ARTICLE IN PRESS

Cognition xxx (xxxx) xxx

Contents lists available at ScienceDirect

Cognition

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

This paper is a part of special issue "Special Issue in Honour of Jacques Mehler, Cognition's founding editor".

Small-range numerical representations of linguistic sounds in 9- to 10-month-old infants

Silvia Benavides-Varela [a,b,*], Natalia Reoyo-Serrano [a]

[a] *Department of Developmental Psychology and Socialisation, University of Padova, Padova, Italy*
[b] *Department of Neuroscience, University of Padova, Padova, Italy*

ARTICLE INFO

Keywords:
Small numbers
Auditory
Language
Infants
Syllables

ABSTRACT

Coordinated studies provide evidence that very young infants, like human adults and nonhuman animals, readily discriminate small and large number of visual displays on the basis of numerical information. This capacity has been considerably less studied in the auditory modality. Surprisingly, the available studies yielded mixed evidence concerning whether numerical representations of auditory items in the small number range (1 to 3) are present early in human development. Specifically, while newborns discriminate 2- from 3-syllable sequences, older infants at 6 and 9 months of age fail to differentiate 2 from 3 tones. This study tested the hypothesis that infants can represent small sets more precisely when listening to ecologically relevant linguistic sounds. The aim was to probe 9- to 10-month-olds' ($N = 74$) ability to represent sound sets in a working memory test. In experiments 1 and 2, infants successfully discriminated 2- and 3-syllable sequences on the basis of their numerosity, when continuous variables, such as individual item duration, inter-stimulus duration, pitch, intensity, and total duration, were controlled for. In experiment 3, however, infants failed to discriminate 3- from 4-syllable sequences under similar conditions. Finally, in experiment 4, infants were tested on their ability to distinguish 2 and 3 tone sequences. The results showed no evidence that infants discriminated these non-linguistic stimuli. These findings indicate that, by means of linguistic sounds, infants can access a numerical system that yields precise auditory representations in the small number range.

1. Introduction

An early and essential enterprise for infants is to organize information present in the environment along the dimensions relevant for the specific task they are facing. Some authors argue that the organization of this rich input might be initially guided by early functioning biases, and the existence of perceptual analyzers that help the young learner identify the entities (e.g., objects, events, sounds, etc.) in core conceptual domains (Carey, 2011). Two prime examples of core cognition that are relevant for the present work concern: 1) the innate sensitivity to numerosity, namely an abstract property defined by the number of discriminable units within sets, and 2) the infants' preparedness to process and acquire language. The current study combines these two domains by investigating the infants' ability to represent small number sets in the auditory domain, focusing on discrimination of speech units (i.e., syllables).

1.1. Numerical cognitive systems

Young infants share with human adults, and non-human animals a subset of numerical skills that are considered the evolutionary foundation of more complex numerical reasoning [evidence in infants: de Hevia, 2016; de Hevia et al., 2017; Feigenson, Carey, & Hauser, 2002; Hyde & Spelke, 2011; in adults: Ansari, Lyons, van Eimeren, & Xu, 2007; Benavides-Varela et al., 2018; Castaldi, Piazza, Dehaene, Vignaud, & Eger, 2019; Hyde & Spelke, 2009, Semenza & Benavides-Varela, 2018; in animals: Agrillo & Bisazza, 2018; Agrillo, Dadda, & Bisazza, 2007; Rugani, Cavazzana, Vallortigara, & Regolin, 2013; Rugani & De Hevia, 2017; Rugani, Regolin, & Vallortigara, 2008; Rugani, Vallortigara, & Regolin, 2013; Vallortigara, Chiandetti, Rugani, Sovrano, & Regolin, 2010], and that are not dependent upon language (Dehaene, 2001; Gelman & Butterworth, 2005). They possess two cognitive systems for encoding numerical information (Feigenson, Dehaene, & Spelke, 2004; Xu, 2003): a parallel individuation or object file system (OFS), specifically dedicated to precisely tracking small numbers (generally three or

* Corresponding author at: Università degli Studi di Padova, Via Venezia 8, 35131 Padova, Italy.
E-mail address: silvia.benavidesvarela@unipd.it (S. Benavides-Varela).

https://doi.org/10.1016/j.cognition.2021.104637
Received 29 August 2020; Received in revised form 11 February 2021; Accepted 13 February 2021

Please cite this article as: Silvia Benavides-Varela, Natalia Reoyo-Serrano, *Cognition*, https://doi.org/10.1016/j.cognition.2021.104637

fewer) and an analog or approximate number system (ANS) that operates in relation to larger numbers, but that under certain circumstances (e.g., Cordes & Brannon, 2009a; Coubart, Streri, de Hevia, & Izard, 2015; Rugani, Vallortigara, & Regolin, 2014; VanMarle & Wynn, 2009) might be responsible for representing all set sizes in an approximate manner (for reviews, see Hyde, 2011; Mou & vanMarle, 2014; Posid & Cordes, 2015).

