Cognition 244 (2024) 105663

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Abstract processing of syllabic structures in early infancy

Chiara Santolin [a,*], Konstantina Zacharaki [a,b], Juan Manuel Toro [a,c], Nuria Sebastian-Galles [a]

[a] *Center for Brain and Cognition, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Carrer Ramon Trias Fargas 25-27, 08005, Barcelona, Spain*
[b] *ESADE Business School, Ramon Llull University, Avenida de Pedralbes, 60-62, 08034, Barcelona, Spain*
[c] *Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluis Companys, 23, 08010, Barcelona, Spain*

ARTICLE INFO

Keywords:
Syllable
Linguistic structure
Abstract processing
Learning
Infancy

ABSTRACT

Syllables are one of the fundamental building blocks of early language acquisition. From birth onwards, infants preferentially segment, process and represent the speech into syllable-sized units, raising the question of what type of computations infants are able to perform on these perceptual units. Syllables are abstract units structured in a way that allows grouping phonemes into sequences. The goal of this research was to investigate 4-to-5-month-old infants' ability to encode the internal structure of syllables, at a target age when the language system is not yet specialized on the sounds and the phonotactics of native languages. We conducted two experiments in which infants were first familiarized to lists of syllables implementing either CVC (consonant-vowel-consonant) or CCV (consonant-consonant-vowel) structures, then presented with new syllables implementing both structures at test. Experiments differ in the degree of phonological similarity between the materials used at familiarization and test. Results show that infants were able to differentiate syllabic structures at test, even when test syllables were implemented by combinations of phonemes that infants did not hear before. Only infants familiarized with CVC syllables discriminated the structures at test, pointing to a processing advantage for CVC over CCV structures. This research shows that, in addition to preferentially processing the speech into syllable-sized units, during the first months of life, infants are also capable of performing fine-grained computations within such units.

1. Introduction

A long tradition of research shows that syllables are one of the primary processing units of speech from the earliest stages of language acquisition in infants. Newborns discriminate speech sequences (bi or tri-syllabic utterances) differing in the number of syllables rather than in the number of phonemes, and show no sensitivity to variations in the number of phonemes composing bi-syllabic utterances (Bijeljac-Babic, Bertoncini, & Mehler, 1993). This evidence suggests preferential tracking of syllables rather than phonemes at birth (see also Bertoncini, Floccia, Nazzi, & Mehler, 1995). Additionally, newborns automatically synchronize their brain activity to the syllable frequency rate (Fló, Benjamin, Palu, & Dehaene-Lambertz, 2022) in a similar way as it is observed in adults (Poeppel, 2003), and are sensitive to violations of universal constraints on syllable structure, discriminating syllables like *blif* and *lbif* (Gómez et al., 2014). Other studies reveal that 2-month-old infants form speech representations that most likely match syllables rather than phonemes (Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, & Mehler, 1988; Jusczyk & Derrah, 1987). These studies showed that infants familiarized with syllables sharing a common phoneme, such as a vowel as in *bi* or *mi*, did not seem to encode the common phoneme as a shared property among the syllables. In fact, infants equally discriminated test syllables having both a new vowel and a new consonant, suggesting that they did not categorize individual phonemes but, rather, the whole syllabic unit (see also Eimas, 1999 for results on 3–4-month-olds). Although some phonetic knowledge seems to be available in the first months of life (Gennari, Marti, Palu, Fló, & Dehaene-Lambertz, 2021), this set of evidence indicates that early processing and representation of speech are better structured in units that roughly correspond to syllables. Such units are key building blocks for further linguistic processing (e.g., Friederici, 2005; Werker, 2018). Interestingly, the current literature leaves open the possibility that, at such a young age, infants may be able to perform more complex computations on syllables than those described so far. In the present study, we aim at advancing this line of research, exploring whether young infants can encode the internal structure of syllables.

* Corresponding author at: Centre for Brain and Cognition, University Pompeu Fabra, c/ Ramon Trias Fargas 25-27, Barcelona 08005, Spain.
E-mail address: chiara.santolin@upf.edu (C. Santolin).

https://doi.org/10.1016/j.cognition.2023.105663
Received 27 September 2021; Received in revised form 2 November 2023; Accepted 3 November 2023
Available online 20 December 2023

In linguistics, syllables are defined as abstract constituents with hierarchical structures,[1] whose main function is grouping speech segments into sequences (e.g., Halle & Vergnaud, 1980; Hockett, 1955). The structure of syllables typically comprises a nucleus, the mandatory element, and onset and coda, optional elements located at the edges of the syllable. The identity of these components is mainly determined by the Sonority Sequencing Principle (SSP), which imposes restrictions on how phonemes (consonants and vowels) are combined into syllables based on their intensity. According to the SSP, the syllable nucleus comprises phonemes with the highest intensity, usually vowels, whereas onset and coda include phonemes with lower intensity, typically consonants (e.g., Selkirk, 1984; Clements, 1990; De Lacy, 2007; Parker, 2008; Prince & Smolensky, 1993/2004; see Ridouane, 2008 for alternatives). Such ordering creates temporal variations in the signal that allow listeners to chunk sounds into syllables.

