants no longer discriminate (Bosch et al., 2000; Pallier et al., 1997; Sebastián-Gallés & Soto-Faraco, 1999 among others). In contrast, Catalan natives (at 4.5 and 8 months of age) have no difficulties perceiving this contrast.

There is a generalized consensus in that infants' sensitivity to the distributional properties of the speech sounds is critical in the perceptual narrowing. The emergence of phonetic categories would result from infants' computation of the distributional properties of the speech sounds. In a seminal article, Maye et al. (2002) found that infants as early as 6 months can extract distributional regularities quickly when presented with sounds from a /da/ – /ta/ continuum. In the familiarization, infants were presented with a unimodal or a bimodal distribution, based on the frequency of each token in the continuum, resulting in one or two frequency peaks. In the test, infants only discriminated the stimuli when they had been previously familiarized with a bimodal distribution. The earliest evidence of sensitivity to distributional learning has been found with 2- to 3-month-old Dutch infants with vowels (Wanrooij et al., 2014). Infants were trained with either a unimodal or a bimodal /ae/–/e/ distribution of the English vowel contrast and tested in an oddball paradigm using EEG. The participants showed a greater mismatch response after the bimodal exposure. The relevance of distributional learning has received extensive support not only from behavioral studies, but also neurophysiological data as well as computational modeling (Adriaans & Swingley, 2012; McMurray et al., 2009; Schatz et al., 2021; Vallabha et al., 2007, among others).

The critical observation supporting the existence of perceptual narrowing is that at an early age infants show discrimination abilities of a certain phonetic contrast not present in the infants' environment and that at a later age (a few months later) such discrimination abilities are no longer observed. The fact that some phonetic contrasts show a gradual decline at a group level has been attributed primarily to the existence of individual differences, with some infants showing earlier (or later) narrowing than others. The frequency of appearance of a phoneme in the native language also plays a role in the shape of perceptual narrowing (Anderson et al., 2003). For instance, infants become worse at discriminating a coronal, non-native pair of sounds [such as /ʈɜ/–/(ʈɜ) for English-speaking infants] than a dorsal one (such as /k'ɜ/–/q'ɜ/), due to the former being more frequent than the latter (Tobias, 1959). Therefore, it can be assumed that phoneme categories emerge as the consequence of computations over a protracted period of time and a considerable amount of input. If so, it is reasonable to suppose that before perceptual narrowing effects can be observed, infants possess some information about the distributional properties of the sounds of their language as they have been collecting such information on them.

Different studies have shown that infants prefer listening to lists of words that follow the properties of their native language. Jusczyk, Friederici et al. (1993) showed that 9-month-olds, but not 6 month-olds, prefer listening to lists of words that contain phonemes or combinations of phonemes only found in their native language (Experiment 1, Experiment 3), or conforming only to the phonotactics of the native language (Experiment 4) using phonemes found in both languages, that is, English and Dutch (see also Friederici & Wessels, 1993; Jusczyk, Cutler et al., 1993; Jusczyk et al., 1994; Sebastián-Gallés & Bosch, 2002 among others for similar patterns of preference). If before perceptual narrowing, infants possess some distributional knowledge of the sounds of their language of exposure, we expect infants to prefer listening to lists conforming native distributional properties over non-native ones.

To this end, we capitalize on the significant differences in the distribution of vowels in Catalan and Spanish. Catalan and Spanish[1] are two Romance languages that present quite different distributions of vocalic sounds. Spanish has a 5-vowel system, whereas Catalan has 8 vowels (Carbonell & Llisterri, 1992; Martínez-Celdrán et al., 2003). The main phonological distinction between the two is that Catalan has vowel reduction; meaning that in unstressed positions /e/, /ɛ/, /a/ are reduced to a /ə/ and /o/ and /ɔ/ are reduced to a /u/, whereas /i/ and /u/ can be found in both stressed and unstressed positions. The use of vowel reduction only in Catalan results in conspicuous differences in the vowel distribution of the two languages. Because mid vowels can only occur in stressed syllables, they are relatively infrequent: only 16% of all vowels in Catalan are mid vowels (e–ɛ, o–ɔ), while they are almost three times more frequent in Spanish (49%). Conversely, central-low vowels are twice more frequent in Catalan (61%) than in Spanish (29%). Figure 1 shows the specific frequency of appearance of each vowel (see Zacharaki & Sebastian-Galles, 2021, for a more complete description of the phonological properties of these two languages, see also Albareda-Castellot et al., 2011). If 4- to 5-month-old infants are sensitive to the distributional properties of the vowel sounds of their language of exposure, we expect to find a preference for lists of words instantiating the native pattern.

2 | MATERIALS AND METHODS

The method, data analysis, and criteria for exclusion of participants were pre-registered on the OSF (Open Science Framework) database before analyzing the data (Zacharaki & Sebastian-Galles, 2020, https://doi.org/10.17605/OSF.IO/2VR85).

2.1 | Participants

Forty-eight 4.5-month-old monolingual infants participated in this experiment. All participants were healthy and full-term (>37 gestation weeks) coming from either a Catalan (n = 23, range = 123–169 days, M = 148, SD = 15, 12 female, M = 98% exposure to Catalan, SD = 3.8%) or a Spanish (n = 25, range = 121–166 days, M = 141, SD = 13, 13 female, M = 97% exposure to Spanish, SD = 5.2%) speaking family. The linguistic profile was calculated using an adapted version of a language questionnaire (Bosch & Sebastian-Galles, 2001). If participants had less than 85% exposure to the dominant language, they were excluded from the final sample. Participants were recruited from private hospitals found in Barcelona (Spain) or online. Twenty-five additional infants were tested but not included in the final sample for the following reasons: not enough data (n = 5), fussy/cried (n = 5), failure to calibrate (n = 3), bilinguals (n = 3), low weight at birth (n = 3), technical error (n = 3), outliers (n = 2),[2] health issues (n = 1). The present research was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the clinical research ethics committee at the Parc de la Salut Marc.