These two systems show different developmental trajectories and functioning features. The ANS is functional from birth on (Izard, Sann, Spelke, & Streri, 2009) and obeys Weber's Law; that is, the discriminability of two values is dependent upon their ratio, not the absolute difference between two values (Dehaene, 1997). Thus, it is easier to discriminate 5 from 10 items (a 1:2 ratio) than 10 from 15 items (a 2:3 ratio), despite the fact that the absolute difference between the values (i.e., 5) is the same in both cases. Moreover, the ANS gradually increases in precision throughout development. Newborns need a 1:3 ratio to discriminate large numerosities. For example, they discriminate 4 vs. 12, but not 4 vs. 8 (Izard et al., 2009), but older infants progressively discriminate finer ratios. By 6 months of age they discriminate numerosities in a 1:2 ratio, and a 2:3 ratio at 9 to 10 months of age (e.g., Brannon, Suanda, & Libertus, 2007; Libertus & Brannon, 2010; Lipton & Spelke, 2003, 2004; McCrink & Wynn, 2007; Wood & Spelke, 2005; Xu & Arriaga, 2007; Xu & Spelke, 2000). Children 3- to 4-years old distinguish a 3:4 ratio (Halberda & Feigenson, 2008), and human adults are able to discriminate numbers with at least a 7:8 ratio (Barth, Kanwisher, & Spelke, 2003) and possibly up to a 9:10 or 10:11 ratios (Halberda & Feigenson, 2008).

The OFS also operates from birth, and its limits are not ruled by ratio but by set size. Beyond this limit, infants fail to even encode that there is more than 1 object in the set (Feigenson et al., 2004; Feigenson & Carey, 2003, 2005; Feigenson, Carey, & Hauser, 2002). Newborns seem able to track and represent a maximum of 2 individual objects in parallel (Coubart, Izard, Spelke, Marie, & Streri, 2014), 10- and 12-month-old infants can represent up to 3 objects (Feigenson, Carey, & Hauser, 2002; Feigenson, Carey, & Spelke, 2002) and human adults about 4 discriminable objects (Trick & Pylyshyn, 1994). The OFS was originally proposed to account for how adults simultaneously attend to small number of visual objects that are briefly stored in memory (Kahneman, Treisman, & Gibbs, 1992; Trick & Pylyshyn, 1994). It works by tracking spatiotemporal information, property/kind changes, and object features such as color, size, and shape, to identify each new object that is introduced into a scene. Subsequent accounts have argued for the existence of "objecthood" in other sensory modalities as well, including auditory perception. According to this view, auditory events are perceived as discrete sound objects defined in terms of their spectrotemporal dimensions (e.g. Griffiths & Warren, 2004; Kubovy & Van Valkenburg, 2001) as well as their location in space (e.g. Warren & Griffiths, 2003).

The number of objects that can be simultaneously tracked in this temporary store is implicitly represented and limited, giving rise to the finite, yet often more precise, representation of small collections.

The precision of infants' numerosity representations in the small numerosity range, which is the focus of the present work, has been previously investigated in the visual modality, in cross-modal studies, as well as with auditory stimuli (see Cantrell & Smith, 2013, for a review). Although it is generally stated that this representation is independent of the format of the stimuli (e.g., Feigenson, 2007), researchers found mixed results particularly inherent to the auditory modality.

1.2. Small numerosity discrimination in the visual modality and across modalities

Several studies using visual stimuli attested that infant discriminations of exclusively small sets can be more precise than predicted by the Weber's law. For example, infants at birth (e.g., Antell & Keating, 1983), at 4 months of age (Starkey & Cooper, 1980), as well as 5-month-old infants (Van Loosbroek & Smitsman, 1990; Wynn, 1992), 7-month-olds (Cordes & Brannon, 2009b), 10 to 12-month-olds (Feigenson, Carey, & Hauser, 2002; Feigenson, Carey, & Spelke, 2002; Strauss & Curtis, 1981), and 12 to 14-month-olds (Feigenson & Carey, 2005) are all able to discriminate a 2:3 ratio in small sets (2 vs. 3 objects), but before the age of 9 months are incapable of discriminating the same ratio in large sets (e.g., 6 vs. 9).

Infants' capacity to discriminate visual objects in the small range has been documented by multiple studies using diverse paradigms. For instance, when visually habituated or familiarized to displays of a given number of items (e.g., two), infants will subsequently display statistically different looking behaviors towards displays containing a new number (e.g., three) than at new displays containing the habituated number. Additionally, 5-month-old infants can perform simple numerical computations, anticipating the numerical outcomes of physical operations such as the addition or removal of an object from a small array. For instance, if infants are shown a 1 + 1 addition, they look longer (i.e., appear surprised) at the result of 3 items than of 2 items (Wynn, 1992). Moreover, when given the opportunity to choose between containers with 2 or 3 food items (crackers), 10 and 12-month-old infants discriminate between these items and spontaneously choose the bucket containing 3 crackers, namely the set with more items (Feigenson, Carey, & Hauser, 2002; Feigenson, Carey, & Spelke, 2002). Furthermore, at 12–14 months of age, in a manual search paradigm, infants notice whether the number of objects retrieved from a box (e.g., 2) is different from the number of objects originally placed into the box (e.g., 3). This is evidenced by their longer manual searching behaviors in this condition, compared to when they had originally seen only two objects placed into the box (Feigenson & Carey, 2005).

Studies pointed out that sometimes, when sets are small, infants tend to respond to continuous variables that covary with number (Clearfield & Mix, 1999, 2001; Feigenson, Carey, & Spelke, 2002; Mix, Huttenlocher, & Levine, 2002; Mix, Levine, & Huttenlocher, 1997). This does not undermine the claim that discrimination in infants is number-based, but rather suggests that stringent stimulus controls are necessary before applying strong interpretations. Indeed, there is also evidence suggesting that changes in number are salient to the infant when non-numerical variables are strictly controlled for both small and large sets alike (Cordes & Brannon, 2008, 2009b; Wynn, Bloom, & Chiang, 2002).