A handful of studies show that young infants can form quite detailed representations of the phonemes composing syllables, and the patterns involving such phonemes. Four-month-old infants track sound patterns defining consonant-vowel (CV) syllables (e.g., fricative consonants always followed nasal vowels), and recognize such patterns in new syllables (Cristia, Seidl, & Gerken, 2011). Similarly, 3-month-olds categorize syllables with a constant phoneme (e.g., *bi*, *be*, *ba*, etc.) associating them with a single visual object, and associating syllables starting with a different phoneme (e.g., *di*, *de*, *da*) with a different object (Mersad & Dehaene-Lambertz, 2016; Mersad, Kabdebon, & Dehaene-Lambertz, 2021). Little is known, however, about whether or not infants around this age are able to form abstract representations of the syllabic structure as a whole, regardless of its individual constituents (phonemes or combinations of phonemes).

The present research is aimed at exploring this issue, investigating whether 4-to-5-month-old infants can track the internal structure of syllables (which determines the ordering of its constituents) and recognize such structure instantiated in novel syllables. We designed two experiments in which infants were familiarized with a list of either CVC or CCV syllables (for example, *bor, kos* as CVC, *plu, spa* as CCV) then tested on their ability to discriminate between these same two structures instantiated in new syllables. The experiments differ in the degree of phonological similarity between syllables presented at familiarization and test. In Experiment 1, stimuli share a higher degree of phonological overlap between familiarization and test; in Experiment 2, the phonological similarity between stimuli is minimal, thus providing a more stringent test of generalization. Note that infants are good at generalizing patterns implemented with speech sounds to novel exemplars (e.g., Marcus, Vijayan, Bandi Rao, & Vishton, 1999), and do so from birth (Gervain, Macagno, Cogoi, Peña, & Mehler, 2008; see Saffran & Kirkham, 2018 for a review). If infants are able to form a representation of the syllabic structures they hear at familiarization (of CVC and CCV), they should discriminate between test syllables. We selected a target age range that precedes the establishment of native phoneme categories (Kuhl, 1979; Werker & Tees, 1984) and phonotactic knowledge of the language(s) of exposure (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Mattys & Jusczyk, 2001). This allows us to investigate general sensitivities to syllabic structures before narrowing to language-specific features takes place.

2. Materials & methods

2.1. Experiment 1

2.1.1. Participants

Forty-eight infants of 4.5- to- 5.5 months of age participated in the study. Twenty-four infants participated in the CVC condition (mean age: 148 days), and 24 infants in the CCV condition (mean age: 150 days). Nineteen infants were excluded from the final sample because of difficulty coding head-turns (6), fussiness/crying (5), test performance >2 standard deviations from the mean (4), parental interference (2), data not meeting the inclusion criterion (set prior to data collection) of at least 8 trials with looking times >2 s (1), and technical error (1). Forty-one out of 48 infants were exposed to Spanish and Catalan, either exclusively/predominantly to one of them (classified as monolinguals) or to the two of them (classified as bilinguals; Bosch & Sebastian-Galles, 2001; Sebastian-Galles & Santolin, 2020). Seven infants were exposed to another language in combination with Catalan or Spanish: French (3 participants), Italian (2), Portuguese (1), Galician (1). All infants were healthy and full term, with no history of hearing or vision problems. The study was conducted in accordance with ethical standards of the Declaration of Helsinki and the protocol was approved by the local ethical committee (Parc de la Salut Mar, Barcelona). Caregivers signed a consent form before participating in the experiment and a small gift was given to them at the end of the experiment.

2.1.2. Stimuli

Ten CVC syllables and ten CCV syllables were used for the familiarization lists. Syllables were generated by combining phonemes such that onsets remained constant between syllable types (CVC and CCV) and all syllables adhered to the SSP (defined above). Since humans are sensitive to violations of the SSP at birth (Gómez et al., 2014), it was crucial to take it into account when designing our stimuli. Syllables were recorded by a female speaker, and concatenated into six strings. Syllables appeared once within each string, and were presented in random order. Strings were concatenated into a 3-min familiarization list, in which each string was repeated twice. Each syllable appeared twelve times during familiarization. There was 1-s silence break between syllables; consecutive repetitions of trials with the same syllable were avoided. See Table 1 for familiarization and test syllables. A test trial consisted of an individual syllable repeated twelve times maximum (see Section 2.3 for information about trial duration). There were twelve test trials, three for each of the four test items, presented in random order. Consecutive repetitions of the same trial were avoided. All infants heard the same test trials regardless of the familiarization condition. In half of the test trials, CVC test items were presented (whose structure was familiar for infants familiarized with CVC, and unfamiliar for infants familiarized with CCV); in the other half of the test trials, CCV test items were presented (whose structure was familiar for infants familiarized with CCV, and unfamiliar for infants familiarized with CVC). Strings and lists of syllables were created using Praat (version 6.0.20). Duration (0.458 s) and intensity (65 dB) of individual syllables were normalized across stimuli.

