Received: 2 October 2023 | Revised: 22 April 2024 | Accepted: 15 May 2024

DOI: 10.1111/desc.13533

Developmental Science WILEY

RESEARCH ARTICLE

Infants' abilities to segment word forms from spectrally degraded speech in the first year of life

Irene de la Cruz-Pavía[1,2] | Monica Hegde[3] | Laurianne Cabrera[3] | Thierry Nazzi[3]

[1]Faculty of Social and Human Sciences, Universidad de Deusto, Bilbao, Spain

[2]Basque Foundation for Science Ikerbasque, Bilbao, Spain

[3]INCC UMR 8002, CNRS, F-75006, Université Paris Cité, Paris, France

Correspondence
Irene de la Cruz-Pavía, Faculty of Social and Human Sciences, Universidad de Deusto, Bilbao, Spain.
Email: irene.delacruz.pavia@deusto.es

Monica Hegde, INCC UMR 8002, CNRS, F-75006, Université Paris Cité, Paris, France.
Email: hegdms0@gmail.com

Funding information
Agence Nationale de la Recherche, Grant/Award Number: ANR-17-CE28-0008 DESIN; ANR's French Investissements d'Avenir—Labex EFL Program, Grant/Award Number: ANR-10-LABX-0083; Basque Foundation for Science Ikerbasque, the Basque Government, Grant/Award Number: IT1439-22; MCIN/AEI/10.13039/501100011033 and by "European Union NextGenerationEU/PRTR", Grant/Award Number: RYC2021-03395-I

Abstract

Infants begin to segment word forms from fluent speech—a crucial task in lexical processing—between 4 and 7 months of age. Prior work has established that infants rely on a variety of cues available in the speech signal (i.e., prosodic, statistical, acoustic-segmental, and lexical) to accomplish this task. In two experiments with French-learning 6- and 10-month-olds, we use a psychoacoustic approach to examine if and how degradation of the two fundamental acoustic components extracted from speech by the auditory system, namely, temporal (both frequency and amplitude modulation) and spectral information, impact word form segmentation. Infants were familiarized with passages containing target words, in which frequency modulation (FM) information was replaced with pure tones using a vocoder, while amplitude modulation (AM) was preserved in either 8 or 16 spectral bands. Infants were then tested on their recognition of the target versus novel control words. While the 6-month-olds were unable to segment in either condition, the 10-month-olds succeeded, although only in the 16 spectral band condition. These findings suggest that 6-month-olds need FM temporal cues for speech segmentation while 10-month-olds do not, although they need the AM cues to be presented in enough spectral bands (i.e., 16). This developmental change observed in infants' sensitivity to spectrotemporal cues likely results from an increase in the range of available segmentation procedures, and/or shift from a vowel to a consonant bias in lexical processing between the two ages, as vowels are more affected by our acoustic manipulations.

KEYWORDS
infants, spectral resolution, speech segmentation, temporal modulations, vocoded speech, word forms

Research Highlights

- Although segmenting speech into word forms is crucial for lexical acquisition, the acoustic information that infants' auditory system extracts to process continuous speech remains unknown.

Irene de la Cruz-Pavía and Monica Hegde contributed equally to this study.

- We examined infants' sensitivity to spectrotemporal cues in speech segmentation using vocoded speech, and revealed a developmental change between 6 and 10 months of age.
- We showed that FM information, that is, the fast temporal modulations of speech, is necessary for 6- but not 10-month-old infants to segment word forms.
- Moreover, reducing the number of spectral bands impacts 10-month-olds' segmentation abilities, who succeed when 16 bands are preserved, but fail with 8 bands.

1 | INTRODUCTION

Before infants can learn words and their meanings, they need to develop the ability to extract word forms from the speech they hear in their environment. This process, referred to as word form segmentation, is anything but trivial given that speech is by and large a continuous signal. Through the segmentation process, infants build phonological representations of potential word forms, that is, the words' sound patterns, and subsequently establish connections between these sound forms and their corresponding meanings. Word form segmentation is thus a crucial building block necessary for lexical acquisition and for accessing the syntactic and semantic structure of language. While many psycholinguistic studies have explored when word segmentation emerges in development and the linguistic cues used to segment words across a variety of languages, we still do not understand the specific acoustic properties of the speech signal that infants rely on for successful word form segmentation. Knowing that the auditory system is still maturing during infancy (Moore, 2002), it is thus possible that the weight of specific auditory cues recruited in speech segmentation changes throughout this process. By investigating the auditory underpinnings of this central aspect of lexical acquisition, the current study seeks to deepen our understanding of how infants' auditory system extracts the acoustic components of the speech signal to process continuous speech and allow language acquisition.

A wealth of psycholinguistic studies has investigated the ability to segment word forms from fluent speech during infancy. This ability to rapidly extract repeated word forms from sentences has been consistently found between 8 and 12 months, with studies occasionally demonstrating word form segmentation in 6- or even 4-month-olds (e.g., in English: Jusczyk & Aslin, 1995; Spanish: Bosch et al., 2013; German: Höhle & Weissenborn, 2003; Dutch: Houston et al., 2000; Parisian French: Nazzi et al., 2006; and Canadian French: Polka & Sundara, 2012). Interestingly, the moment in which this ability emerges seems to vary somewhat across languages (e.g., 4 months in French: Berdasco-Muñoz et al., 2018, 7.5 months in English: Jusczyk & Aslin, 1995). Although the source of this apparent crosslinguistic variation remains unclear, it might at least partially result from prosodic differences between languages, specifically in their rhythm (Berdasco-Muñoz et al., 2018; Jusczyk & Aslin, 1995).

In parallel to determining *when* infants begin to segment words from speech, a substantial body of work has investigated *how* they achieve this feat, that is, which cues infants use for word form segmentation. These studies point to four main sources of information being used across languages, namely, statistical, prosodic, and acoustic-segmental cues, and prior lexical knowledge. Infants have been found to detect statistical regularities in the input, computing for instance transition probabilities between syllables. Tracking this statistical cue allows infants as young as 5.5 months to extract word forms from speech, as high transition probabilities indicate cohesive units, that is, words, while dips in these probabilities signal word boundaries (for English: Saffran et al., 1996; for French: Hoareau et al., 2019; Mersad & Nazzi, 2012; for Dutch: Johnson & Tyler, 2010), although limits to this ability have also been found (Johnson & Tyler, 2010; Mersad & Nazzi, 2012). From around 7 months of age, infants also begin to rely on a number of prosodic cues to segment speech, including pitch accent (for English: Nazzi, et al., 2005), major prosodic boundaries (for English: Gout et al., 2004), and rhythm (Jusczyk, Houston et al., 1999; Nazzi et al., 2006). This latter cue varies cross-linguistically, as different languages have different basic rhythmic units: while English-learning infants rely on strong-weak feet (Jusczyk, Houston et al., 1999), French-learning infants rely on syllables (Goyet et al., 2013; Nazzi et al., 2006, 2014; Nishibayashi et al., 2015). Statistical and prosodic cues are the first segmentation cues shown to be available to infants, presumably due to the fact that they do not require much knowledge of the native language. Indeed, even newborn infants can track transition probabilities (Fló et al., 2022; Teinonen et al., 2009), are able to discriminate languages based on rhythm (Byers-Heinlein et al., 2010; Nazzi, Bertoncini et al., 1998), distinguish words based on pitch contour (Nazzi, Floccia et al., 1998), group syllables into prosodically well-formed sequences (Abboub et al., 2016), and perceive prosodic (and syntactic) boundaries (Christophe et al., 1994, 2001).