Young infants also discriminate small number of sets in tasks that include multimodal input (e.g., Coubart et al., 2014; Jordan, Suanda, & Brannon, 2008). Pioneer studies of this kind showed that 6- to 8-month-old infants differentially looked at a display with a number of visual items congruent with the number of sounds they heard, compared to displays that showed a number of objects incongruent with the number of sounds played (e.g., 2 visual items and 3 sounds or vice versa; Starkey, Spelke, & Gelman, 1983, 1990; Kobayashi, Hiraki, & Hasegawa, 2005). Similarly, Jordan and Brannon (2006) found that 7-month-old infants preferentially attended to visual displays with the same number of faces as the number of adult humans they heard speaking, when a small number of voices and faces (two or three) were presented. Moreover, Féron, Gentaz, and Streri (2006) showed that 5-month-old infants successfully discriminate 2 vs. 3 objects using a transfer paradigm from the tactile to the visual modality, thus suggesting an amodal representation of numbers. Together, these studies suggest that young infants also succeed in discriminating exclusively small sets (e.g., two from three items) when redundant information is provided.

1.3. Small numerosity discrimination in the auditory modality

A few studies have addressed the question concerning whether numerical representations of small sets of solely auditory stimuli are present early in human development. These studies have yielded mixed evidence.

In one of the first studies of its kind, Bijeljac-Babic, Bertoncini, and Mehler (1993) presented 4-day-old infants with naturally produced multisyllabic utterances while controlling for total sound duration. The authors tested whether newborns were able to detect a change in the number of syllables using the high-amplitude sucking procedure. The procedure builds on the infants' interest for certain stimuli, which is translated into an elevated sucking rate. The results showed that infants increased their sucking rate from habituation to test when they heard novel natural utterances containing a different number of syllables (i.e., 2 or 3 syllables). By contrast, a control group that heard novel words containing the same number of syllables in the habituation and test phases, did not show significant changes in sucking behavior. The study demonstrated that newborns were able to detect and distinguish sounds in the small number range.

However, subsequent research exploring numerical discrimination in the auditory modality found, unlike Bijeljac-Babic and colleagues, that older infants were unable to make the 2- versus 3-item distinction. First, Lipton and Spelke (2004) used a head-turn preference procedure to test infants' sensitivity to large and small numerosities in auditory stimuli (bells, whistles, chirps, buzzes, drums, and horns). Their study provided evidence that 9-month-old infants discriminate large number of sounds differing by a 2:3 ratio (i.e., 8 vs. 12 and 4 vs. 6 sounds), but not small number of sounds differing by the same ratio (2 vs. 3 sounds). The infants' failure to discriminate 2 and 3 sounds was interpreted as evidence that the OFS, operating in the small numerosity range, was not functional for comparing small numerosities in the auditory modality.

A subsequent study testing the hypothesis that infants can use analog magnitudes to represent small values in the auditory domain (VanMarle & Wynn, 2009) reported that 7-month-old infants discriminated two from four tones (a 1:2 ratio) but failed to discriminate a more difficult (2:3) ratio, namely two from three tones. The authors argued that infants' discrimination of small numbers of auditory events could be achieved by accessing the ANS. Under this view, it is the ratio difficulty -not the incapacity to use the OFS- that determined the infants' incapacity to distinguish 2 from 3 sounds. Although the studies of Lipton and Spelke (2004) and VanMarle and Wynn (2009) provided a different theoretical explanation to account for the infants' behavior, both converged on the fact that infants' fail to discriminate a number of auditory items in the small range, that is otherwise distinguishable by infants of similar ages in the visual and audio-visual modalities.

1.4. The present study

A conundrum raised by the overall pattern of findings described above is why newborns succeed in the 2 versus 3 sound distinction whereas older infants, at 7 months and 9 months of age, fail to do so.

One possible explanation is that sensitivity to numerical distinctions in the auditory modality declines during ontogeny. Lowering in sensitivity, also known as *perceptual narrowing*, has been described in infancy for various developmental processes (e.g., Kelly et al., 2007; Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009; Werker & Tees, 1984). In the language domain, for example, the phonemes of different languages can be discriminated by young infants until about the first half of the first year of life (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). However, as the infant grows and acquires experience with native perceptual inputs, the sensitivity to phonemes of non-native languages (to which they are never exposed to) declines (Werker & Tees, 1984). Under a similar developmental scenario, older infants, unlike newborns, might fail to discriminate 2 from 3 sounds perhaps because they experience less numerical discriminations tasks in the auditory system. Alternatively, infants may be capable of precisely representing and distinguishing small sets of sounds under some circumstances, for instance, if presented with ecologically relevant and meaningful sounds, like linguistic ones.