[1] Unless specified, when Catalan is mentioned, we refer to the Eastern dialect, spoken in the metropolitan area of Barcelona. As for Spanish, we always refer to the Standard Castilian one.

[2] Based on the suggestion of one of the reviewers and the editor, we have eliminated 2 participants from the sample due to their mean Looking time being more than two 2SD above the mean in the non-native condition. The analysis including these two participants can be found in Annex 1. The analysis including the two outliers yielded the same pattern of significances.

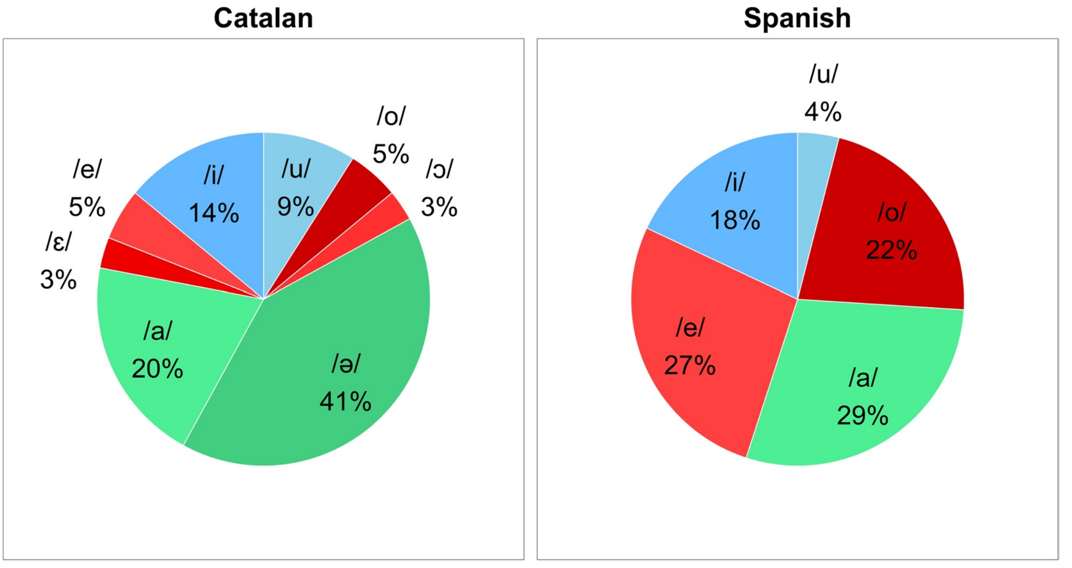

FIGURE 1 Distribution (in percentage) of Catalan and Spanish vowels. The height of the tongue during articulation (F1) is color coded. High vowels appear in blue, mid vowels in red, and low vowels in green. The placement of the vowels in the pie charts approaches the one in the vowel space (F1/F2)

TABLE 1 Acoustic properties of the lists and percentage of each vowel category (F1), standard deviation is inside the parentheses

	Pitch			Duration	% Vowels		
	Mean (Hz)	Min	Max	Mean	Low	Mid	High
Catalan	262 (16.9)	221	301	0.84 (0.1)	61	16	23
Spanish	265 (16)	222	304	0.86 (0.1)	29	49	22

2.2 | Stimuli

The stimuli were 240 nonsense words which were organized in 10 lists for each language Catalan and Spanish respectively, that is, 20 lists in total. Ten of the words in each list were disyllabic and two were trisyllabic. A female bilingual Catalan-Spanish speaker, highly proficient in the two languages, was recorded. The speaker is a musician, and she was taught phonetics in school, which enabled her to articulate the words fittingly. The speaker had no detectable accent in either language. The stimuli were recorded in a soundproof room in the 'Laboratori de Recerca en Infancia' at the University Pompeu Fabra using an Audio-Tecnica microphone (AT2050) at a sampling rate of 44,100 Hz. We used Audacity® recording and editing software (Mazzoni, 1999) to record the stimuli and Praat (Boersma & Weenink, 2019) to extract the words from the recording and analyze their pitch and frequency. We made sure that our stimuli were not different in terms of pitch nor intensity (see Table 1). The speaker was instructed to say the words in an infant-directed manner and always to stress the penultimate syllable in each word.

Each list mimicked the vowel distribution of either Catalan (Rafel, 1980) or Spanish (Alcina & Blecua, 1975). For instance, as shown in Table 1, 61% of the vowels in the Catalan list occupy a central-low position in the vowel space, while only 29% occupy such position in Spanish. The same **'CV**CV/CV**'CV**CV structure was used in the two languages to ensure that the lists differed only in terms of their vowels (both structures are very common in Catalan and Spanish). The consonantal structure across the two languages was the same, the only difference was the frequency and placement of the vowels (Table 2). We used Wuggy (Keuleers & Brysbaert, 2010), a multilingual pseudoword generator, to create phonotactically legal structures in Spanish. The same consonantal structure was used across the two languages, and all the words were also phonotactically permissible in Catalan. The speaker was instructed to use only the five common vowels of the two languages /a, e, i, o, u/ for all of the lists and avoid the Catalan vowels /ɛ, ə, ɔ/, so as to eliminate providing additional linguistic cues to the participants. The order of the words within each list was randomized across participants. After each word, a silence of different duration was added to reach 2 s in total for each pseudoword, that is, the duration of the word plus the amount of silence needed.

2.3 | Procedure

Infants were tested in a sound-attenuated, dimly lit room in the 'Laboratori de Recerca en Infancia' at the University Pompeu Fabra. The testing room was equipped with an ASUS VE276N (size: 27″, resolution: 1920 × 1080) and a Tobii Pro Spectrum eyetracker (120 Hz sampling rate). Two M-Audio AV 30 loudspeakers were placed left and right from the monitor, hidden behind a beige curtain. A SONY HC9 camera was placed 20 cm above the monitor allowing the experimenter to check on the infant while testing. The participants were seated on a baby chair facing the monitor at a 65 cm distance. The baby chair was adaptable, and height and distance were adjusted for each participant depending on their size. The caregiver was seated in a chair behind the participant and was instructed to avoid any contact with the infant and to look towards the floor or to have their eyes closed. The experimenter controlled the experiment outside the testing room using a custom-made script on PsychoPy (Peirce et al., 2019).