2.1.3. Procedure

We used the Head-turn Preference Procedure. Infants were seated on

Table 1
Stimuli used at Familiarization and Test in Experiment 1 and 2.

	Experiment 1		Experiment 2	
	CVC	CCV	CVC	CCV
Familiarization	*bor*	*bro*	*bor*	*bro*
	dor	*dro*	*kin*	*kni*
	fil	*fli*	*sam*	*sma*
	fur	*fru*	*gel*	*gle*
	gol	*glo*	*pus*	*psu*
	kal	*kla*	*dor*	*dro*
	ker	*kre*	*sen*	*sne*
	pul	*plu*	*sum*	*smu*
	tar	*tra*	*kas*	*ksa*
	bel	*ble*	*bil*	*bli*
Test	*pal*	*pla*	*sap*	*spa*
	tor	*tro*	*kos*	*kso*

[1] In this article, we refer to 'phonological syllable' as defined by Hockett (1955) and Halle and Vergnaud (1980).

a caregiver's lap in a soundproof booth equipped with three computer screens while the caregiver listened to music over headphones to avoid interferences with the infant's behavior. Screens were placed in front (at 125 cm distance) and on the two sides (at 100 cm distance each) of the infant. The familiarization phase had a fixed duration of 3 min, during which infants were presented with either the CVC or the CCV list. During the first 2 min of familiarization, infants heard the stimuli and watched a video-clip displaying neutral images (i.e., slow-moving clouds) on the central screen. During the last minute, a colored pinwheel moved across the screens to give infants the opportunity to practice head-turns to each side (e.g., Bouchon, Floccia, Fux, Adda-Decker, & Nazzi, 2015; Schott, Mastroberardino, Fourakis, Lew-Williams, & Byers-Heinlein, 2021)[2]. At the beginning of each test trial, the pinwheel was displayed on the central screen until the infant fixated on it. At that point, a trained experimenter (blind to the stimuli) terminated the central pinwheel and triggered the appearance of one of the side-pinwheels. As soon as the infant made a head-turn towards the side pinwheel, one of the test items was played until the infant looked away for >2 s, or until 24 s had elapsed (maximum trial duration). Once a trial ended, the central pinwheel reappeared. Test items were delivered by loudspeakers placed closed to the side screens. The procedure was set up using WISP, a custom-designed MATLAB software (Olson, 2017), which was also used to code looking times (see, for example, Thiessen & Saffran, 2007; Lew-Williams, Pelucchi, & Saffran, 2011; Santolin & Saffran, 2019, for similar procedures). Looking time data were analyzed using R (version 4.2.2; RStudio Team, 2019). Stimuli and datasets are available at osf.

2.1.4. Results

We ran a 2 × 2 mixed ANOVA with condition (familiarization to CVC vs. CCV) as a between-subjects factor, and test structure (CVC vs. CCV) as a within-subjects factor. The critical interaction *condition by test structure* reached statistical significance ($F(1,46) = 8.812$, $p = .005$, $\eta^2G = .066$, suggesting that infants' looking times at test differed as a function of familiarization (see Fig. 1). Main effects were not statistically significant (*test structure*: $F(1,46) = .079$, $p = .780$, $\eta^2G = .000$; *condition*: $F(1,46) = 2.155$, $p = .149$, $\eta^2G = .029$).

Looking times within each condition were examined with paired-samples *t*-tests. Infants familiarized with CVC syllables looked for a significantly longer amount of time to CVC test syllables than to CCV test syllables ($t(23) = 2.750$, $p = .011$, $d = .561$). In contrast, infants familiarized with CCV failed to show a significant familiarity preference ($t(23) = 1.666$, $p = .109$, $d = -.340$). See Table 2 for average looking times for each condition.

2.2. Experiment 2

2.2.1. Participants

Forty-eight infants of 4.5- to- 5.5 months of age participated in the study. Twenty-four infants participated in the CVC condition (mean age: 152,2 days), and 24 infants in the CCV condition (mean age: 144,73 days). Seventeen infants were excluded from the final sample because of difficulty coding head-turns (3), fussiness/crying (4), data not meeting the inclusion criterion (set prior to data collection) of at least 8 trials with looking times >2 s (7), and technical error (3). Thirty-two out of 48 infants were exposed to Spanish and Catalan, either exclusively/predominantly to one of them (classified as monolinguals) or to both (classified as bilinguals). Twelve infants were exposed to another language in combination with Catalan or Spanish: English (3 participants), French (3), Russian (2), Tagalog (1), Bulgarian (1), Czech (1). Three infants were exposed to German and Dutch (1), German only (1), English and Italian (1). One infant was exposed to Catalan, English and Arabic. All infants were healthy and full term, with no history of hearing or vision problems. As for Experiment 1, the study was conducted in accordance with ethical standards (Declaration of Helsinki) and the protocol was approved by the local ethical committee (Parc de la Salut Mar, Barcelona). Caregivers signed a consent form before participating in the experiment and a small gift was given to them at the end of the experiment.