In this article, we ask whether accurately representing the number of sounds (i.e., whether a sequence contains specifically 2 or 3 syllables) is part of infants' knowledge at 9 to 10 months of age. To assess this empirical question, we employed the two-alternative looking paradigm. This procedure was initially developed for studies of auditory and visual categorization (McMurray & Aslin, 2004), and was subsequently employed to test various cognitive abilities in infants, including bilingual speech perception (Albareda-Castellot, Pons, & Sebastián-Gallés, 2011), same/different abstract representations (Addyman & Mareschal, 2010), rule learning (Hochmann, Benavides-Varela, Fló, Nespor, & Mehler, 2018; Hochmann, Benavides-Varela, Nespor, & Mehler, 2011; Kovács & Mehler, 2009a, 2009b), and the encoding of word sounds (Benavides-Varela & Mehler, 2015). Specifically, the present study adapted the original procedure of Kovács and Mehler (2009a, 2009b). In the familiarization phase, 9- to 10-month-olds were simultaneously presented with a wide set of two- and three-syllable sets. Following each sequence, an attractive toy appeared on the left or the right side of the screen (Fig. 2). The speech sequence's numerosity predicted the location of the toy. For example, if the sequence contained 2 syllables the toy appeared on the right side of the screen, whereas if the sequences contained 3 syllables, it appeared on the left side. Then, infants were tested on their ability to generalize with novel items. In these test trials, the toy did not appear after the speech sequence.

There are three crucial aspects of this kind of experimental design that are unique for studies on auditory numerical perception in infants. Most of these aspects imply a higher burden on working memory and attentional resources than those taxed by traditional habituation/familiarizion paradigms. First, whilst previous studies varied the stimuli during the habituation/familiarization but kept the stimuli constant during the test phase, in the present study the stimuli vary from trial to trial during both the familiarization and test. This impedes infants to use a single speech sequence for discrimination in either phase. Varying the tokens across trials also avoids infants detecting co-occurrences of puppets and sequences and forming specific links between them. Such rapid object–sound associations have been reported in infants as young as 3 months of age (Friedrich & Friederici, 2017), and could have interfered with the aims of the study. Second, the present study tested the infants' ability to simultaneously track what is common and what differs between two sets of stimuli in parallel, while previous studies familiarized infants with only one set and then tested whether infants reacted to/discriminated a single change in numerosity in the test phase. Thus, besides providing a measure of discrimination, intermixing two sets consents to explore the infants' ability to grasp the task-relevant feature (i.e. number of sounds) for organizing multiple stimuli online. Arguably, this offers a better approximation of whether and how infants sort out information in real and more complex learning situations. Third, during the test phase, like in previous studies, infants are required to generalize an abstract numerical property to new instances; here in addition, they might reliably look to the correct side of the screen only if they can remember that a given abstract property determined the location of a toy in the previous phase. This memory task is not involved in traditional habituation/familiarization tasks and should provide an indication of the infants' ability to store abstract representations and to subsequently perform simple computations over them (e.g. build correct predictions of future events).

In this paradigm, monolingual-learning infants have trouble learning two regularities in parallel (e.g., Benavides-Varela & Mehler, 2015; Hochmann et al., 2011; Hochmann et al., 2018; Kovács & Mehler, 2009a, 2009b), possibly due to limitations in executive functions and working memory (Kovács & Mehler, 2009a, 2009b). Thus, a key signature of discrimination is the generalization of one the two simultaneously presented syllable sets: the one that infants find easier to process. Because infants in our sample were also monolingual, we expected them to generalize the smaller-number sequence.

Two experiments used this testing procedure to study the critical 2 vs. 3 discrimination while controlling for continuous physical properties of sounds, namely item duration, inter-stimulus duration, pitch, and intensity. Moreover, while Experiment 1 also controlled for tempo, Experiment 2 controlled for the total duration of the stimuli in the two sets. Given that infants appeared to discriminate these numerosities, the limits of the infants' auditory numerical representations in the small range were tested in Experiment 3. In particular, the discrimination between 3 and 4 sets of syllables was tested there. Finally, in Experiment 4, the infants' ability to discriminate 2 vs. 3 tone sequences was assessed.

2. Experiment 1: discrimination of 2 and 3 syllables

If speech enjoys a privileged status *vis-á-vis* auditory numerical representations, infants should be able to accurately represent and distinguish small number of syllables in a condition in which similar-age infants fail with non-speech sounds (i.e., 2 vs. 3 sounds).

2.1. Participants

Nineteen infants, monolingual learners of Italian, participated in the study (9 males; aged from 9 months 21 days to 10 months 18 days; M_{age} = 10 months and 5 days). All participants were full term, with no birth complications, and no hearing or visual problems reported. Seven additional infants were excluded from the analysis because of crying or fussiness ($N = 2$), due to side bias ($N = 1$), or because of experimental error ($N = 4$).

Participants of the first three Experiments presented in this study were recruited around the city of Trieste in Italy. The families were contacted by means of a letter sent to their residence. The testing took place between June and August 2010. Information on participants' ethnicity, parental education, income, and occupation was not recorded at the time of testing, but prior work with this population suggests that participants are primarily Caucasian and predominantly come from middle to high socioeconomic status homes. Infants' parents signed the informed consent before the experiment and after they had understood the procedure and all their questions had been answered. The Ethics Committee of the Scuola Internazionale Superiore di Studi Avanzati in Trieste, where the experiments were conducted, approved the study.