We adapted the procedure employed in Jusczyk, Cutler et al. (1993), which was originally used to test infants' preference towards their predominant stress pattern. We used an eye tracker to make the procedure completely infant-controlled. Areas of interest (AOIs) were defined as enlarged squares around the visual stimuli, 600 × 600 pixels in size (see Figure 2b,c). The size of the AOI was 17% in

TABLE 2 Examples of an experimental list

Spanish	Catalan
/ˈbane/	/ˈbuna/
/geˈrota/	/giˈrota/
/ˈgiðo/	/ˈguða/
/ˈgira/	/ˈgari/
/ˈlape/	/ˈlipa/
/loˈtiɣe/	/laˈteɣa/
/ˈmaβo/	/ˈmaβa/
/ˈneβa/	/ˈnaβa/
/ˈnoɲe/	/ˈnoɲa/
/ˈpoβa/	/ˈpiβa/
/ˈreði/	/ˈraða/
/ˈtuki/	/ˈtaka/

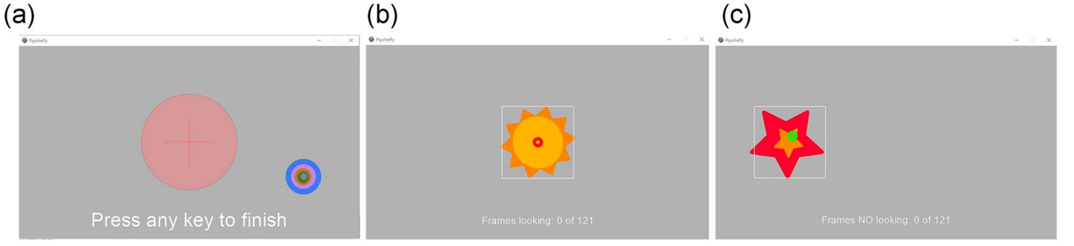

FIGURE 2 (a) Circle shown before calibration as feedback for the participant's position, (b) attention getter shown at the beginning of each trial, (c) visual attractor while listening to the lists of words. Only the colored pictures were shown inside the testing room, the white squares were only visible on the experimenter's monitor

relation to the whole screen size. After the infant was seated in the baby chair, a circle with a cross in the middle and a random moving ball appeared on the screen. Once the participant looked at the screen, another circle appeared that had to be green and to overlap with the preexisting circle. If the participant was not placed properly, too far from the screen or lower/higher than it should, the circle was red. In this manner, we got immediate feedback on the positioning of the participant and corrected when needed before performing the calibration (see Figure 2a). When the participant was placed suitably, a 5-point calibration was performed. If the participant reached the minimum of 3 valid points of calibration, a rotating sun appeared in the middle of the screen accompanied by a recurring ping-like sound. Once the participants had looked at the sun for two accumulative seconds, it disappeared, and a rotating star appeared on either side of the screen and one of the lists was played simultaneously. Each trial consisted of the sun (attention-getter) and star (visual attractor) sequence. Infants completed a practice phase with four lists (two in each language), followed by a testing phase with a maximum of 16 trials (eight in Catalan and eight in Spanish). The auditory stimuli were presented at 65 dBs. The ordering of the lists was pseudorandomized, with the restrictions that the first list in the practice and test phases were in the native language of the participant and that no more than three lists of the same language could appear sequentially. If the participant looked outside the predefined AOI for two consecutive seconds, the trial ended. The experiment ended if the participant completed all the test trials or if they became fussy or stopped looking at the screen. The experiment lasted approximately 7 min. Caregivers signed a consent form before the experiment and a small gift was given to them at the end of the session.

2.4 | Data analyses

Looking times were calculated using the eyetrackingR package (Dink & Ferguson, 2015) in RStudio (RStudio Team, 2019), report (Makowski, 2018), and a custom-made script. Looking times were calculated only when the participants were looking inside the AOIs. Looking anywhere else was considered as looking away time. For a trial to be valid, the minimum looking time was 1.25 s[3] that corresponds to the longest duration of our nonsense words. For a participant to be valid, they had to have at least four valid trials in the test phase and a balanced number in each condition, at least two

[3] The reviewers suggested that this criterion might be too short, so we repeated the analyses with a threshold of 2.5 s per trial. The analyses with the new threshold with the whole sample and without the outliers can be found in Annexes 2A and 2B. The analyses yielded the same pattern of results.

for the native condition and two for the non-native. If participants had more than 4 valid trials, they had to have an approximately equal number of trials per condition.

Looking times were analyzed with linear mixed models, using *lme4* (Bates et al., 2015). The dependent variable was looking time in seconds during the test phase. The fixed effects were stimuli (Native vs. Non-native, coded as −0.5 and 0.5) and Language Profile (Catalan vs. Spanish, coded as −0.5 and 0.5). Participant number was added as a random effect. The fit of the models was evaluated using the 'anova' function. The pie plot (Figure 1) was generated using 'ggplot2' (Wickham, 2016) and the violin plots (Figure 3) using the R package 'ggstatsplot' (Patil, 2021).

3 | RESULTS

Mean looking times to the two types of stimuli, that is, native and non-native, are shown in Figure 3 separated according to Language Profile. We fitted a linear mixed model (estimated using REML and BOBYQA optimizer) to predict Looking Times with Stimuli and Language Profile as fixed effects (formula: AOI_sec ~ Stimuli * Language Profile). The model included Participant as random effect (formula: ~1 | Participant.Number). The model's total explanatory power is weak (conditional $R^2 = 0.13$) and the part related to the fixed effects alone (marginal R^2) is of 0.01. The output of the model is summarized in Table 3. The model includes Language Profile (Catalan, Spanish) and the interaction with Stimuli following the recommendations of the reviewers.