2.2.2. Stimuli

Ten CVC syllables and ten CCV syllables were used for the Familiarization lists (see Table 1). Syllables were created such that the identity of phonemes in onset, nucleus and coda positions in familiarization and test syllables was not a cue for discrimination. The same phoneme could not appear more than three times in the same position in the familiarization list. For instance, /s/ appeared twice in coda position in the CVC list, and twice in onset position in the CCV list (see Table 1). Given such low frequency, it is very unlikely that /s/ could have been informative to disentangle test syllables *kos* and *kso*. To achieve this balance, we used a larger pool of phoneme combinations to generate syllables. Importantly, our stimuli did not violate the Sonority Sequencing Principle. All the other features of the stimuli, and the structure of test trials were identical to Experiment 1.

2.2.3. Procedure

Identical to Experiment 1.

2.2.4. Results

We ran the same analyses as Experiment 1. The 2 × 2 mixed ANOVA showed a significant interaction *condition by test structure* ($F(1,46) = 4.267$, $p = .044$, $\eta^2G = .023$) suggesting that infants' looking times at test differed as a function of familiarization (see Fig. 1). The critical interaction was examined with paired-sample *t*-tests, which revealed that infants familiarized with CVC syllables looked for a significantly longer amount of time to CCV test syllables than to CVC test syllables ($t(23) = -2.169$, $p = .040$, $d = -.443$), thus showing a novelty preference. Infants familiarized with CCV syllables did not show discrimination at test ($t(23) = -.594$, $p = .558$, $d = .121$). See Table 2 for average looking times. The main effect of *condition* was also statistically significant $F(1,46) = 5.671$, $p = .021$, $\eta^2G = .083$) which could be driven by the overall longer looking times showed by infants of the CCV condition.

3. Discussion

We investigated the ability of 4-to-5-month-old infants to encode and generalize the internal structure of syllables. Overall, infants succeeded in the task, being able to tease apart syllabic structures at test, even when test syllables were implemented by new combinations of phonemes. In both experiments, the significant interaction between condition (familiarization to CVC or CCV) and test structure (CVC and CCV) shows that infants behaved differently based on the syllabic structure to which they were familiarized. Only infants familiarized to CVC successfully discriminated novel test syllables. We suggest that such CVC asymmetry may reflect a processing advantage for CVC over CCV structures, as CVC is more frequent than CCV across many languages (e.g., Frota, Freitas, Vigário, & Martins, 2005; Gorman & Gillam, 2003; Harris, 1983; Kenstowicz, 1986; Lass, 1984; Levelt, Schiller, & Levelt, 2000; Potet, 1995; Prieto Vives & Bosch Baliarda, 2006). As far as we are aware, the languages our infants were exposed to show such asymmetry.

To the best of our knowledge, this research provides the first evidence of abstract processing of syllabic structure in the first few months of life. To succeed in the task, infants must have encoded some abstract properties of the syllables, recognizing their internal structure regardless of the phoneme combination used to implement them. In Experiment 1, infants may have used the overlap between some phonemes composing familiarization and test syllables as a cue to recognize the familiar structure at test (e.g., /r/ and /l/ in final position in CVCs in

[2] The subdivision of the 3 familiarization minutes was slightly different for 4 infants, who spent 45 s (rather than 1 min) practicing head-turns. Their average number of head-turns (3.25) was almost identical to the other infants (3.29).