2.2. Stimuli

Linguistic stimuli were consonant-vowel (CV) syllables repeated either two (i.e., CV_CV) or three times (i.e., CV_CV_CV). The syllables were synthesized with the female voice of the MBROLA Italian database IT4 (Dutoit, Pagel, Pierret, Bataille, & Van der Vrecken, 1996). The duration of each syllable was 200 ms and a monotonous pitch of 240 Hz was used. The intensity of the stimuli was 70 dB. There were pauses of 250 ms between syllables. By keeping pitch, duration, and intensity constant, we precluded infants from grouping or chunking the syllable sequences on the basis of acoustic features (see Bion, Benavides-Varela, & Nespor, 2011 showing these abilities by the age of 7 months). Moreover, by maintaining constant the rate of syllable presentation we prevented participants from using this continuous temporal property for discrimination of the two sets (although to our knowledge no study has directly tested the infants' capacity to discriminate different speech rates). Twenty different syllables [sa, fi, be, to, ku, ka, le, ni, zo, tu, ba, ve, ri, mo, fu, la, se, pi, gu, do] were used. For each participant, twelve different syllables were randomly chosen for the familiarization trials (repeated either 2 or 3 times) and the remaining eight were exclusively used in the test phase, for generalization. The syllables chosen for the familiarization and test varied randomly across participants. A summary of the properties of the stimuli used in the experiments is presented in Fig. 1. Details can be found in the Supplementary materials-1.

The visual stimuli were three pictures of colorful puppets. Each of the puppets appeared inside one of two white squares displayed on the left and the right side of the screen. The squares had a side-length of 8 cm, positioned at a distance of 13.5 cm. The puppets loomed from 4 cm to 7 cm inside the squares for 2 s. The colorful puppets were used as visual reinforcement. They were accompanied by a tinkling bell of 300 ms presented with a delay of 800 ms with respect to the onset of the visual stimulus.

2.3. Procedure

The experiment consisted of a familiarization phase of 20 trials and a test phase of 8 trials (Fig. 2). Trials in the familiarization phase started with a display of a central animated visual attractor and two white squares, one on the left and one on the right side of the screen. When the infant fixated the central attractor, a syllable set was played. After the offset of the stimulus, the central attractor disappeared, leaving only the two white squares visible for 1 s. Subsequently, a looming puppet appeared on one of the two white squares for 2 s. The side of the presentation of the puppet was contingent on the sequence: the 2-syllables sequence predicted its appearance in one of the squares (e.g., left), while the 3-syllables sequence predicted the puppet' s appearance in the other square (e.g. right). The pairing of the sequences with puppets' locations was counterbalanced across participants. Two-syllable and three-syllable sequences were presented in an interleaved pseudo-random order, ensuring that each of the sequences was presented no more than three times in a row. The sound-puppet pairing was random. Different syllables were presented across trials to avoid infants associating the appearance of the puppet with a specific token. Moreover the

Experiment	Stimuli properties							Graphical Example
	***Tokens** (see Supplementary Materials for details)*	***Number of elements***	***Token pitch***	***Token duration***	***Token intensity***	***ISI duration***	***Sequence duration***	
1	20 different syllables	2-Syllables	240Hz	200ms	70dB	250ms	650ms	
		3-Syllables	240Hz	200ms	70dB	250ms	1100ms	
2		2-Syllables	240Hz	269ms	70dB	337ms	875ms	
		3-Syllables	240Hz	159ms	70dB	199ms	875ms	
3		3-Syllables	240Hz	200ms	70dB	250ms	1100ms	
		4-Syllables	240Hz	200ms	70dB	250ms	1500ms	
4	20 different pure sine waves	2-Tones	Range: 250-1300 Hz	200ms	70dB	250ms	650ms	
		3-Tones	Range: 250-1300 Hz	200ms	70dB	250ms	1100ms	

Fig. 1. Properties of the stimuli used in the Experiments.

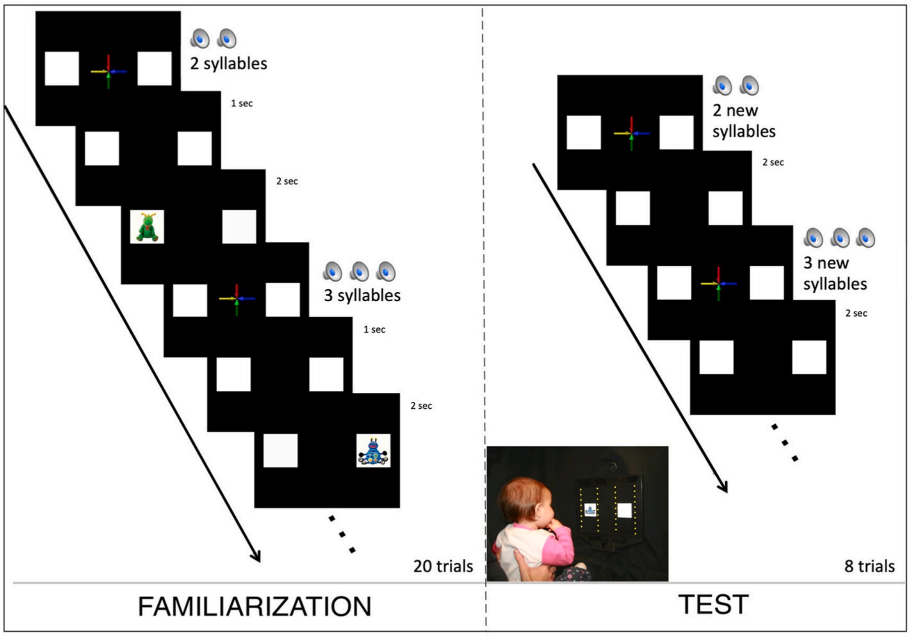

Fig. 2. Schematic diagram of the procedure and the timeline of the trials. The small box in the bottom shows the set up. The yellow dotted lines superimposed in the screen exemplify the horizontal division (left, middle, right) used for the coding. The lines are for illustration purposes and were not displayed to infants.

same syllables were used for the 2 and 3 syllable sequences across trials, so that infants could not use the identity of the syllables to form the categories.