As their knowledge and experience with (the native) language accumulates, infants begin to make use of at least three sources of acoustic-segmental information to segment words, that is, cues involving units smaller than the word, such as individual sounds (or segments): (a) coarticulation, that is, the overlap of adjacent articulations while producing the speech sounds, by around 8 months of age (for English: Johnson & Jusczyk, 2001; for French: Nishibayashi et al., 2015), (b) phonotactic constraints, that is, the language-specific restrictions on

the sequences of phonemes allowed within- and between-words, by 9 months of age (for English: Mattys & Jusczyk, 2001; for French: Gonzalez-Gomez & Nazzi, 2013), and (c) allophonic variation, that is, the context-dependent variants of a given phoneme, by 10.5 months of age (for English: Jusczyk, Hohne et al., 1999b). Finally, by 6 to 8 months, infants can use their first known word forms (e.g., *Mommy*) to segment the words adjacent to them (Bortfeld et al., 2005; Mersad & Nazzi, 2012).

In sum, a substantial body of work to date has investigated a number of sources of information present in the continuous speech stream that infants may use to segment word forms. These studies have established that some speech cues are used earlier than others in infancy and, further, that these cues are used in combination. Furthermore, the relative perceptual weight of these procedures changes with age, although a detailed developmental path remains to be traced (see debate between Johnson & Jusczyk, 2001, and Thiessen & Saffran, 2003, regarding the use of rhythmic units and transition probabilities). To this aim, the current study investigates the emergence of word segmentation in infancy from a psychoacoustic approach, asking if and how infants use key physical acoustic properties of the speech signal to succeed at this challenging task.

Psychoacoustic models describe the decomposition and analysis of the speech signal by the peripheral and central auditory systems. This approach characterises perceptual mechanisms involved in speech processing that are consistent with the physiological processing of speech by the auditory system. According to current psychoacoustic models (Dau et al., 1997a, 1997b; Moore, 2012), the auditory system encodes speech by first extracting the spectral information of the signal via a series of filters present in the cochlea, then modeling the output signals of these filters as narrow bands that are modulated in amplitude over time. Temporal information is processed simultaneously at two time scales. The slower variations of amplitude over time are known as Amplitude Modulation (AM) or acoustic "temporal envelope" of speech, while the faster oscillations in instantaneous frequency that take place close to the center frequency of each auditory filter are known as Frequency Modulation (FM or acoustic "temporal fine structure"; Moore, 2012). Spectral and temporal (AM and FM) information are hence the two primary acoustic components extracted from speech (and all other nonspeech acoustic signals) by the auditory system. The perception of pitch, loudness, and timbre is driven by changes in these two acoustic components. The fundamental role in speech perception of these components is supported by a wealth of psychoacoustic studies using *vocoders*, which are speech analysis-synthesis tools developed to imitate auditory processing. Crucially, vocoders allow a selective manipulation of the complex spectro-temporal properties of the speech signal, and specifically of three parameters: spectral resolution, that is, the number of spectral bands containing the temporal modulations, which, when reduced, diminish the fine spectral details of the signal, and temporal resolution by reducing AM and/or FM information.

Studies investigating the role of these spectro-temporal parameters in speech perception have revealed that spectral cues and FM convey information on pitch (i.e., F0), stress, place of articulation, and voice quality (Rosen, 1992; Smith et al., 2002; Xu et al., 2005), while AM conveys information about syllabic rate, rhythm, and manner of articulation. Moreover, current evidence suggests that AM cues might play a particularly important role in speech perception at the lexical level. Indeed, studies using vocoders with adult participants show that AM information suffices for near-perfect word recognition, even when carried by a signal comprising only four spectral bands (Shannon et al., 1995; Zeng et al., 2005). As for younger listeners, two studies have assessed word processing based on degraded vocoded speech. One pioneering study with 27-month-old toddlers showed that they successfully associate known words to their corresponding pictures when presented with vocoded sentences in which FM cues are removed but AM cues are carried in only eight spectral bands, but they fail if AM cues are presented in only two bands (Newman & Chatterjee, 2013). In turn, 34-month-old toddlers succeed at associating new words to object pictures when FM cues are removed from the signal, although they require AM cues to be presented in a greater number of spectral bands, that is, 16 instead of 8 (Newmanet al., 2020).

Taken together, the above results demonstrate that the AM of the signal presented in a small number of bands is sufficient to allow adults and toddlers to extract lexical information, and that younger listeners require more spectral information (i.e., a greater number of bands) than adults, with the exact number of bands varying as a function of the stimuli presented (known vs. new words). Here, we seek to determine the granularity required at the AM and FM levels of analysis for young infants to succeed at segmenting word forms from the speech signal.

To investigate this question, we tested French-learning infants at 6 and 10 months of age (experiments 1 and 2, respectively), and manipulated the number of spectral bands in which AM cues are preserved (8 or 16 bands, see Figure 1) while replacing FM by a pure tone carrier in each band. The two age groups were selected in light of previous studies revealing important developmental changes in infants' phonological acquisition between these ages. First, infants start becoming attuned to the vowels of their native language at around 6 months of age (Kuhl et al., 1992; Trehub, 1976), and to the consonants at around 10 months (Werker & Tees, 1984). Second, and in line with these changes, French-learning infants tested in a segmentation task give more weight to vowels when recognizing word forms at 6 months, but give more weight to consonants from 8 months of age. Thus, while 6-month-olds treat a consonant mispronunciation as being more similar to a familiar target than a vowel mispronunciation, they show the opposite pattern at 8 months of age (Nishibayashi & Nazzi, 2016), showing that they have developed an adult-like consonant-bias in lexical processing (Nishibayashi & Nazzi, 2016). As spectral degradation affects adults' identification of vowels to a greater extent than consonants (Xu et al., 2005), word recognition might be more affected in the group of 6-month-olds, who still give more weight to vowels as compared with consonants in lexical processing tasks.

The number of bands (8 or 16) used for the vocoded stimuli were selected based on the previous studies reviewed above, which showed that toddlers succeed at lexical-related tasks using these same conditions, with more challenging tasks requiring a greater number of bands (Newman & Chatterjee, 2013; Newman et al., 2020). Based on these studies, we predicted that infants would more likely succeed at