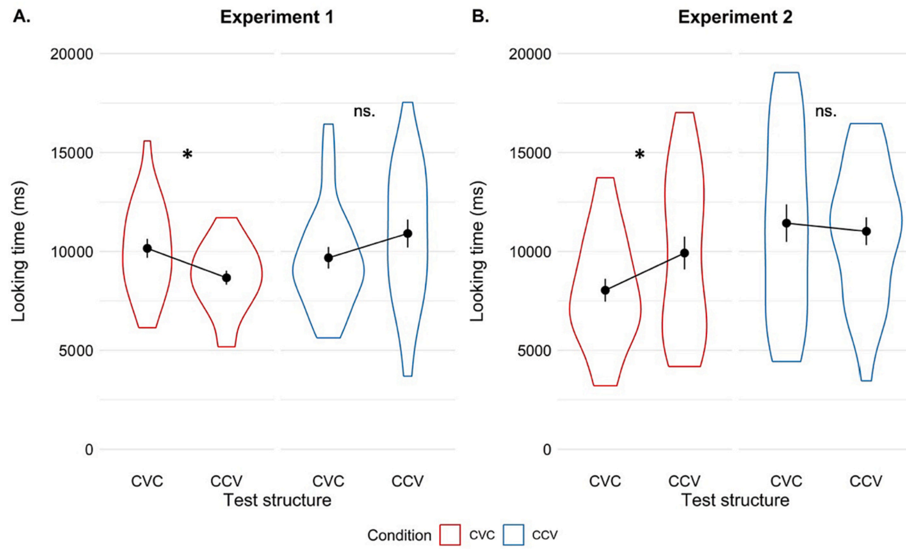

Fig. 1. Average looking times at test for Experiment 1 (A) and Experiment 2 (B), for each condition. Red violins represent looking times of infants familiarized with CVC syllables. Blue violins represent looking times of infants familiarized with CCV syllables. Black dots indicate group means, error bars indicate standard errors. Asterisks indicate significant differences within conditions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Summary of looking times

		Experiment 1		Experiment 2	
Condition		CVC	CCV	CVC	CCV
Test structure	CVC	**10.160** (2.321)	**9.678** (2.655)	**8.039** (2.820)	**11.431** (4.622)
	CCV	**8.671** (1.715)	**10.910** (3.451)	**9.922** (4.049)	**11.018** (3.423)

Average looking times at Test divided by experiment (1 and 2) and condition (the syllabic structure to which infants were familiarized, CVC or CCV). Values are seconds (standard deviations).

familiarization and test stimuli). Experiment 2 was designed to test this hypothesis using a new set of syllables in which phoneme position in familiarization and test stimuli was better balanced, and unlikely to provide a cue for discrimination. Results of Experiment 2 validate our hypothesis, and support Experiment 1's as infants recognized the familiar syllabic structure even when the phonological overlap between familiarization and test syllables was minimal, thus showing generalization.

What type of information did infants use to track syllabic structures? It is possible that infants processed temporal variations of intensity (sonority) that determine the position of phonemes within syllables. In CVC, there is a rise in sonority from onset (consonant) to nucleus (vowel), and a fall from nucleus to coda (another consonant) whereas in CCV there is just a rise from onset to nucleus. Infants may have picked up on these patterns, and recognized them in the new set of test syllables. This possibility is supported by research showing that newborns are sensitive to sonority variations in speech stimuli (Gómez et al., 2014). It is also plausible that infants detected and generalized just a portion of the syllabic structure rather than the whole syllable to succeed in our task. At this early age, vowels (mostly located at the nucleus) have a special status for infants who mainly rely on vocalic information to initially parse the language input (Mehler, Dupoux, Nazzi, & Dehaene-Lambertz, 1996; Zacharaki & Sebastian-Galles, 2021, 2022) and recognize words (Benavides-Varela, Hochmann, Macagno, Nespor, & Mehler, 2012; Bouchon et al., 2015). Another possibility is that infants may have detected changes in duration between vocalic and consonantal intervals. Such intervals indicate the onset and offset of vowels and consonants (and their clusters) composing a syllable (Ramus, Nespor, & Mehler, 1999; see also Gasparini, Langus, Tsuji, & Boll-Avetisyan, 2021). For instance, CCV has a longer consonantal interval in the onset with respect to CVC. It is possible that infants were sensitive to such patterns and used them to discriminate at test. Our results pave the way to further studies aimed at disentangling all these possibilities.

We observe an opposite direction of preference through which infants expressed discrimination across experiments. Interpreting direction of preference in infant studies is notoriously challenging. Both familiarity and novelty preferences coexist in the literature, within the same research areas and even when the same experimental paradigm is used (e.g., Houston-Price & Nakai, 2004; Hunter & Ames, 1988; Kosie et al., 2023; Santolin, Garcia-Castro, Zettersten, Sebastian-Galles, & Saffran, 2021). Factors such as infants' age, familiarization duration and infants' prior experience with the laboratory setting cannot account for our results as they were constant across experiments. In an attempt to explore this difference in direction of preference across experiments, we ran an exploratory analysis of the familiarization phase of Experiments 1 and 2. We offline coded infants' looking times at the screens while listening to the familiarization materials as a proxy of infants' engagement with the task. We found no differences in the proportion of looking times at the screens, and in the number of times infants disengaged from the screens, across experiments and conditions (*condition by experiment* interaction on the proportion of looking time of the 3-min familiarization: $F(1, 82) = .61$, $p = .437$, $\eta^2G < .01$).[3] These results indicate no a

[3] For more information about analysis, scripts and dataset please see the Supplementary Material.

priori differences between conditions (i.e., between being familiarized with CVC and CCV), and suggest that differences at test can be attributed to learning occurring during familiarization. A possible explanation of the novelty preference obtained in Experiment 2 could be related to the increased variability of the materials. Most of the phoneme combinations implementing test syllables were not experienced during familiarization, and may have attracted infants' interest (i.e., looking time) for the newest type of stimulus (CCV).