The test trials were similar to the familiarization trials, except that after each syllable set no puppet was displayed in either side of the screen. Only the two white squares were visible. Two seconds after the sequence offset, the next trial started. Each of the sets was presented four times in a pseudo-random order. The first four trials consisted in two trials per condition.

2.4. Data acquisition

Infants' gaze was recorded with an eye-tracker (TOBII 1750). The eye tracker was integrated into a 17-in. TFT screen, where the stimuli were presented via an IMAC 10.1 running PsyScope X software (http://psy.ck.sissa.it/). A loudspeaker was placed behind the screen for the presentation of the acoustic stimuli. Infants were seated on their parent's lap at about 50 cm distance from the monitor. A hidden video-camera was used to observe the infant's behavior (see small box in Fig. 2).

2.5. Data analysis

2.5.1. Coding and primary dependent measures

To assess whether infants discriminated and generalized the numerosity property, we measured their looking behavior after hearing a new item. The primary measure consisted in coding the direction of the first look in each trial, and separately for the familiarization (see Supplementary materials -2) and the test phases. The direction of the first look was determined by dividing the screen into three parts of equal size: left, middle and right (see small box in Fig. 2). If infants fixated either the left or the right side of the screen, their gaze was coded as correct or incorrect according to the auditory sequence presented.

A normalized difference score was then used as dependent variable. The score takes into consideration the first look behavior of the 8 trials in the test phase. It was obtained by subtracting the number of trials in which the infant fixated first the incorrect side from the number of trials in which they fixated first the correct side, divided by the total number of trials in which the infants fixated either side.

$$\text{Normalized difference score} = \frac{(\#\text{correct looks-}\#\text{incorrect looks})}{(\#\text{correct looks} + \#\text{incorrect looks})}$$

Normalized difference scores range between −1 and 1. Positive difference scores indicate that infants identified the distinctive numerical features and singled out the side of the screen assigned to each sequence. Negative scores indicate that they searched for the puppet in the incorrect side of the screen. Scores around zero indicated that infants showed no consistent pattern of responses.

Statistically, infants' performance was first compared against chance using separate *t*-tests for the 2 and for the 3-syllable sequences. Then, a paired t-test was used to compare the infants' performance in the two types of sequences. Unless otherwise specified, one-tailed tests were computed reflecting the directional hypotheses stated in the introduction. Namely, that infants would succeed or perform above chance (not different from chance), and that their performance would be better for the smaller-number sequence (2-syllables sequence).

Spurious looking times were excluded from the analyses, using the same criterion adopted by Kovács' original study (Kovács, 2008; Kovács & Mehler, 2009a, 2009b). Thus, looking times $\leq$ 80 ms were not considered a fixation. Average fixation times were calculated and reported for each study.

The eight test trials might induce extinction effects because the puppet did not appear after the speech sequence. Thus, additional analyses considering only the results of the first four test trials were also computed (See Supplementary materials -3).

2.5.2. Additional measures

We calculated three additional indices of infants' looking behavior in the test phase. First, we consider the "correctness" or the proportion of time the participant look to the correct window as compared to the total time the infant look at the correct and incorrect sides in each trial. A second measure was the longest look, namely the side of the screen in which infants fixated longer within the 2 s after hearing a new item and before the start of the next trial. A trial was coded as correct if the infant looked longer to the correct side. The trial was coded as incorrect if the infant looked longer to the incorrect side of the screen. Then, difference scores were computed in the same way as it was done for the first look behavior. Finally, a measure of latency was extracted, namely the time elapsed from the end of the presentation of the auditory stimuli until the first fixation in either the left or the right side of the screen. Data from trials with incorrect responses were discarded from the latter analysis.

2.6. Results

Infants fixated either side of the screen 72.9% of the total number of test trials on average. The mean duration of the first fixation was 597 ± 349 ms. In the two-syllable sequences infants' mean difference score was 0.40, which was significantly greater than chance [$t(18) = 2.60$; $p = .01$; Cohen's d = 0.59]. For the three-syllable sequences infants' mean difference score was −0.09 and did not differ significantly from chance, [$t(18) = -0.53$; $p > .05$; Cohen's d = 0.12]. A paired t-test showed that infants obtained significantly higher difference scores for the 2-syllables than for the 3-syllable sequences [$t(18) = 1.73$; $p = .05$; Cohen's d = 0.40]. The results are presented in Fig. 3A.

The analyses over the proportion of correct looking time revealed a comparable finding to that observed in the first look analysis. The proportion of correct looking was significant for the 2-syllable sequences [Mean = 0.65; $t(18) = 2.76$; $p = .007$; Cohen's d = 0.63], but did not reach significance for the 3-syllable trials [Mean = 0.50; $t(18) = 0.06$; $p = .48$; Cohen's d = 0.01]. The results are depicted in Fig. 3B.

The analysis over the longest look also yielded similar results to those obtained with the previous measures. In the two-syllable sequences infants' mean difference score (0.34) was significantly greater than chance [$t(18) = 2.64$; $p = .008$; Cohen's d = 0.61] whereas in the three-syllable sequences infants' mean difference score (0.11) did not differ significantly from chance, [$t(18) = 0.74$; $p > .05$; Cohen's d = 0.17]. The results are presented in Fig. 3C.