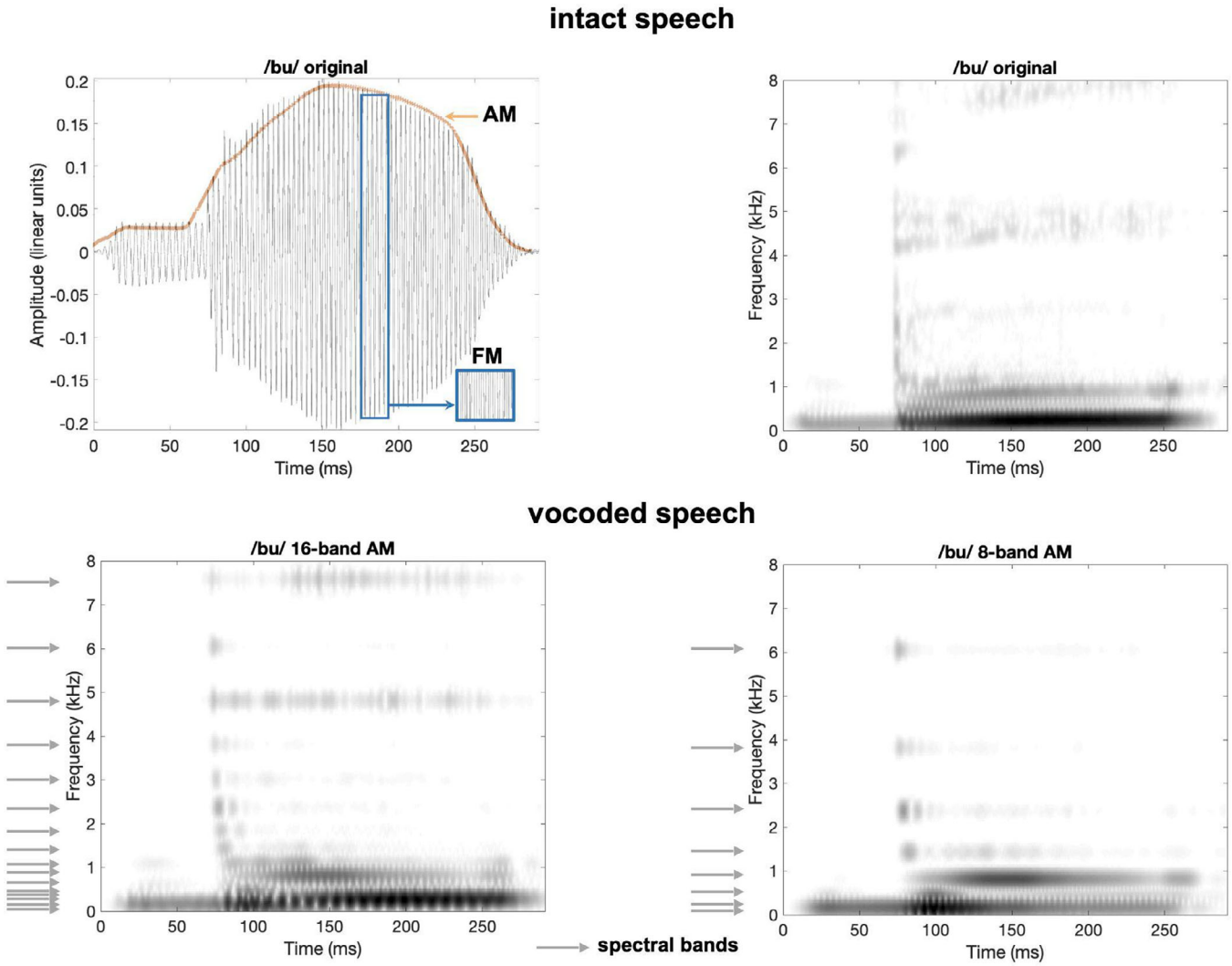

FIGURE 1 The upper panels depict the waveform (left) and spectrogram (right) of one of the French words (*bout* /bu/) used in experiments 1 and 2, in intact speech. Its AM and a sample of its FM are depicted in the waveform (orange line and blue box, respectively). The lower panels depict the spectrograms of the vocoded stimuli used in experiments 1 and 2. The grey arrows depict the number of spectral bands in which AM information was preserved, either 16 (left) or 8 (right).

detecting a previously heard word form when presented with 16 bands compared to 8 bands. Additionally, as mentioned above, we might find age differences, as 6-month-olds rely primarily on statistical and prosodic information to segment intact speech, while 10-month-olds have a wider range of segmentation cues available. Ten-month-olds might hence be more skilled at compensating the loss of FM and spectral information in the signal, while 6-month-olds might be more affected by the reduction of prosodic and phonetic information that results from degrading these spectro-temporal cues.

2 | EXPERIMENT 1

The present experiment was based on a study conducted by Nishibayashi and Nazzi (2016). In that study, French-learning infants were familiarized with two monosyllabic target words embedded in passages, and then presented with four different test trials, corresponding to 20 repetitions of each of the familiar targets or of two previously unheard "control" words. The 6- and 8-month-old infants succeeded at segmenting the two familiarized target words. In a pilot experiment (reported in Appendix 1), we used the same design to test 6-month-old infants with degraded stimuli. Specifically, we removed FM cues and filtered the signal to yield eight spectral bands. Results failed to establish successful segmentation in that condition. Accordingly, in the present study, we simplified Nishibayashi and Nazzi's (2016) design, and familiarized infants with one rather than two target words. As a result, the familiarization time of the target word doubled, which we expected to facilitate segmentation.

Previous studies on word segmentation presenting infants with intact stimuli in the two-target word design have usually found longer listening times to familiar targets as compared to novel controls (Nishibayashi et al., 2015), though a few exceptions are reported in the literature (Bosch et al., 2013, experiment 1; Goyet et al., 2013, experiment 3). This is the first study to date examining infants' segmentation

of vocoded speech. Therefore, we cannot draw specific predictions of their listening behavior. Further, the scarce evidence available exposing French-learning infants to vocoded speech using other tasks yielded mixed results. The 6-month-old infants tested on their syllable discrimination abilities showed a novelty effect (i.e., longer listening times to novel items) when presented with intact stimuli, but an impact of familiarization length when presented with 4-band AM vocoded speech (Cabrera et al., 2013): a familiarity effect (i.e., longer listening times to familiar items) after a 1-min-long familiarization, but a novelty preference after a 2-min-long familiarization. At any rate, if infants in the present study are able to segment the target words presented in the passages, they should exhibit different looking times to target and control words in the test phase. Equal looking times would in turn suggest that 6-month-old infants are not able to segment words from fluent speech when spectral resolution and FM information have been degraded.

2.1 | Methods

2.1.1 | Participants

Forty-eight 6-month-old infants participated in the experiment. All infants were born full-term and were being raised in monolingual French-speaking families. Half of the infants were tested with the 8-band stimuli (8-band group), and the other half with the 16-band stimuli (16-band group, see stimulus section for further details). The 8-band group consisted of 24 infants (14 girls, 10 boys; mean age: 6;13; age range: 6;02–6;28). Data from 18 additional infants were excluded from analysis due to fussiness or crying (11 infants), parental interference (3 infants), online coding difficulties or errors (2 infants), too short looking times (1 infant), or mean looking times 2.5 SD above or below the group mean (1 infant). The 16-band group consisted of 24 infants (12 girls, 12 boys; mean age: 6;18; age range: 6;04–7;01). Data from 15 additional infants were excluded from analysis due to fussiness or crying (9 infants), falling asleep (2 infants), computer error (2 infants), or mean looking times 2.5 SD above or below the group mean (2 infants). All parents gave informed consent before their infant's participation in accordance with the local Ethic Committee.

2.1.2 | Stimuli

Stimuli consisted of four monosyllabic words with a consonant-vowel structure (/bu/ *bout* "piece"; /fø/ *feu* "fire"; /py/ *pus* "pus" and *pu*—"stink" and past participle of the verb "can"; /vo/ *veau* "calf" and *vos* "your" (plural)) selected from the set originally used by Nishibayashi and Nazzi (2016).

Nishibayashi and Nazzi (2016) created a passage for each of the four words, to be used during familiarization. Each passage comprised six sentences (mean sentence length: 11 syllables) that contained the target word toward their beginning (three) or end (three). Within passages, targets were always preceded and followed by different syllables, in order to neutralize information regarding syllable co-occurrences (see Appendix 2). The passages were recorded by a French-native female talker in mild infant directed speech, and were 16 s long. In addition to the passages, the talker produced 20 repetitions of each word in isolation, which were pasted together to create one 20-s-long audio file for each of the four words, to be used in the test phase. The four words were grouped into the pairs /vo/—/py/ and /fø/—/bu/, which constituted the target and control word pairs (see procedure below). Within pairs, the two words differed in the voicing and manner of articulation of their consonants, and in the roundness and backness of their vowels.

We then manipulated the spectro-temporal properties of the original stimuli—passages and files of isolated tokens—by processing them with a vocoder similar to the one used in Cabrera et al. (2014) and de la Cruz-Pavía et al. (2023), which degraded their spectral resolution and FM cues. Specifically, we reduced spectral resolution by passing each speech sound through a bank of fourth-order gammatone filters. Stimuli used for the 8-band group were passed through 8 filters (i.e., 8 spectral bands), while stimuli for the 16-band group were passed through 16 filters.