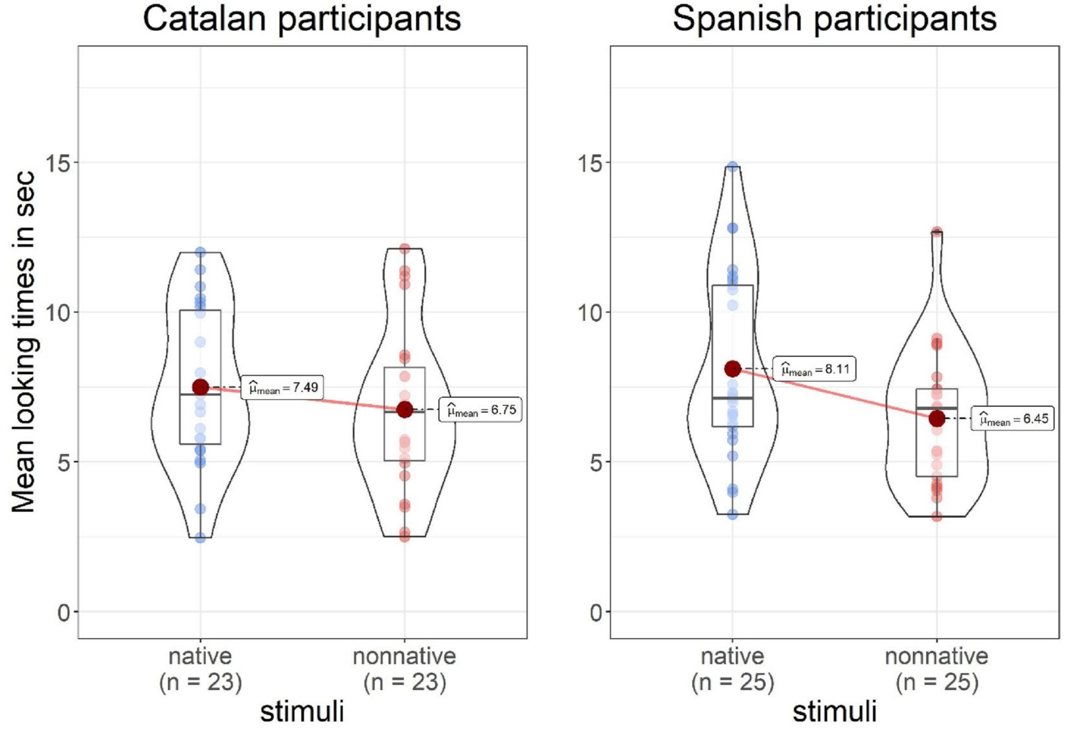

FIGURE 3 Individual participants' mean looking times to the native and non-native lists separated by Language Profile. Red large dots stand for the group mean, boxplots stand for the interquartile difference, and horizontal lines for the median

TABLE 3 Output of the model with Language Profile

Predictors	Estimates	CI	p
(Intercept)	7.35	6.67–8.03	**<0.001**
Stimuli1	−1.16	−2.00 to −0.32	**0.007**
Language Profile1	0.11	−1.25–1.48	0.871
Stimuli1 * Language Profile1	−0.89	−2.57–0.79	0.298
Random effects			
σ^2	26.89		
$\tau_{00\ \text{Participant.Number}}$	3.45		
ICC	0.11		
$N_{\text{Participant.Number}}$	48		
Observations	591		
Marginal R^2/conditional R^2	0.013/0.125		

Note: The significant p-values are in bold inside the tables.

The model's intercept, when all the predictors are centered at 0, is at 7.35 [95% CI (6.67, 8.03), $t(585) = 21.11$, $p < 0.001$], meaning that participants looked at the lists of words on average for 7.35 s. We found that the effect of Stimuli is statistically significant and negative [beta = −1.16, 95% CI (−2.00, −0.32), $t(585) = -2.71$, $p = 0.007$]. Participants paid attention to the native lists 1.16 s more than to the non-native ones. We found that the effect of Language Profile is statistically nonsignificant and positive [beta = 0.11, 95% CI (−1.25, 1.48), $t(585) = 0.16$, $p = 0.871$]. The interaction between Language Profile on Stimuli was also statistically nonsignificant and negative [beta = −0.89, 95% CI (−2.57, 0.79), $t(585) = -1.04$, $p = 0.298$]. Standardized parameters were obtained by fitting the model on a standardized version of the dataset. 95% Confidence Intervals (CIs) and p values were computed using the Wald approximation.

The best fitted model was the one including only Stimuli as a fixed effect. We fitted a linear mixed model (estimated using REML and BOBYQA optimizer) to predict Looking Times with Stimuli (formula: AOI_sec ~ stimuli). The model included Participant as it is described above random effect (formula: ~1 | Participant.Number). The model's total explanatory power is weak (conditional $R^2 = 0.12$) and the part related to the fixed effects alone (marginal R^2) is of 0.01. The output of the model is summarized in Table 4.

The model's intercept, when stimuli is centered at 0, is at 7.36 [95% CI (6.68, 8.03), $t(587) = 21.34$, $p < 0.001$], meaning that participants looked at the lists of words on average for 7.36 s. We found that the effect of Stimuli is statistically significant and negative [beta = −1.18, 95% CI (−2.02, −0.34), $t(587) = -2.76$, $p = 0.006$]. Participants paid attention to the native lists 1.18 s more than to the non-native ones. Standardized parameters were obtained by fitting the model on a standardized version of the dataset. 95% Confidence Intervals (CIs) and p values were computed using the Wald approximation.

4 | DISCUSSION

The objective of the present study was to investigate whether infants show sensitivity to the native language vowel system around the fourth month of life, an earlier time than first signs of perceptual narrowing for vowels, that is, 6 months, have been reported (Kuhl et al., 1992; Polka & Werker, 1994).