Finally, the results considering response latencies revealed no significant modulation across conditions [$t(12) = -0.26$; $p > .05$; Cohen's d = 0.07, two-tails]. The mean response latencies were 960 ± 271 ms and 1004 ± 468 ms for the 2 and 3 syllable sequences respectively. Six infants were not included in this statistical analysis because they did not provide correct responses in at least one of the conditions.

2.7. Discussion

Experiment 1 provides evidence that 9- to 10-month-old infants discriminate two from three syllable sequences when the potentially confounding continuous variables of syllable duration, inter-stimulus duration, pitch, tempo, and intensity are controlled. Moreover, the results showed that the infants were able to associate a two-syllables sequence to the corresponding side of the screen, suggesting that this property is easier to learn with respect to the three-syllables sequence, presumably because it imposes a lower load on working memory capacities.

Apparently, infants were able to differentiate between the two sequences based on their ability to extract a common perceptual pattern from each set of stimuli. One possible explanation is that they classify strings according to the number of syllables. However, one should abstain from accepting this interpretation too hastily. As a consequence of equalizing tempo (delay between the onsets of two consecutive events) the 2-syllable stimuli were 0.45 s shorter that the 3-syllable stimuli. Studies on auditory perceptual abilities indicate that duration discrimination takes several years to complete maturation (Elfenbein, Small, & Davis, 1993; Jensen & Neff, 1993), and that durational cues for segmentation and grouping of auditory stimuli require substantial language experience (Bion et al., 2011; de la Mora, Nespor, & Toro, 2013). This questions the fact that infants used duration as the primary cue to discrimination in the present study. Nevertheless, other studies found that infants are sensitive to small differences in temporal intervals. For example, 10-month-old infants can detect deviations of inter-tone intervals of 1.5 s and 0.5 s, as evidenced by electrophysiological measures (Brannon, Roussel, Meck, & Woldorff, 2004). Moreover, 6-month-olds can discriminate 5 s and 1 s audio-visual events in behavioral studies (VanMarle & Wynn, 2006), and 10 month-olds distinguish events lasting 1 s vs. 1.5 s (Brannon et al., 2007). It is thus an open question whether differences in duration may have constituted one property that contributed at determining the 9- to 10-month-old infants' performance in the present study. Experiment 2 tested this possibility directly.

3. Experiment 2: infants' ability to differentiate between sets of 2 and 3 syllables of the same overall duration

This experiment aimed at establishing whether infants' ability to discriminate sequences on the basis of the number of syllables persists when durational differences are removed. To achieve this, the experiment used stimuli similar to those in Experiment 1, but the durations of the 2- and 3-syllable sequences were equated. In particular, we submitted the two sets to extension and compression procedures in such a way that duration could no longer be a consistent cue on which infants might discriminate the two types of stimuli. If the infants base their discrimination mainly on non-numerical attributes, their performance should be affected when the durational differences between the two sets are eliminated. On the other hand, if the infants' responses are based on their numerical competence, then they should replicate the same pattern of responses as in Experiment 1.

3.1. Participants

Participants were twenty-two infants (8 females; aged from 9 months 17 days to 10 months 19 days; M_{age} = 10 months 2 days), monolingual learners of Italian. All participants were full term, with no birth

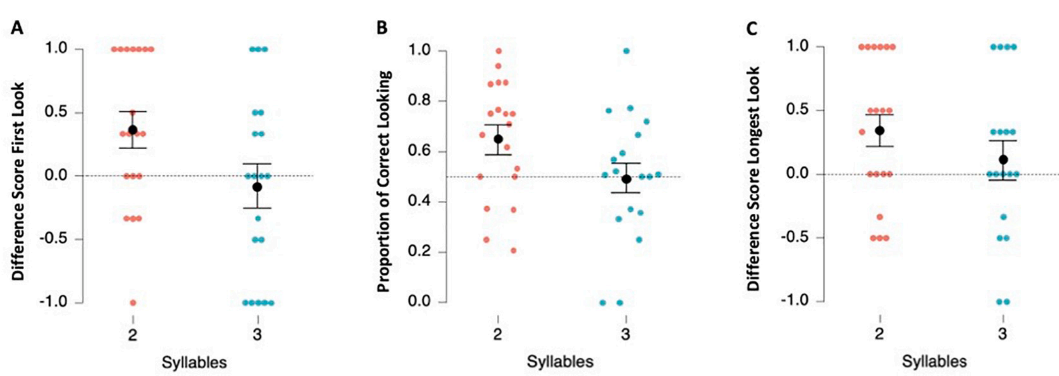

Fig. 3. Main findings of Experiment 1 - Infants' ability to differentiate between sets of 2 and 3 syllables. A. Normalized difference scores computed over first look measures. B. Proportion of looking time towards the correct side of the screen. C. Normalized difference scores computed over the longest look. Colored dots represent individual participants' scores in each condition, the black dot indicates the group mean, and bars depict standard errors of the mean. The dotted line in the middle depicts chance level.

complications, and no hearing or visual problems reported. Three additional infants were excluded from the analysis because of crying or fussiness.

3.2. Stimuli

The stimuli were the same used in the previous experiment, except that the sequences were edited in Praat (Boersma, 2001) to have the same duration. The algorithm shortened the 3-syllable sequences and lengthened the 2-syllable sequences so that all the sequences lasted 875 ms. The visual stimuli were identical to those used in the previous experiment.

3.3. Procedure and data analysis

The procedure and analyses were identical to those used in the previous experiment.