Gammatone filters were 4-ERB_N wide (equivalent-rectangular bandwidth) for the 8 band stimuli, and 2-ERB_N wide for the 16 band stimuli, with center frequency (Fc) uniformly spaced along an ERB_N-number scale ranging from 80 to 8020 Hz. The ERB scale (Glasberg & Moore, 1990; Moore, 2003) allows simulation of the bandwidth of cochlear filters of the normal ear. The Hilbert transform was applied in each band to extract the AM and FM components. The AM was lowpass-filtered using a zero-phase Butterworth filter (36 dB/octave roll-off, see Ardoint & Lorenzi, [2010], and Cabrera et al., [2014]) and a cut-off frequency of $ERB_N/2$. The original FM carriers were replaced by pure tone carriers centered at the Fc of each spectral band. In each band, the new carrier was multiplied by the filtered AM function. The modulated speech signals—8 bands in the 8-band group, 16 bands in the 16-band group—were then added up and the level of the resulting speech signal was equalized in root mean square value as the input signal.

2.1.3 | Procedure and design

The study took place at the Babylab of the Université Paris Cité's Integrative Neuroscience and Cognition Center (France). It was conducted in a sound-attenuated booth with dim lights, and we used a variant of the headturn preference procedure designed by Jusczyk and Aslin (1995), experiment 3 and used by Nishibayashi and Nazzi (2016). The booth contained three pegboard panels placed at its two sides and front. A green light was mounted on the front panel. Directly below the central light was placed a video camera (its lens mounted on a hole on the panel), in order to monitor the infants' looking behavior during the study. A red light and a hidden loudspeaker were mounted on each of the side panels. Infants were seated on a caregiver's lap, who was in turn seated on a chair placed at the center of the booth. Caregivers listened to masking music over headphones during the study, which

prevented them from hearing the stimuli. An experimenter, placed outside the booth and wearing headphones with masking music, watched the video of the infants on a TV screen, monitoring their looking behavior. Via a button box, the experimenter started and ended the flashing of the central and sidelights and the presentation of the stimuli as a function of the infants' head turns. The computer stored the direction and duration of the infants' looking times automatically.

The study consisted of a familiarization phase followed immediately by a test phase. Both during familiarization and test, each trial started with the green central light flashing silently to attract the infants' attention. Once the infant fixated centrally, the green light was extinguished and one of the red lights started blinking on one of the side panels. When the infant turned at least 30° toward the side light, the stimulus for that trial began to play. Stimuli were played until the end or until the infant looked away from the light for more than 2 s. After this, a new trial began. If an infant's looking time, that is, their orientation time toward the blinking light, was shorter than 1.5 s in a given trial, the trial was immediately replayed and the infant's looking time during the first presentation was discarded. Infants with more than three such short looks were not retained for analysis. Also, if looking time during the repetition of the trial was again shorter than 1.5 s, the trial did not repeat, and the infant was not retained for analysis. These exclusion criteria were applied following Nishibayashi and Nazzi (2016), and are reported as "too short looking times" in the Participants section.

During the familiarization phase, infants listened to a passage containing a target word, which was repeated until they accumulated 1 min of looking time to the target. Once an infant reached this criterion, familiarization ended and the test phase began. It consisted of 12 trials, split into 3 blocks of 4 trials each. In each of the three test blocks, infants heard two trials presenting the 20-repetition list corresponding to the target word, that is, the word presented repeatedly in the passage heard in familiarization, and two trials presenting the 20-repetition list of a control word, presented in semirandom order. Specifically, in each block, target and control were presented once on the left and once on the right side of the booth. Moreover, there were no more than two consecutive trials of the same kind—target/control—or presented on the same side.

Infants were familiar with one of four lists, each containing a different target word. Each of the four words was used as a target for a quarter of the infants and as a control for another quarter (i.e., List 1: target word /vo/, control word /py/; List 2: target /fø/, control /bu/; List 3: target /py/, control /vo/; List 4: target /bu/, control /fø/).

3 | RESULTS AND DISCUSSION

All analyses were conducted in R (version 4.3.1, R Core Team, 2019). The upper panel of Figure 2 depicts the mean raw looking times (LTs) to the target and control trials of the two groups of 6-month-olds. Note that the raw LTs were log-transformed prior to analysis (Csibra et al., 2016). We fit linear mixed effects models (lme4 package). We started by fitting the conceptually most complex model, which explained the dependent variable LTs (log-transformed), with the interaction of the fixed effects of Vocoder (8 bands, 16 bands), and Condition (familiar, novel), and Participant as random factor, and allowed Condition to vary randomly by Participant. We then built decreasingly complex models, running analyses of variances (ANOVAs) to compare pairs of models. The simplest model that fits the data included only the random factor Participant (lmer(LTs_log ~ (1 | Participant), data = data, REML = F); the results of the model, as well as the results of the most complex model that converged are reported in the Supporting Information, and mean raw LTs are reported in Table 1).

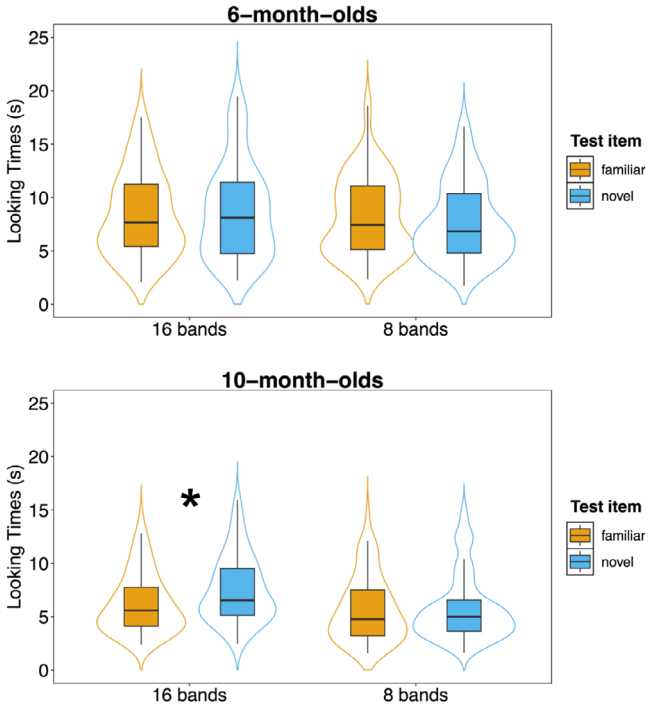

FIGURE 2 The two panels depict box-plots showing participants' median looking times in seconds (central line within each plot) and their upper and lower quartiles, in response to familiar and novel trials (in orange and blue, respectively), at 6 (upper panel) and 10 (lower panel) months, in the 16 and 8 band conditions. Standard deviations are depicted by the violin plots and asterisks depict significant differences.

In sum, these results fail to show that 6-month-old infants segmented target word forms from vocoded speech in which only AM cues were preserved, while FM cues were removed and the number of spectral bands reduced. Moreover, doubling the number of spectral bands (from 8 to 16) did not appear to aid their segmentation.

4 | EXPERIMENT 2

In the present experiment, we examined whether older infants' accumulated language exposure and maturation of their auditory system allows them to segment word forms from a speech signal comprising AM cues extracted in a limited number of frequency bands. Specifically, we tested 10-month-old infants, as by this age they have amassed substantial knowledge of the phonology of their native language

TABLE 1 Mean looking times in seconds and standard error to the familiar and control test items by age and vocoder condition.

	6-Month-olds		10-Month-olds	
8 Bands	Familiar	Novel	Familiar	Novel
	8.18 s, ± 0.43	7.67 s, ± 0.39	5.70 s, ± 0.35	5.75 s, ± 0.35
16 Bands	Familiar	Novel	Familiar	Novel
	8.35 s, ± 0.42	8.67 s, ± 0.47	6.22 s, ± 0.33	7.56 s, ± 0.38

(Nishibayashi & Nazzi, 2016; Werker & Tees, 1984), using the exact same procedure and stimuli as in experiment 1.