TABLE 4 Output of the best fitted model

Predictors	Estimates	CI	*p*
(Intercept)	7.36	6.68–8.03	**<0.001**
Stimuli1	−1.18	−2.02 to −0.34	**0.006**
Random effects			
σ^2	26.89		
$\tau_{00\ \text{Participant.Number}}$	3.35		
ICC	0.11		
$N_{\text{Participant.Number}}$	48		
Observations	591		
Marginal R^2/conditional R^2	0.011/0.121		

Note: The significant *p*-values are in bold inside the tables.

We ran a preference procedure using the contrasting naturally occurring frequency distribution of vowels found in Spanish and Catalan. The results showed that infants prefer listening to lists that resemble the incidence of vowels in their native language over lists that do not. Our findings are robust given that the same pattern of results was observed in all the different analyses reported in this manuscript and in the annexes.

Previous research investigating the establishment of the native phoneme system has compared infants' responses to native and non-native contrasts at different points in time. The fact that infants can show discrimination at an early age, and no longer a few months later has been interpreted as indicating that the acquisition of the native phoneme system started at some point between the two testing times. The result of our investigation suggests that such approaches may have underestimated infants' linguistic knowledge. However, our results do not allow us to conclude if perceptual narrowing, measured as the decline of discrimination to non-native contrasts has already started at the age we have tested. As mentioned, Bosch and Sebastián-Gallés (2003) showed that Catalan and Spanish 4.5-month-olds, the same population tested here, were equally able to discriminate the /e/–/ɛ/ contrast, native and non-native, respectively; however, at 8 months of age, only infants learning Catalan continued to discriminate the contrast. The authors concluded that perceptual narrowing to the native vowel space had not started at 4.5 months. It is possible that behavioral paradigms based on procedures measuring recovery of attention after habituation/familiarization exposures may not provide sufficiently fine-grained measures to detect the onset of the characteristic decline of non-native contrasts in perceptual narrowing. Electrophysiological measures might offer an alternative approach, as some adult studies have reported changes in neurophysiological responses preceding changes in behavior (Tremblay et al., 1998) and the same possibility has been postulated for infants (Cheour et al., 1998). However, the difficulty of getting sufficiently robust measures at the individual level, together with the individual differences in the time-course of perceptual narrowing, may make this approach methodologically quite challenging.

It is worth noticing that the distribution of vowels in Catalan and Spanish has been based on phoneme counts and therefore constitutes a proxy to the actual distribution of speech tokens (phones). It is well established that there is significant overlap in the acoustic realizations of vowels categorized as belonging to different categories, both in adult (Hillenbrand et al., 1995) and in infant speech (Adriaans & Swingley, 2017; Kuhl et al., 1997; Werker et al., 2007). See Morrison (2014) for example, showing the overlap of Spanish /e/ tokens with Catalan /e/ and /ɛ/ ones.

We have assumed that infants have used the distribution of vowels in the lists to prefer one type of list over the other. It might be argued that infants used other kinds of information also present in the stimuli. Infants could have used information about specific sounds; however, the lists in the two languages used the same consonants and vowels, so infants could not have used such information to prefer one list over another. A second type of information refers to specific properties of some words. Although the vowels and consonants were the same in both lists, because of the difference in phonotactic rules between Catalan and Spanish, on average half of the words in each list were only permissible in Spanish, while the other half were permissible in both languages. The fact that Catalan has vowel reduction implies that only one mid-vowel can appear in a word, that is, the stimulus/go'bete/, cannot be a word in Catalan, since /o/ and /e/ are reduced when not stressed in Catalan (it should be /gu'betə/; Toro et al., 2011, showed that Catalan adults, but not Spanish ones, use this type of information to segment words in continuous speech). It is highly unlikely that infants might have used this information to prefer one type of list over the other. First because the effect should have been present only in Catalan-learning infants (for Spanish-learning infants, all stimuli are phonotactically permissible words in their language of exposure), and our results did not show an interaction between stimuli and language of exposure. Second, because the emergence of preference towards lists of words that contain more frequent phonotactic patterns has not been reported before 9–10 months of age (Gonzalez-Gomez & Nazzi, 2015; Jusczyk, Friederici et al., 1993; Jusczyk et al., 1994; Nazzi et al., 2009). A third possibility is that infants may have paid more attention to stressed than unstressed syllables. This is also unlikely because such bias would make the differences disappear between the two languages, as vowel reduction only applies to unstressed syllables, therefore no preference should be evident. Finally, a fourth type of information that infants could have paid attention to are word edges. Infants might have paid more attention to the first syllable (primacy effect) or the last one (recency effect) in each pseudoword. This effect has been shown previously with 3-month-old infants when presented with speech sequences of five syllables (Hochmann et al., 2016). The vowel information gathered from only the first or last syllable from our stimuli coincides with naturally occurring vowel distribution in the whole language, therefore we cannot be certain on if infants computed distributions over all the stimuli or just the edges. This question remains to be answered in future experiments, specifically designed to answer this issue.

We propose that before the emergence of the phonetic system, as a consequence of perceptual narrowing, infants compute the frequency distribution of sounds and are sensitive to speech mirroring the properties of the native language. Such a process would take place over months, likely starting prenatally. Moon et al. (2013) showed that newborns prefer listening to vowels they had been exposed to prenatally (native language) over phonemes they have never been exposed to before (foreign language), so sensitivity to speech sounds frequency can be detected already at birth. We also propose that this type of representations would contribute to the more refined language discrimination abilities (beyond the use of rhythmic/prosodic information) observed in infants around 4–5 months of age. Zacharaki and Sebastian-Galles (2021) recently reported that 4-5-month-old infants can discriminate between two dialects of Catalan only differing in the distributional properties of their vocalic systems.

Our results do not provide information on the nature of information infants compute distributional information over. There is no doubt that frequency of appearance is an important factor of native phoneme category formation (Cristia et al., 2011; Jusczyk, 1993) and in the decline of perception non-native phonemes (Anderson et al., 2003). Infants acquire first the most frequent native phonemes, such as coronal stops in English, than less frequent ones, such as dorsal stops, and consequently lose the ability to discriminate between the more frequent non-native sounds first (Anderson et al., 2003).