Our study extends previous research demonstrating that infants are highly sensitive to syllabic units very early in development (e.g., Eimas, 1999; Jusczyk, Kennedy, & Jusczyk, 1995), which is interpreted as showing a holistic representation of the speech (see Halle & Cristia, 2012 for a review). In the current study, we took a further step and showed that 4-to-5 month-olds can encode and generalize the internal structure of syllables, being able to disentangle new syllables implementing such structure from other syllables instantiating a different structure. Our results also point to a processing advantage for CVC. The present research shows that, in addition to preferentially processing the speech into syllable-sized units, during the first months of life, infants are also capable of performing fine-grained computations *within* such units.

Previous studies show that infant sensitivity to syllables emerges with little language experience, and before important milestones of language acquisition are achieved (e.g., Bertoncini & Mehler, 1981; Bijeljac-Babic et al., 1993; Jusczyk & Derrah, 1987). In parallel, research on other species (e.g., Chen & Ten Cate, 2015; Takahasi, Yamada, & Okanoya, 2010; Toro & Trobalón, 2005) and computational modelling (e.g., Doyle & Levy, 2013; Frank, Goldwater, Griffiths, & Tenenbaum, 2010) suggest that syllable-sized units can be used for the organization and segmentation of speech in the absence of language knowledge. Our findings add up to the large body of work showing that some of the general mechanisms at the ontogeny of language acquisition are shared with other animals (e.g., Petkov & Ten Cate, 2020; Rey, Minier, Malassis, Bogaerts, & Fagot, 2019; Saffran et al., 2008; Santolin, Rosa-Salva, Regolin, & Vallortigara, 2016; Santolin, Rosa-Salva, Vallortigara, & Regolin, 2016; Sonnweber, Ravignani, & Fitch, 2015; Toro, 2016; Toro, Trobalón, & Sebastián-Gallés, 2003). We interpret the results of the current research within a similar framework. If infants are using changes in sonority or in sound duration (vocalic/consonantal intervals) to detect and generalize the internal structure of syllables, it is plausible that this early stage of syllable processing does not necessarily require much language experience.

These findings contribute to advance our understanding of the type of computations infants can perform on syllables early in life. This, in turn, could shed light on the nature of the privileged status of syllables as key units of speech perception.

CRediT authorship contribution statement

Chiara Santolin: Conceptualization, Methodology, Investigation, Formal analysis, Visualization, Writing – original draft. **Konstantina Zacharaki:** Investigation, Formal analysis, Visualization, Writing – review & editing. **Juan Manuel Toro:** Conceptualization, Writing – review & editing. **Nuria Sebastian-Galles:** Conceptualization, Methodology, Supervision, Writing – review & editing, Funding acquisition.

Declaration of Competing Interest

None.

Data availability

Data are uploaded in OSF (see Materials & methods).

Acknowledgments

We give special thanks to Luca Bonatti for the helpful feedbacks on this work. We thank Valentina Ferrario for the offline coding of the familiarization phase, and Gonzalo Garcia-Castro and Ege Ekin Özer for inspiring discussions on these results. We are grateful to Pam Miller for proofreading the manuscript, and Xavier Mayoral, Silvia Blanch, Katia Pistrin of the Laboratori de Recerca en Infancia for their invaluable help during data collection. We are extremely grateful to the infants and the families who participated in the study. C.S. was supported by a Juan de la Cierva post-doctoral fellowship (IJC2019-041548-I/AEI/10.13039/501100011033). N.S.G. was supported by the project PID2021-123416NB-I00/MICIN/AEI/10.13039/501100011033/FEDER, financed by the Spanish Ministerio of Science and Innovation,the State Research Agency and the European Regional Development Fund, and by ICREA Academia 2018 award (P-5993) financed by the Catalan Institution for Research and Advanced Studies (ICREA). J.M.T was supported by the project PID2021-123973NB I00/MICIN/AEI/10.13039/501100011033/FEDER, financed by the Spanish Ministerio of Science and Innovation, the State Research Agency and the European Regional Development Fund. N.S.G and J.M.T. were also supported by the project 2021 SGR 00911 financed by the Catalan Generalitat AGAUR.

Appendix A. Supplementary data

Supplementary data to this article can be found at OSF.

References

Benavides-Varela, S., Hochmann, J. R., Macagno, F., Nespor, M., & Mehler, J. (2012). Newborn's brain activity signals the origin of word memories. *Proceedings of the National Academy of Sciences, 109*(44), 17908–17913.

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General, 117*(1), 21.

Bertoncini, J., Floccia, C., Nazzi, T., & Mehler, J. (1995). Morae and syllables: Rhythmical basis of speech representations in neonates. *Language and Speech, 38*(4), 311–329.

Bertoncini, J., & Mehler, J. (1981). Syllables as units in infant speech perception. *Infant Behavior and Development, 4*, 247–260.