4.1 | Methods

4.1.1 | Participants

Forty-eight 10-month-old infants participated in the experiment. All infants were born full-term and were being raised in monolingual French-speaking families. The 8-band group comprised 24 infants (12 girls, 12 boys; mean age: 10;14; age range: 10;03–11;05). Data from two additional infants were excluded from analysis due to fussiness or crying (one infant), or mean looking times 2.5 SD above or below the group mean (one infant). The 16-band group comprised 24 infants (10 girls, 14 boys; mean age: 10;21; age range: 10;04–11;07). Data from 11 additional infants were excluded from analysis due to fussiness or crying (6 infants), too short looking times (2 infants) or mean looking times 2.5 SD above or below the group mean (3 infant). All parents gave informed consent before their infant's participation.

4.1.2 | Stimuli, procedure and design

Stimuli, apparatus, lists and procedure were identical to experiment 1.

4.2 | Results and discussion

Raw LTs were log-transformed. Again, we fit linear mixed effects models, first fitting the conceptually most complex model, which was identical to the one used in experiment 1, and then step-by-step decreasingly complex models which were compared in pairs using ANOVAs. The model that best fits the data (i.e., the final model) included the fixed effects of Condition and Vocoder, and the random factor Participant (lmer(LTs_log ~ Vocoder + Condition + (1 | Participant), data = data, REML = F); the results of this final model and the results of the most complex model that converged are reported in the Supporting Information). The function Anova revealed significant effects of Vocoder ($\chi^2(1) = 9.28$, $p = .002$) and Condition ($\chi^2(1) = 4.87$, $p = .028$). We further analyzed these effects using marginal means (using EM Means function), which revealed longer LTs in the 16-band (6.89 s), compared to the 8-band condition (5.72 s; $p = .004$) (see Figure 2), and longer LTs to novel words (6.66 s) as compared with familiar words (5.96 s; $p = .028$). Out of the 48 infants analysed, 32 had longer LTs to novel as compared with familiar test items: 19 in the 16-band condition, and 13 in the 8-band condition (each $n = 24$).

These results establish that 10-month-old infants succeed at segmenting the familiarized word forms from fluent speech that has been acoustically degraded and only contains AM information. When put together, the results of experiments 1 and 2 suggest that this ability emerges between 6 and 10 months of age. In order to confirm this developmental change, we examined whether 6- and 10-month-old infants exhibited significantly different looking times, analyzing separately the groups presented with speech in 8 and 16 bands. We followed the same procedure as in the analysis split by age.

First, we examined the two age groups presented with the stimuli in eight bands. The initial, most complex model tested the interaction of the fixed effects of Age (6 months, 10 months) and Condition (familiar, novel) on the dependent variable LTs (log-transformed), contained Participant as random factor, and allowed Condition to vary randomly by Participant. The final model included the fixed effect of Age, and the random factor Participant (LTs_log ~ Age + (1 | Participant), data = data, REML = F); see Supporting Information for the results of the final, best-fitting model and the most complex model that converged. The function Anova revealed a significant effect of Age ($\chi 2(1) = 19.4$, $p < 0.001$), due to longer LTs at 6 months (7.92 s) than 10 months (5.72 s; $p < 0.001$), as established by a posthoc analysis using marginal means.

A second analysis examined the two age groups presented with stimuli in 16 bands. The conceptually most complex model was identical to the one used in the analysis of eight bands, while the final model included the fixed effects of Age and Condition, their interaction, and the random factor Participant (LTs_log ~ Age × Condition + (1 | Participant), data = data, REML = F). The function Anova revealed significant effects of Age ($\chi^2(1) = 4.61$, $p = 0.032$) and Condition ($\chi^2(1) = 4.64$, $p = 0.031$), and the significant interaction of the two factors ($\chi^2(1) = 3.91$, $p = 0.048$). We further analyzed this interaction using marginal means, which revealed longer looking times to novels as compared with familiar test items, but only in the group of 10-month-olds (familiar: 6.20 s, novel: 7.56, $p = 0.007$; 6-month-olds: familiar: 8.35, novel: 8.66, $p = 0.900$).

The results of these analyses establish that 10-month-old infants succeed at segmenting the familiarized word forms from fluent speech that has been acoustically degraded and only contains AM information but only when presented in the higher number of spectral bands, that is, 16.

5 | GENERAL DISCUSSION

In two experiments with French-learning infants, we investigated the role of spectro-temporal acoustic cues in infants' ability to segment word forms from continuous speech. The 6- and 10-month-old infants listened to vocoded speech in which FM information had been removed (replaced by pure tones), and AM cues were preserved in a reduced number of spectral bands (either 8 or 16 bands). Previous literature reported that AM cues presented in a small number of bands (i.e., four bands) are enough for adults to extract lexical information, while toddlers require AM cues to be presented in a greater number of bands (8 or 16), depending on the task. Using the headturn preference procedure, we examined whether AM cues extracted in a limited number of spectral bands (either 8 or 16) provide sufficient information for young infants to segment word forms from continuous speech. The results of the two experiments showed that, at 10 months, infants successfully segment the word forms despite the degradation of the speech signal, although only when AM was presented within 16 spectral bands. This establishes that, by 10 months of age, FM temporal cues are not necessary for speech segmentation, and also that the number of spectral bands preserved in the signal does impact infants' ability to extract word forms, as found in previous studies with toddlers examining other areas of lexical processing (Newman & Chatterjee, 2013; Newman et al., 2020). By contrast, we did not find evidence that the 6-month-old infants segmented the target word forms from vocoded speech comprising only AM temporal cues, regardless of the number of spectral bands preserved. From an ecological perspective, FM information (>500 Hz), signals fine phonetic and voice-pitch information, while slower AM fluctuations (<500 Hz) convey rhythm and syllabicity (Rosen, 1992; Smith et al., 2002; Xu et al., 2005). Therefore, in our study, fine phonetic and voice-pitch information are lacking but rhythm and syllabicity are preserved in both vocoder conditions. According to our results, 6-month-olds cannot rely solely on rhythm and syllabicity for word form segmentation, even with 16 spectral bands. However, 10-month-olds are capable of segmenting word forms when rhythm and syllabicity information is provided.

Before discussing this developmental change, we would like to highlight two aspects of the discrimination effect found at 10 months in the 16-band condition, namely, the fact that it was a novelty effect, when most studies with intact speech find a familiarity effect. A number of factors have been shown to change direction of preference in infant studies, including in segmentation tasks (e.g., stimulus complexity, duration of familiarization, expertise acquired with age; Bosch et al., 2013; Hunter & Ames, 1988; Thiessen et al., 2005). As ours is the first study to test infants' segmentation preferences using vocoded speech, we have no precedent with which to compare these results. In a syllable discrimination task, Cabrera et al. (2013) found a novelty effect with intact stimuli, a familiarity effect after a 1-min-long familiarization with spectrally degraded speech (four bands), but a novelty preference after 2 min of familiarization with the same spectrally degraded speech. Hence, direction of preference appears to be governed by (some of) the same factors impacting it in intact speech, but further studies are necessary to establish them.

Coming now to the developmental change found in the present research, it aligns with previous literature showing important changes in infants' phonological knowledge between 6 and 10 months of age (Nishibayashi & Nazzi, 2016; Werker & Tees, 1984). While at 6 months infants are starting to acquire the phonological inventory of their language, by 10 months they are becoming increasingly attuned to it (Werker & Tees, 1984). Moreover, by 10 months, French learners have acquired a number of acoustic-segmental procedures that assist them in speech segmentation (i.e., coarticulation, phonotactics, allophonic variation; Gonzalez-Gomez & Nazzi, 2013; Johnson & Jusczyk, 2001; Jusczyk, Hohne et al., 1999; Mattys & Jusczyk, 2001), and have developed an adult-like phonological bias (i.e., a consonant-bias) in word form segmentation (Nishibayashi & Nazzi, 2016). There are two possible, but not mutually exclusive, explanations for the emergence of segmentation of our vocoded stimuli suggested by the contrastive results found at 6 and 10 months of age.