This pattern of behavior has been found with consonants at 8–9 months of age, earlier than the age typically associated with perceptual narrowing of consonants, that is, 10–12 months of age (Werker

& Tees, 1984). These earlier effects of frequency on speech perception align with our hypothesis and results. Our results remain neutral concerning whether infants compute statistics over isolated phones (vowels), over larger speech units (syllable-like), or over specific acoustic properties. The patterns of preference we have observed may have been driven by the differences in the two lists in vowel sounds, syllable-like units, or even by F1 values encoding vowel height.

In conclusion, the present study provides evidence of emergence of native phonetic knowledge at 4.5 months of age, earlier than previously reported. Infants keep track of the most frequent sounds present in their native language and they can use such knowledge when presented with novel stimuli. Future research can shed light on the exact information that infants compute their statistics over.

ACKNOWLEDGMENTS

This research was supported by grants from the Spanish Ministerio de Ciencia, Innovación y Universidades (PGC2018-101831-B-I00), and the Catalan Government [SGR 2017-268 and FI-9015-456763; ICREA (Catalan Institution for Research and Advanced Studies) Academia 2019 award]. Konstantina Zacharaki was supported by a fellowship of the Catalan Government (FI2018). The authors declare no conflicts of interest with regard to the funding source of this study. We thank Carlota Pagès-Portabella for producing the stimuli. We are grateful to Chiara Santolin and Gonzalo Garcia-Castro for their helpful feedback. We thank Xavier Mayoral and Silvia Blanch for their technical support. We also thank Cristina Dominguez and Katia Pistrin for their efforts in recruiting infants. We thank Edward Cousland who proofread the manuscript. We would like to thank the clinics Quirón and Sagrada Familia that allowed us to recruit participants in their premises. We also thank all families and infants who participated in the experiments. Data collection was half-way when the COVID-19 pandemic started. We would like to pay special tribute to the families that collaborated with us under these difficult circumstances.

CONFLICT OF INTEREST

The authors report no conflict of interest.

ORCID

Konstantina Zacharaki https://orcid.org/0000-0001-8264-8457

REFERENCES

Adriaans, F., & Swingley, D. (2012). Distributional learning of vowel categories is supported by prosody in infant-directed speech. In *Proceedings of the Annual Meeting of the Cognitive Science Society (CogSci)*, 34, 72–74.

Adriaans, F., & Swingley, D. (2017). Prosodic exaggeration within infant-directed speech: Consequences for vowel learnability. *The Journal of the Acoustical Society of America*, *141*(5), 3070–3078. https://doi.org/10.1121/1.4982246

Albareda-Castellot, B., Pons, F., & Sebastián-Gallés, N. (2011). The acquisition of phonetic categories in bilingual infants: New data from an anticipatory eye movement paradigm. *Developmental Science*, *14*(2), 395–401. https://doi.org/10.1111/j.1467-7687.2010.00989.x

Alcina, F. J., & Blecua, J. M. (1975). *Gramática española*. Esplugues de Llobregat.

Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A statistical basis for speech sound discrimination. *Language and Speech*, *46*(2–3), 155–182. https://doi.org/10.1177/00238309030460020601

Aslin, R., Jusczyk, P., & Pisoni, D. (1998). *Speech and auditory processing during infancy: Constraints on and precursors to language*. John Wiley & Sons Inc.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer*. http://www.praat.org/

Bosch, L., Costa, A., & Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*, *12*(2), 189–221. https://doi.org/10.1080/09541446.2000.10590222

Bosch, L., & Sebastian-Galles, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, *2*(1), 29–49. https://doi.org/10.1207/s15327078in0201_3

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, *46*(2–3), 217–243. https://doi.org/10.1177/00238309030460020801

Carbonell, J. F., & Llisterri, J. (1992). Catalan. *Journal of the International Phonetic Association*, *22*(1–2), 53–56. https://doi.org/10.1017/S0025100300004618

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, *1*(5), 351–353. https://doi.org/10.1038/1561

Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., Aaltonen, O., Eerola, O., & Näätänen, R. (1995). Mismatch negativity indicates vowel discrimination in newborns. *Hearing Research*, *82*(1), 53–58. https://doi.org/10.1016/0378-5955(94)00164-L

Cristia, A., Mcguire, G. L., Seidl, A., & Francis, A. L. (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. *Journal of Phonetics*, *39*(3), 388–402. https://doi.org/10.1016/j.wocn.2011.02.004

Dink, J. W., & Ferguson, B. (2015). *eyetrackingR: An R library for eye-tracking data analysis*.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303–306. https://doi.org/10.1126/science.171.3968.303

Friederici, A. D., & Wessels, J. M. I. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics*, *54*(3), 287–295. https://doi.org/10.3758/BF03205263

Gonzalez-Gomez, N., & Nazzi, T. (2015). Constraints on statistical computations at 10 months of age: The use of phonological features. *Developmental Science*, *18*(6), 864–876. https://doi.org/10.1111/desc.12279

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*(5), 3099–3111. https://doi.org/10.1121/1.411872

Hochmann, J. R., Langus, A., & Mehler, J. (2016). An advantage for perceptual edges in young infants' memory for speech. *Language Learning*, *66*(September), 13–28. https://doi.org/10.1111/lang.12202

Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics*, *21*(1–2), 3–28. https://doi.org/10.1016/s0095-4470(19)31319-1

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*(3), 675–687. https://doi.org/10.1111/j.1467-8624.1993.tb02935.x

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, *32*(3), 402–420. https://doi.org/10.1006/jmla.1993.1022

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, *33*(5), 630–645. https://doi.org/10.1006/jmla.1994.1030

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633. https://doi.org/10.3758/BRM.42.3.627

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*(5326), 684–686. https://doi.org/10.1126/science.277.5326.684

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, R. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5), 606–608. https://doi.org/10.1126/science.1736364

Makowski, D. (2018). The psycho package: An efficient and publishing-oriented workflow for psychological science. *Journal of Open Source Software*, *3*(22), 2–3. https://doi.org/10.21105/joss.00470

Martínez-Celdrán, E., Fernández-Planas, A. M., & Carrera-Sabaté, J. (2003). Castilian Spanish. *Journal of the International Phonetic Association*, *33*(2), 255–259. https://doi.org/10.1017/S0025100303001373

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111. https://doi.org/10.1016/s0010-0277(01)00157-3

Mazzoni, D. (1999). *Audacity®*.