Bijeljac-Babic, R., Bertoncini, J., & Mehler, J. (1993). How do 4-day-old infants categorize multisyllabic utterances? *Developmental Psychology, 29*(4), 711.

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

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.

Chen, J., & Ten Cate, C. (2015). Zebra finches can use positional and transitional cues to distinguish vocal element strings. *Behavioural Processes, 117*, 29–34.

Clements, G. (1990). The role of the sonority cycle in core syllabification. *Papers in Laboratory Phonology, 1*, 283–333.

Cristia, A., Seidl, A., & Gerken, L. (2011). Learning classes of sounds in infancy. *University of Pennsylvania Working Papers in Linguistics, 17*(1), Article 9. Retrieved from http://reposi-tory.upenn.edu/pwpl/vol17/iss11/19.

De Lacy, P. (2007). The interaction of tone, sonority, and prosodic structure. *The Cambridge Handbook of Phonology*, 281–307.

Doyle, G., & Levy, R. (2013). Combining multiple information types in Bayesian word segmentation. In *Proceedings of the 2013 conference of the north American chapter of the Association for Computational Linguistics: Human language technologies* (pp. 117–126).

Eimas, P. D. (1999). Segmental and syllabic representations in the perception of speech by young infants. *The Journal of the Acoustical Society of America, 105*(3), 1901–1911.

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.

Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition, 117*(2), 107–125.

Friederici, A. D. (2005). Neurophysiological markers of early language acquisition: From syllables to sentences. *Trends in Cognitive Sciences, 9*(10), 481–488.

Frota, S., Freitas, M. J., Vigário, M., & Martins, F. (2005). Prosodic and frequency effects on the development of syllable structure in European Portuguese. In *Paper presentation at 10th international congress for the study of child language, Berlin, Germany*.

Gasparini, L., Langus, A., Tsuji, S., & Boll-Avetisyan, N. (2021). Quantifying the role of rhythm in infants' language discrimination abilities: A meta-analysis. *Cognition, 213*, Article 104757. https://doi.org/10.31219/osf.io/rmn5x