On the one hand, it could stem from an increase in the number of segmentation cues available to infants between 6 and 10 months. At 6 months, segmentation is driven by prosodic and statistical information. Adult studies showed that removing FM cues and reducing spectral information impact prosodic perception by removing F0 and stress related cues, and also impacts phoneme perception, specifically the perception of place of articulation (Smith et al., 2002; Xu et al., 2005; Zeng et al., 2005). Added to 6-month-old infants' limited knowledge of the consonantal inventory of their native language, these manipulations likely impaired their use of prosodic and statistical information to segment the target word forms. By contrast, at 10 months, infants are more advanced in the acquisition of the phonemes of their native language, can use more diverse segmentation cues, and presumably have somewhat greater lexical knowledge. This wider range of available segmentation procedures would have allowed infants at 10 months to compensate for the degradation of the signal, as long as sufficient spectral information is preserved.

Alternatively or in addition, the different results obtained with the 6- and 10-month-olds could stem from a well-attested phonological bias, which undergoes a developmental shift in infancy. A wealth of literature has shown that consonants play a more salient role than vowels in lexical processing in many languages, including French (Nespor et al., 2003; see Nazzi & Cutler, 2019, for a review). This so-called *consonant bias in lexical processing* emerges quite early in development, as found for French, although the exact age at which it emerges appears to vary across languages (Nazzi & Cutler, 2019). Until 6 months, French learners exhibit the reverse vowel bias, giving more weight to vowels in lexical processing tasks. Thus, French-learning 5-month-olds detect mispronunciations of their own name only when the change occurrs in a vowel (e.g. Alix vs. Elix), but not in a consonant (Victor vs. Zictor; Bouchon et al., 2015), and this differential sensitivity in own name recognition changes between 8 and 11 months of age (Von Holzen & Nazzi, 2020). Similarly, French-learning 6-month-olds consider consonant mispronunciations more similar than vowel mispronunciations to a familiar target segmented from a passage, while at 8 months this preference is reversed, evidencing a shift from a vowel to a consonant bias in early speech segmentation (Nishibayashi & Nazzi, 2016).

Importantly, studies investigating phoneme perception with adult listeners reveal that consonant identification is less affected by spectral degradation, as compared with vowel identification (Xu et al., 2005). The absence of evidence of segmentation in the 6-month-olds might hence result from the combination of the French-learning 6-month-olds giving more weight to vowels to segment speech (Nishibayashi & Nazzi, 2016), and our manipulations of the signal affecting vowels to a greater extent than consonants (Xu et al., 2005), as at this age word recognition is more likely to be disrupted in the case of loss of information in vowels as compared with consonants. Conversely, since, by 10 months, French-learning infants preferentially rely on consonants to segment speech, and consonants were better preserved than vowels in our manipulations, they might have been able to extract enough information from the signal to discriminate between the target and novel words. Indeed, a recent study showed that French-speaking adults exhibit a consonant bias in a lexical decision task when presented with vocoded speech in which FM cues were removed and AM cues were preserved only in eight spectral bands (de la Cruz-Pavía et al., 2023).

One possible way to disentangle whether these two potential explanations are cumulative or mutually exclusive would be to present infants with intact FM and spectral cues but degraded AM cues, as degradation of AM cues has been found to impact consonant identification more than vowel identification in adults (Ardoint & Lorenzi, 2010; Drullman et al., 1994; Xu et al., 2005). If the absence of segmentation observed in the groups of 6-month-olds was caused by concurrent vowel degradation and a vowel bias, we would expect to find evidence of segmentation in this new condition. In turn, consonant degradation might prevent the 10-month-olds from segmenting the signal, unless they are able to compensate due to the wider range of segmentation procedures available to them. Future studies will explore the role of different degradations of AM cues in infants' word form segmentation abilities.

Finally, the current results contribute to a better understanding of the role of auditory processing in speech perception during early development. They are in line with a recent study investigating adults' and infants' reliance on fast temporal modulations of speech in phonetic categorisation (Hegde, Nazzi & Cabrera, 2024). In that study, which used vocoded speech, 6- and 10-month-old infants were more sensitive to the degradation of fast temporal cues (i.e., FM and faster AM cues) than adults when detecting changes in vowel or consonant categories. Additionally, 6- and 10-month-olds relied differently on these fast temporal cues for consonant detection, but not for vowel detection, thus indicating a change in the use of temporal modulation in speech perception during the first year of life, as found in the current study. Furthermore, electrophysiological studies have highlighted a gradual maturation in infants' neural capacity to synchronize with AM cues in nonspeech sounds and to AM cues conveying phonetic contrasts in the first year of life (Lorenzini et al., 2023; Liberto et al., 2023; see Cabrera & Lau, 2022 for a review of temporal processing development). Moreover, research exploring AM detection abilities in older children suggests that although the basic sensory mechanisms for detecting temporal modulation cues mature early, the ability to utilize this information continues to develop between the ages of 5 and 11 years (Cabrera et al., 2019). With the current findings that 10-month-olds require 16 spectral bands of information to segment word forms from continuous speech, it becomes clear that developmental changes in acoustic cue weighting for both low and high-level auditory processing begin during the first year of life, but continue in later months/years. These results call thus for more research, combining neural and behavioral measures, to specify the full developmental trajectory of language-related auditory processing.

6 | CONCLUSIONS

Infants begin to extract word forms from fluent speech between 4 and 7.5 months of age (Berdasco-Muñoz et al., 2018; Jusczyk & Aslin, 1995). In French, this ability is well attested at 6 months (Nishibayashi & Nazzi, 2016). Infants use a number of cues–prosodic, statistical, acoustic-segmental–to achieve this feat. Using a psychoacoustic approach, we examine the impact on word form segmentation of degrading the temporal (AM and FM) and spectral information of the speech signal, that is, the two fundamental acoustic components extracted from speech by the auditory system. In two experiments with French-learning infants using vocoded speech, we show that FM information, that is, the fast temporal modulations of the speech signal, is not necessary for 10-month-old infants to succeed at word segmentation, and that at this age infants can even accommodate a certain reduction in the number of spectral bands available (to 16 bands), when the slow temporal modulations of the signal (i.e., AM information) are preserved. In contrast, 6-month-old infants seemingly cannot recover from these degradations of the signal, as a result of their limited range of segmentation procedures, and/or due to a vowel bias in lexical processing, as vowels are more affected by these acoustic manipulations. These results shed light on the acoustic underpinnings of a crucial step in lexical acquisition.

ACKNOWLEDGMENTS

We wish to thank Maxine dos Santos and Viviane Huet for help in recruiting the participants, Coraline Eloy, Adèle Henensal, Anna Lapteva, and Paula Perrineau-Hecklé for help in running a subset of the participants and Julián Villegas for his help with statistical analysis. This research was supported by the Agence Nationale de la Recherche (ANR), France, under Grant No. ANR-17-CE28-0008 DESIN awarded to Laurianne Cabrera, the ANR's French Investissements d'Avenir—Labex EFL Program under Grant No. ANR-10-LABX-0083 awarded to Thierry Nazzi, Irene de la Cruz-Pavía, and Monica Hegde, and by the Basque Foundation for Science Ikerbasque, the Basque Government under Grant No. IT1439-22, and Grant RYC2021-03395-I funded by MCIN/AEI/10.13039/501100011033 and by "European Union NextGenerationEU/PRTR", all three awarded to Irene de la Cruz-Pavía. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Irene de la Cruz-Pavía https://orcid.org/0000-0003-3425-0596

REFERENCES

Abboub, N., Nazzi, T., & Gervain, J. (2016). Prosodic grouping at birth. *Brain and Language*, *162*, 46–59.