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science*, *12*(3), 369–378. https://doi.org/10.1111/j.1467-7687.2009.00822.x

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: A two-country study. *Acta Paediatrica (Oslo, Norway: 1992)*, *102*(2), 156–160. https://doi.org/10.1111/apa.12098

Morrison, G. S. (2014). An acoustic and statistical analysis of Spanish mid-vowel allophones. *Estudios de Fonetica Experimental*, *13*, 12–37.

Nazzi, T., Bertoncini, J., & Bijeljac-Babic, R. (2009). A perceptual equivalent of the labial-coronal effect in the first year of life. *The Journal of the Acoustical Society of America*, *126*(3), 1440–1446. https://doi.org/10.1121/1.3158931

Pallier, C., Bosch, L., & Sebastian-Galles, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, *64*(3), B9–B17. https://doi.org/10.1016/S0010-0277(97)00030-9

Patil, I. (2021). Visualizations with statistical details: The "ggstatsplot" approach. *Journal of Open Source Software*, *6*(61), 3167. https://doi.org/10.21105/joss.03167

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421–435. https://doi.org/10.1037/0096-1523.20.2.421

Rafel, J. (1980). Dades sobre la freqüència de les unitats fonològiques en català. *Estudis Universitaris Catalans XXIII*, *2*, 473–496.

RStudio Team. (2019). *RStudio: Integrated development environment for R*. RStudio, Inc. http://www.rstudio.com/

Schatz, T., Feldman, N. H., Goldwater, S., Cao, X.-N., & Dupoux, E. (2021). Early phonetic learning without phonetic categories: Insights from large-scale simulations on realistic input. *Proceedings of the National Academy of Sciences*, *118*(7), 2001844118. https://doi.org/10.1073/pnas.2001844118/-/DCSupplemental.y

Sebastián-Gallés, N., & Bosch, L. (2002). Building phonotactic knowledge in bilinguals: Role of early exposure. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(4), 974–989. https://doi.org/10.1037/0096-1523.28.4.974

Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, *72*(2), 111–123. https://doi.org/10.1016/S0010-0277(99)00024-4

Tobias, J. (1959). Relative occurrence of phonemes in American English. *Journal of the Acoustical Society of America*, *31*(5), 631. https://doi.org/10.1121/1.1907766

Toro, J. M., Pons, F., Bion, R. A. H., & Sebastián-Gallés, N. (2011). The contribution of language-specific knowledge in the selection of statistically-coherent word candidates. *Journal of Memory and Language*, *64*(2), 171–180. https://doi.org/10.1016/j.jml.2010.11.005

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development*, *47*(2), 466–472. https://doi.org/10.2307/1128803

Tremblay, K., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: Neurophysiological changes during speech-sound training. *NeuroReport*, *9*(16), 3560. https://doi.org/10.1097/00001756-199811160-00003

Tsuji, S., & Cristia, A. (2013). Perceptual attunement in vowels: A meta-analysis. *Developmental Psychobiology*, *56*(2), 179–191. https://doi.org/10.1002/dev.21179

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(33), 13273–13278. https://doi.org/10.1073/pnas.0705369104

Wanrooij, K., Boersma, P., & vanZuijen, T. L. (2014). Fast phonetic learning occurs already in 2-to-3-month old infants: An ERP study. *Frontiers in Psychology*, *5*(FEB). https://doi.org/10.3389/fpsyg.2014.00077

Werker, J. F., & Gervain, J. (2013). Speech perception in infancy: A foundation for language acquisition. In *The Oxford handbook of developmental psychology* (pp. 909–925). Oxford University Press.

Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*, *103*(1), 147–162. https://doi.org/10.1016/j.cognition.2006.03.006

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. https://doi.org/10.1016/s0163-6383(84)80022-3

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. http://ggplot2.org

Yeung, H. H., Chen, K. H., & Werker, J. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, *68*(2), 123–139. https://doi.org/10.1016/j.jml.2012.09.004

Zacharaki, K., & Sebastian-Galles, N. (2020). Early vowel preference of Catalan/Spanish 4-5-month-old infants: A distributional account. https://doi.org/10.17605/OSF.IO/2VR85
Zacharaki, K., & Sebastian-Galles, N. (2021). The ontogeny of early language discrimination: Beyond rhythm. *Cognition*, *213*, 104628. https://doi.org/10.1016/j.cognition.2021.104628

How to cite this article: Zacharaki, K., & Sebastian-Galles, N. (2022). Before perceptual narrowing: The emergence of the native sounds of language. *Infancy*, *27*(5), 900–915. https://doi.org/10.1111/infa.12486

APPENDIX

Annex 1: Analysis with outliers ($n = 50$)

The analysis reported in Annex 1 include all valid participants and the two outliers that were spotted during the review process. Originally, we had included them because the pattern of results did not change by doing so, and they were not considered as outliers when the mean was calculated over all looking times regardless of stimuli.

We fitted a linear mixed model (estimated using REML and BOBYQA optimizer) to predict Looking Times with Stimuli and Language Profile as a fixed effect (formula: AOI_sec ~ Stimuli * Language Profile). The model included Participant as random effect (formula: ~1 | Participant. Number). The model's total explanatory power is moderate (conditional $R^2 = 0.20$) and the part related to the fixed effects alone (marginal R^2) is of 0.01. The output of the model is summarized in Table A1.

TABLE A1 Output of the model

Predictors	Estimates	CI	p
(Intercept)	7.71	6.88 to 8.54	**<0.001**
Stimuli1	−1.07	−1.90 to −0.23	**0.012**
Language Profile1	−0.65	−2.31 to 1.01	0.445
Stimuli1 * Language Profile1	−1.07	−2.74 to 0.60	0.209
Random effects			
σ^2	27.99		
$\tau_{00\ \text{Participant.Number}}$	6.48		
ICC	0.19		
$N_{\text{Participant.Number}}$	50		
Observations	623		
Marginal R^2/conditional R^2	0.013/0.199		

Note: The significant p-values are in bold inside the tables.