Gennari, G., Marti, S., Palu, M., Fló, A., & Dehaene-Lambertz, G. (2021). Orthogonal neural codes for speech in the infant brain. *Proceedings of the National Academy of Sciences, 118*(31), Article e2020410118.
Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences, 105*(37), 14222–14227.
Gómez, D., Berent, I., Benavides-Varela, S., Bion, R. A., Cattarossi, L., Nespor, M., & Mehler, J. (2014). Language universals at birth. *Proceedings of the National Academy of Sciences, 111*(16), 5837–5841.
Gorman, B. K., & Gillam, R. B. (2003). Phonological awareness in Spanish: A tutorial for speech—Language pathologists. *Communication Disorders Quarterly, 25*(1), 13–22.
Halle, M., & Vergnaud, J. (1980). Three dimensional phonology. *Journal of Linguistic Research, 1*(1), 83–105.
Halle, P., & Cristia, A. (2012). Global and detailed speech representations in early language acquisition. In S. Fuchs, M. Weirich, D. Pape, & P. Perrier (Eds.), *Speech planning and dynamics* (pp. 11–38). Frankfurt am Main: Peter Lang.
Harris. (1983). *Syllable structure and stress in Spanish: A nonlinear analysis*. Cambridge, MA: MIT Press.
Hockett, C. (1955). *A manual of phonology*. Chicago: University of Chicago Press.
Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development: An International Journal of Research and Practice, 13*(4), 341–348.
Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli*, vol. 5*. *Advances in Infancy Research* (pp. 69–95). Norwood, NJ: Ablex.
Jusczyk, P., & Derrah, C. (1987). Representation of speech sounds by young infants. *Developmental Psychology, 23*(5), 648.
Jusczyk, P., Friederici, A., Wessels, J., Svenkerud, V., & Jusczyk, A. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32*(3), 402–420.
Jusczyk, P. W., Kennedy, L. J., & Jusczyk, A. M. (1995). Young infants' retention of information about syllables. *Infant Behavior and Development, 18*(1), 27–41.
Kenstowicz, M. (1986). Notes on syllable structure in three Arabic dialects. *Revue Québécoise de Linguistique, 16*(1), 101–127.
Kosie, J., Zettersten, M., Abu-Zhaya, R., Amso, D., Babineau, M., Baumgartne, H., … Casey, Z. (2023). *ManyBabies 5: A large-scale investigation of the proposed shift from familiarity preference to novelty preference in infant looking time*.
Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *The Journal of the Acoustical Society of America, 66*(6), 1668–1679.
Lass, R. (1984). *Phonology: An introduction to basic concepts*. Cambridge University Press.
Levelt, C., Schiller, N., & Levelt, W. (2000). The acquisition of syllable types. *Language Acquisition, 8*(3), 237–264.
Lew-Williams, C., Pelucchi, B., & Saffran, J. R. (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science, 14*(6), 1323–1329.
Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283*(5398), 77–80.
Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78*(2), 91–121.
Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, G. (1996). *Coping with linguistic diversity: The infant's viewpoint. Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 101–116).
Mersad, K., & Dehaene-Lambertz, G. (2016). Electrophysiological evidence of phonetic normalization across coarticulation in infants. *Developmental Science, 19*(5), 710–722.
Mersad, K., Kabdebon, C., & Dehaene-Lambertz, G. (2021). Explicit access to phonetic representations in 3-month-old infants. *Cognition, 213*, 104613.
Olson, R. (2017). Wisconsin Infant Studies Program. Retrieved from https://bitbucket.org/rholson1/wisp/src/default/.
Parker, S. (2008). Sound level protrusions as physical correlates of sonority. *Journal of Phonetics, 36*(1), 55–90.
Petkov, C., & Ten Cate, C. (2020). Structured sequence learning: Animal abilities, cognitive operations, and language evolution. *Topics in Cognitive Science, 12*(3), 828–842.
Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication, 41*(1), 245–255.
Potet, J. P. G. (1995). Tagalog monosyllabic roots. *Oceanic Linguistics*, 345–374.
Prieto Vives, P., & Bosch Baliarda, M. (2006). The development of codas in Catalan. *Catalan Journal of Linguistics, 5*, 237–272.
Prince, V., & Smolensky. (1993/2004). *Optimality Theory: Constraint Interaction in Generative Grammar*.
Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*(3), 265–292.
Rey, A., Minier, L., Malassis, R., Bogaerts, L., & Fagot, J. (2019). Regularity extraction across species: Associative learning mechanisms shared by human and non-human primates. *Topics in Cognitive Science, 11*(3), 573–586.
Ridouane, R. (2008). Syllables without vowels: Phonetic and phonological evidence from Tashlhiyt Berber. *Phonology, 25*(2), 321–359.
RStudio Team. (2019). RStudio: Integrated development environment for R RStudio, Inc. Retrieved from http://www.rstudio.com/.
Saffran, J., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F., & Cushman, F. (2008). Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition, 107*(2), 479–500.
Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology, 69*, 181–203.
Santolin, C., Garcia-Castro, G., Zettersten, M., Sebastian-Galles, N., & Saffran, J. R. (2021). Experience with research paradigms relates to infants' direction of preference. *Infancy, 26*(1), 39–46.
Santolin, C., Rosa-Salva, O., Regolin, L., & Vallortigara, G. (2016). Generalization of visual regularities in newly hatched chicks (Gallus gallus). *Animal Cognition, 19*(5), 1007–1017.
Santolin, C., Rosa-Salva, O., Vallortigara, G., & Regolin, L. (2016). Unsupervised statistical learning in newly hatched chicks. *Current Biology, 26*(23), R1218–R1220.
Santolin, C., & Saffran, J. R. (2019). Non-linguistic grammar learning by 12-month-old infants: Evidence for constraints on learning. *Journal of Cognition and Development, 20*(3), 433–441.
Schott, E., Mastroberardino, M., Fourakis, E., Lew-Williams, C., & Byers-Heinlein, K. (2021). Fine-tuning language discrimination: Bilingual and monolingual infants' detection of language switching. *Infancy, 26*(6), 1037–1056.
Sebastian-Galles, N., & Santolin, C. (2020). Bilingual acquisition: The early steps. *Annual Review of Developmental Psychology, 2*, 47–68.
Selkirk, E. (1984). On the major class features and syllable theory. In M. Aronoff, & R. Oehrle (Eds.), *Language sound structure* (pp. 107–136). Cambridge Mass: MIT Press.
Sonnweber, R., Ravignani, A., & Fitch, W. T. (2015). Non-adjacent visual dependency learning in chimpanzees. *Animal Cognition, 18*(3), 733–745.
Takahasi, M., Yamada, H., & Okanoya, K. (2010). Statistical and prosodic cues for song segmentation learning by Bengalese finches (Lonchura striata var. domestica). *Ethology, 116*(6), 481–489.
Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning & Development, 3*, 73–100.
Toro, J. M. (2016). Something old, something new: Combining mechanisms during language Acquisition. *Current Directions in Psychological Science, 25*(2), 130–134.
Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception &Ppsychophysics, 67*(5), 867–875.
Toro, J. M., Trobalón, J. B., & Sebastián-Gallés, N. (2003). The use of prosodic cues in language discrimination tasks by rats. *Animal Cognition, 6*, 131–136.
Werker, J. F. (2018). Perceptual beginnings to language acquisition. *Applied PsychoLinguistics, 39*(4), 703–728.
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.
Zacharaki, K., & Sebastian-Galles, N. (2021). The ontogeny of early language discrimination: Beyond rhythm. *Cognition, 213*, Article 104628.
Zacharaki, K., & Sebastian-Galles, N. (2022). Before perceptual narrowing: The emergence of the native sounds of language. *Infancy, 27*(5), 900–915.