Ardoint, M., & Lorenzi, C. (2010). Effects of lowpass and highpass filtering on the intelligibility of speech based on temporal fine structure or envelope cues. *Hearing Research*, *260*(1–2), 89–95.

Berdasco-Muñoz, E., Nishibayashi, L.-L., Baud, O., Biran, V., & Nazzi, T. (2018). Early segmentation abilities in preterm infants. *Infancy*, *23*(2), 268–287.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science*, *16*(4), 298–304.

Bosch, L., Figueras, M., Teixidó, M., & Ramon-Casas, M. (2013). Rapid gains in segmenting fluent speech when words match the rhythmic unit: Evidence from infants acquiring syllable-timed languages. *Frontiers in Psychology*, *4*, 106.

Bouchon, C., Floccia, C., Fux, T., Adda-Decker, M., & Nazzi, T. (2015). Call me alix, not elix: Vowels are more important than consonants in own-name recognition at 5 months. *Developmental Science*, *18*(4), 587–598.

Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The roots of bilingualism in newborns. *Psychological Science*, *21*(3), 343–348.

Cabrera, L., Bertoncini, J., & Lorenzi, C. (2013). Perception of speech modulation cues by 6-month-old infants. *Journal of Speech, Language and Hearing Research*, *56*(6), 1733–1744.

Cabrera, L., Tsao, F.-M., Gnansia, D., Bertoncini, J., & Lorenzi, C. (2014). The role of spectro-temporal fine structure cues in lexical-tone discrimination for french and mandarin listeners. *The Journal of the Acoustical Society of America*, *136*(2), 877–882.

Cabrera, L., Varnet, L., Buss, E., Rosen, S., & Lorenzi, C. (2019). Development of temporal auditory processing in childhood: Changes in efficiency rather than temporal-modulation selectivity. *The Journal of the Acoustical Society of America*, *146*(4), 2415–2429.

Cabrera, L., & Lau, B. K. (2022). The development of auditory temporal processing during the first year of life. *Hearing, Balance and Communication*, *20*(3), 155–165. https://doi.org/10.1080/21695717.2022.2029092

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *The Journal of the Acoustical Society of America*, *95*(3), 1570–1580.

Christophe, A., Mehler, J., & Sebastián-Gallés, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy*, *2*(3), 385–394.

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical treatment of looking-time data. *Developmental Psychology*, *52*(4), 521–536.

Dau, T., Kollmeier, B., & Kohlrausch, A. (1997a). Modeling auditory processing of amplitude modulation. i. detection and masking with narrow-band carriers. *The Journal of the Acoustical Society of America*, *102*(5), 2892–2905.

Dau, T., Kollmeier, B., & Kohlrausch, A. (1997b). Modeling auditory processing of amplitude modulation. ii. spectral and temporal integration. *The Journal of the Acoustical Society of America*, *102*(5), 2906–2919.

de la Cruz-Pavía, I., Eloy, C., Perrineau-Hecklé, P., Nazzi, T., & Cabrera, L. (2023). Consonant bias in adult lexical processing under acoustically degraded listening conditions. *JASA Express Letters*, *3*(5), 055206.

Di Liberto, G. M., Attaheri, A., Cantisani, G., Reilly, R. B., Ní Choisdealbha, Á., Rocha, S., Brusini, P., & Goswami, U. (2023). Emergence of the cortical encoding of phonetic features in the first year of life. *Nature Communications*, *14*(1). https://doi.org/10.1038/s41467-023-43490-x

Drullman, R., Festen, J. M., & Plomp, R. (1994). Effect of reducing slow temporal modulations on speech reception. *The Journal of the Acoustical Society of America*, *95*(5), 2670–2680.

Fló, A., Benjamin, L., Palu, M., & Dehaene-Lambertz, G. (2022). Sleeping neonates track transitional probabilities in speech but only retain the first syllable of words. *Scientific Reports*, *12*(1), 4391.

Glasberg, B. R., & Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*(1-2), 103–138.

Gonzalez-Gomez, N., & Nazzi, T. (2013). Effects of prior phonotactic knowledge on infant word segmentation: The case of nonadjacent dependencies. *Journal of Speech, Language, and Hearing Research*, *56*(3), 840–849.

Gout, A., Christophe, A., & Morgan, J. L. (2004). Phonological phrase boundaries constrain lexical access ii. Infant data. *Journal of Memory and Language*, *51*(4), 548–567.

Goyet, L., Nishibayashi, L.-L., & Nazzi, T. (2013). Early syllabic segmentation of fluent speech by infants acquiring french. *PLoS One*, *8*(11), e79646.

Hegde, M., Nazzi, T., & Cabrera, L. (2024). An auditory perspective on phonological development in infancy. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1321311

Hoareau, M., Yeung, H. H., & Nazzi, T. (2019). Infants' statistical word segmentation in an artificial language is linked to both parental speech input and reported production abilities. *Developmental Science*, *22*(4), e12803.

Höhle, B., & Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Developmental Science*, *6*(2), 122–127.

Houston, D. M., Jusczyk, P. W., Kuijpers, C., Coolen, R., & Cutler, A. (2000). Cross-language word segmentation by 9-month-olds. *Psychonomic Bulletin & Review*, *7*, 504–509.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*, *5*, 69–95.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, *44*(4), 548–567.

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, *13*(2), 339–345.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1–23.

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics*, *61*, 1465–1476.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in english-learning infants. *Cognitive Psychology*, *39*(3-4), 159–207.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608.

Lorenzini, I., Labendzki, P., Basire, C., Hababou-Bernson, M., Calcus, A., & Cabrera, L. (2023). Neural processing of auditory temporal modulations in awake infants. *The Journal of the Acoustical Society of America*, *154*(3), 1954–1962. https://doi.org/10.1121/10.0020845

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*(2), 91–121.

Mersad, K., & Nazzi, T. (2012). When mommy comes to the rescue of statistics: Infants combine top-down and bottom-up cues to segment speech. *Language Learning and Development*, *8*(3), 303–315.

Moore, B. C. (2003). Speech processing for the hearing-impaired: successes, failures, and implications for speech mechanisms. *Speech Communication*, *41*(1), 81–91.

Moore, B. C. (2012). *An introduction to the psychology of hearing*. Brill.

Moore, D. R. (2002). Auditory development and the role of experience. *British Medical Bulletin*, *63*(1), 171–181.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 756.

Nazzi, T., & Cutler, A. (2019). How consonants and vowels shape spoken-language recognition. *Annual Review of Linguistics*, *5*, 25–47.

Nazzi, T., Dilley, L. C., Jusczyk, A. M., Shattuck-Hufnagel, S., & Jusczyk, P. W. (2005). English-learning infants' segmentation of verbs from fluent speech. *Language and Speech*, *48*, 279–298.

Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior and Development*, *21*(4), 779–784.

Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring french: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, *54*(3), 283–299.

Nazzi, T., Mersad, K., Sundara, M., Iakimova, G., & Polka, L. (2014). Early word segmentation in infants acquiring Parisian French: task-dependent and dialect-specific aspects. *Journal of Child Language*, *41*(3), 600–633.

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e linguaggio*, *2*(2), 203–230.

Newman, R., & Chatterjee, M. (2013). Toddlers' recognition of noise-vocoded speech. *The Journal of the Acoustical Society of America*, *133*(1), 483–494.

Newman, R. S., Morini, G., Shroads, E., & Chatterjee, M. (2020). Toddlers' fast-mapping from noise-vocoded speech. *The Journal of the Acoustical Society of America*, *147*(4), 2432–2441.

Nishibayashi, L.-L., Goyet, L., & Nazzi, T. (2015). Early speech segmentation in french-learning infants: Monosyllabic words versus embedded syllables. *Language and Speech*, *58*(3), 334–350.