The model's intercept when all the predictors are centered at 0, is at 7.71 (95% CI [6.88, 8.54], $t(617) = 18.20$, $p < 0.001$), meaning that participants looked at the lists of words on average for 7.7 s. We found that the effect of Stimuli is statistically significant and negative (beta = −1.07, 95% CI [−1.90, −0.23], $t(617) = -2.50$, $p = 0.012$). Participants paid attention to the native lists 1.07 s more than to the non-native ones. We found that the effect of Language Profile is statistically nonsignificant and negative (beta = −0.65, 95% CI [−2.31, 1.01], $t(617) = -0.76$, $p = 0.445$). The interaction between

Language Profile on Stimuli was also not statistically nonsignificant and negative (beta = −1.07, 95% CI [−2.74, 0.60], $t(617) = -1.26$, $p = 0.209$). Standardized parameters were obtained by fitting the model on a standardized version of the dataset. 95% Confidence Intervals (CIs) and p-values were computed using the Wald approximation.

Annex 2A: Analysis with 2.5 s as a threshold of valid trials ($n = 46$)

Per the suggestion of the reviewers, we reanalyzed the data by setting a new threshold of the minimum looking time per trial to 2.5 s. Due to this new threshold we had to eliminate four participants (3 Catalan Monolinguals and 1 Spanish Monolingual) who did not reach this new criterion.

We fitted a linear mixed model (estimated using REML and BOBYQA optimizer) to predict Looking Times with Stimuli and Language Profile (formula: AOI_sec ~ Stimuli * Language Profile). The model included Participant.Number as random effect (formula: ~1 | Participant.Number). The model's total explanatory power is moderate (conditional $R^2 = 0.19$) and the part related to the fixed effects alone (marginal R^2) is of 0.02. The output of the model is summarized in Table A2.

TABLE A2 Output of the model

Predictors	Estimates	CI	p
(Intercept)	8.98	8.15 to 9.81	**<0.001**
Stimuli1	−1.44	−2.35 to −0.52	**0.002**
Language Profile1	−0.85	−2.51 to 0.80	0.313
Stimuli1 * Language Profile1	−0.58	−2.41 to 1.25	0.537
Random effects			
σ^2	27.04		
$\tau_{00\ \text{Participant.Number}}$	5.54		
ICC	0.17		
$N_{\text{Participant.Number}}$	46		
Observations	504		
Marginal R^2/conditional R^2	0.021/0.188		

Note: The significant p-values are in bold inside the tables.

The model's intercept when all the predictors are centered at 0 is at 8.98 (95% CI [8.15, 9.81], $t(498) = 21.25$, $p < 0.001$), meaning that participants looked at the lists of words on average for 8.98 s. We found that the effect of Stimuli is statistically significant and negative (beta = −1.44, 95% CI [−2.35, −0.52], $t(498) = -3.08$, $p = 0.002$). Participants paid attention to the native lists 1.44 s more than to the non-native ones. We found that the effect of Language Profile is statistically nonsignificant and negative (beta = −0.85, 95% CI [−2.51, 0.80], $t(498) = -1.01$, $p = 0.313$). The interaction between Language Profile on Stimuli is statistically nonsignificant and negative (beta = −0.58, 95% CI [−2.41, 1.25], $t(498) = -0.62$, $p = 0.537$). Standardized parameters were obtained by fitting the model on a standardized version of the dataset. 95% Confidence Intervals (CIs) and p-values were computed using the Wald approximation.

Annex 2B: Analysis with 2.5 s as a threshold of valid trials without outliers ($n = 44$)

We also ran the analysis without the outliers (2 Catalan monolinguals) as presented in the main paper to allow for a better comparison.

We fitted a linear mixed model (estimated using REML and BOBYQA optimizer) to predict Looking Times with Stimuli and Language Profile (formula: AOI_sec ~ Stimuli * Language Profile). The model included Participant.Number as random effect (formula: ~1 | Participant.Number). The model's total explanatory power is weak (conditional $R^2 = 0.11$) and the part related to the fixed effects alone (marginal R^2) is of 0.02. The output of the model is summarized in Table A3.

TABLE A3 Output of the model

Predictors	Estimates	CI	p
(Intercept)	8.61	7.92–9.29	**<0.001**
Stimuli1	−1.53	−2.45 to −0.60	**0.001**
Language Profile1	−0.03	−1.40–1.34	0.968
Stimuli1 * Language Profile1	−0.39	−2.25–1.46	0.678
Random effects			
σ^2	25.99		
$\tau_{00\ \text{Participant.Number}}$	2.77		
ICC	0.10		
$N_{\text{Participant.Number}}$	44		
Observations	473		
Marginal R^2/conditional R^2	0.020/0.115		

Note: The significant p-values are in bold inside the tables.

The model's intercept when all the predictors are centered at 0 is at 8.61 (95% CI [7.92, 9.29], $t(467) = 24.66$, $p < 0.001$), meaning that participants looked at the lists of words on average for 8.61 s. We found that the effect of Stimuli is statistically significant and negative (beta = −1.53, 95% CI [−2.45, −0.60], $t(467) = -3.23$, $p = 0.001$). Participants paid attention to the native lists 1.53 s more than to the non-native ones. We found that the effect of Language Profile is statistically nonsignificant and negative (beta = −0.03, 95% CI [−1.40, 1.34], $t(467) = -0.04$, $p = 0.968$). The interaction between Language Profile on Stimuli is statistically nonsignificant and negative (beta = −0.39, 95% CI [−2.25, 1.46], $t(467) = -0.42$, $p = 0.678$). Standardized parameters were obtained by fitting the model on a standardized version of the dataset. 95% Confidence Intervals (CIs) and p-values were computed using the Wald approximation.