Nishibayashi, L.-L., & Nazzi, T. (2016). Vowels, then consonants: Early bias switch in recognizing segmented word forms. *Cognition*, *155*, 188–203.

Polka, L., & Sundara, M. (2012). Word segmentation in monolingual infants acquiring canadian english and canadian french: Native language, cross-dialect, and cross-language comparisons. *Infancy*, *17*(2), 198–232.

R Core Team. (2019) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *336*(1278), 367–373.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*(5234), 303–304.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, *416*(6876), 87–90.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*, 21.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53–71.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental Psychology*, *39*(4), 706.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development*, *47*(2), 466–472.

Von Holzen, K., & Nazzi, T. (2020). Emergence of a consonant bias during the first year of life: New evidence from own-name recognition. *Infancy*, *25*(3), 319–346.

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.

Xu, L., Thompson, C. S., & Pfingst, B. E. (2005). Relative contributions of spectral and temporal cues for phoneme recognition. *The Journal of the Acoustical Society of America*, *117*(5), 3255–3267.

Zeng, F.-G., Nie, K., Stickney, G. S., Kong, Y.-Y., Vongphoe, M., Bhargave, A., Wei, C., & Cao, K. (2005). Speech recognition with amplitude and frequency modulations. *Proceedings of the National Academy of Sciences*, *102*(7), 2293–2298.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: de la Cruz-Pavía, I., Hegde, M., Cabrera, L., & Nazzi, T. (2024). Infants' abilities to segment word forms from spectrally degraded speech in the first year of life. *Developmental Science*, *27*, e13533. https://doi.org/10.1111/desc.13533

APPENDIX 1

Pilot experiment

The present pilot experiment is based on the task originally designed by Nishibayashi and Nazzi (2016), in which infants are familiarized with *two* different target words (as opposed to one, as in experiments 1 and 2 above) embedded into passages using vocoded speech, and then tested on occurrences of the two familiar targets mixed with two previously unheard vocoded words.

METHODS

Participants

Twenty-four 6-month-old infants participated in the experiment (12 girls, 12 boys; mean age: 6;15; age range: 6;00–6;30). All infants were born full-term and were being raised in monolingual French-speaking families. Data from seven additional infants were excluded from analysis due to equipment failure (two infants), experimenter error (one infant), insufficient looking times (two infants), or mean looking times 2.5 SD above or below the group mean (two infants). Parents gave informed consent before their infant's participation.

Stimuli

The stimuli differed from those of experiments 1 and 2 in that they consisted of eight (rather than four) monosyllabic words. As in experiments 1 and 2, these were selected from the set originally selected and used by Nishibayashi and Nazzi (2016). The eight words were arranged into four pairs, such that words in a pair differed in their consonants' voicing and place of articulation, and in the roundness and place of articulation of their vowels (/vo/ *veau* "calf" and *vos* "your" (plural)—/py/ *pus* "pus," and *pu*—"stink" and past participle of the verb "can"; /vø/ *vœu* "wish"—/pu/ *poux* "louse"; /fo/ *faux* "fake"—/by/ *but* "aim"; /fø/ *feu* "fire"—/bu/ *bout* "piece").

We then processed the stimuli with a vocoder, using the exact same parameters as in experiments 1 and 2′s 8-band group. That is, FM cues were replaced by pure tone carriers, while AM cues were preserved, and the signal was passed through eight filters, resulting in eight spectral bands.

Procedure

While the apparatus was identical, procedure differed from experiments 1 and 2. Following Nishibayashi and Nazzi (2016), infants listened to two passages in familiarization, each containing a different target word. The two passages were presented in semirandom order until infants accumulated 30 s of looking time to each. Once an infant reached this criterion for one of the passages, the second one was played repeatedly until the criterion was reached. Infants were tested with one of four lists, each corresponding to a different pair of target words. During the test phase, infants were presented with four different trials, each corresponding to 20 repetitions of a single word (the two target words, i.e., those presented in the passages heard in familiarization, or the two control words, i.e., previously unheard words). These four trials were repeated in three test blocks. Each of the eight words that comprised the stimuli occurred with equal frequency as target and control across infants.

RESULTS

The raw LTs were log-transformed prior to analysis. We fit linear mixed effects models (lme4 package). We started by fitting the conceptually most complex model, which explained the dependent variable Looking Times, with the interaction of the fixed effect of Condition (familiar, novel), and Participant as a random factor, and allowed Condition to vary randomly by Participant. We then built decreasingly complex models, comparing pairs of models using ANOVAs. The simplest model that fit the data included only the random factor Participant (modelLTs3 < - lmer(LTs ~ + (1 | Participant))); the results of the model and the results of the most complex model that converged are reported in the Supporting Information).

In sum, these results fail to show that the 6-month-olds segmented target word forms from vocoded speech in which only AM cues were preserved, while spectral information was reduced to eight bands and FM cues were removed.

APPENDIX 2

Passages used in the familiarization phase of experiments 1 and 2 and the pilot experiment reported in Appendix 1. Passages were originally created by Nishibayashi and Nazzi (2016, Appendix 2)

Note: Only the passages containing the target words /vo/, /py/, /fø/, and /bu/ were used in experiments 1 and 2; all passages were used in the pilot experiment reported in Appendix 1.

/vo/: *Le veau de lait est délicieux en cette saison. J'aurais préféré un ris de veau poivré. Ce veau vient de naître dans la ferme voisine. La vache vient de mettre bas d'un petit veau blanc. Un veau gambade dans le nouvel enclos de Georges. Le jeune fermier s'occupe de ses veaux avec amour.*

/py/: *Du pus sortit de ses nombreuses plaies béantes. Les enquêteurs ont trouvé des traces de pus séché. Son pus a été enlevé par des antibiotiques. Tes blessures ne montrent plus un pus jaunâtre. Le pus se présenta de l'éraflure de tout le coude. Ils ont toujours prélevé leur pus au centre.*

/fø/: *Ce feu a pris au troisième et quatrième étage. J'ai vu de belles choses avec des feux de Bengale. Un feu s'est déclaré en Inde et au Pakistan. L'incendie est provoqué par le feu d'une poubelle. Trop de feu dénature la qualité de la viande. Le taxi n'a pas attendu son feu rouge.*

/bu/: *Le bout du tunnel n'est pas si loin que ça. Ils mettent sur le côté tous leurs bouts durs. Des bouts de robinet traînent par terre chez lui. Elle m'a volontiers donné son bout rassis. Ce bout en bois s'est enflammé tout de suite. Les délégués lui ont remis un bout mural.*

/vø/: *Mon vœu le plus cher est la paix dans le monde. Leurs dieux pourront accorder trois vœux innocents. Ce vœu est réalisable par le mage bleu. Tu vas avoir droit à des vœux incroyables. Un vœu sera souhaité au cours de la nuit. Les génies donnent le choix entre deux vœux magiques.*

/pu/: *Les poux sont redoutés dans les maternelles. Je sais comment enlever ces poux hargneux. Le pou est très résistant aux shampoings normaux. Tu aurais dû éviter d'attraper des poux bruns. Leurs poux ne sont plus qu'un mauvais souvenir. Il a toujours eu la force d'un pou de combat.*

/fo/: *Ce faux a été découvert au musée. Il est difficile de distinguer le faux du vrai. Son faux fut démasqué par la police. Je veux savoir lequel est un faux tout de suite. Leur faux est bien dissimulé dans les bois. Interpol a mis la main sur ton faux hier.*

/by/: *Le but est de trouver le dernier élément. Je ne sais pas si j'atteindrai mon but premier. Son but semble bien malhonnête à première vue. Ils ne savent jamais quels sont leurs buts nobles. Ce but n'est pas louable dans cet hôpital. Vous ne voulez pas connaître ces buts cachés.*