3.4. Results

Infants fixated the center of the screen 82.6% of the total number of test trials on average. The mean duration of the first fixation was 650 ± 265 ms. Two infants did not provide data in the 2-syllables tests, and two different infants did not provide data in the 3-syllable sequences. Thus, 20 infants were included in each of the analyses. In the sequences with 2 syllables, infants' mean difference score was 0.36, which was significantly greater than 0 [$t(19) = 2.36$; $p = .02$; Cohen's $d = 0.52$]. For the 3-syllable sequences infants' mean difference score was −0.18, which did not differ from chance [$t(19) = -1.43$; $p > .05$; Cohen's $d = 0.32$]. Moreover, paired *t*-test showed that infants obtained significantly higher difference scores for the 2-syllables than for the 3-syllable sequences [$t(17) = 3.31$; $p = .002$; Cohen's $d = 0.78$]. The results are presented in Fig. 4A.

The results considering the proportion of correct looking times are depicted in Fig. 4B. The analyses also indicated that infants performed significantly above chance for the 2-syllable sequences [Mean = 0.63; $t(19) = 2.75$; $p = .007$; Cohen's $d = 0.62$], but did not reach significance for the 3-syllable sequences, [Mean = 0.45; $t(19) = 0.80$; $p = .22$; Cohen's $d = 0.18$].

The analysis over the longest look showed results consistent with those obtained in with the previous measures. In the 2-syllable sequences infants' mean difference score (0.39) was significantly greater than chance [$t(19) = 3.41$; $p = .002$; Cohen's $d = 0.76$] whereas in the 3-syllable sequences infants' mean difference score (−0.16) did not differ significantly from chance, [$t(19) = -1.09$; $p > .05$; Cohen's $d = 0.24$]. The results are presented in Fig. 4C.

Finally, considering response latencies, the results showed no effects across conditions. The mean response latencies were 750 ± 300 ms and 982 ± 451 ms for the 2 and 3 syllable sequences respectively. There were no significant differences between the two [$t(13) = -1.48$; $p > .05$; Cohen's $d = 0.41$, two-tails]. Four infants did not provide correct responses in at least one of the conditions.

3.5. Discussion

The results of Experiment 2 confirmed those of Experiment 1. They corroborate that infants discriminate 2- from 3-syllable sequences and find the 2-syllable regularity more prominent or easier to learn with respect to the 3-syllable one. Furthermore, the findings of Experiment 2 suggest that the infants' responses do not merely rely on the duration of the stimuli. On the contrary, the results indicate that infants continue to exhibit a reliable discrimination, very similar to that observed in Experiment 1. Thus, their performance appears to be unaffected when durational properties of the stimuli were eliminated.

Altogether, the results of Experiments 1 and 2 suggest that 9- to 10-month-old infants can make exact discriminations amongst two small numerosities, when instantiated over syllables. Taken together, previous and present results suggest the discrimination of speech units in the small number range is present at birth (Bijeljac-Babic et al., 1993) and remains invariant for the first months of life. Experiment 3 tested for this capacity further, identifying the syllable sequences for which 9- to 10-month-olds can no longer make a distinction.

4. Experiment 3: infants' ability to differentiate between 3 and 4 syllables

The goal of Experiment 3 was to identify the limits of the infants' ability to discriminate auditory sequences of linguistic stimuli. Thus, we tested 9- to 10-month-olds in a more difficult task in which they were presented with 3- and 4-syllable sets.

4.1. Participants

Participants were seventeen infants (10 males; aged from 9 months 20 days to 10 months 25 days; $M_{age} = 10$ months 2 days), monolingual learners of Italian. All participants were full term, with no birth complications, and no hearing or visual problems reported. Additional 3 infants were excluded from the analysis because of crying or fussiness ($N = 1$), or because of experimental error ($N = 2$).

4.2. Stimuli

The stimuli were constructed as in Experiment 1, except that the sequences had either 3 or 4 syllables.

4.3. Procedure and data analysis

The analyses and procedure were identical to those performed in the previous experiments.

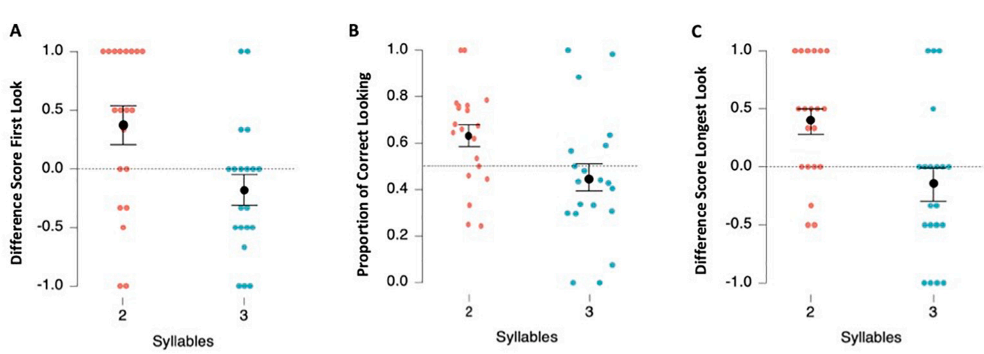

Fig. 4. Main findings of Experiment 2 – testing infants' ability to differentiate between sets of 2 and 3 syllables of the same duration. A. Normalized difference scores computed over first look measures. B. Proportion of looking time towards the correct side of the screen. C. Normalized difference scores computed over the longest look. Colored dots represent individual participants' scores in each condition, the black dot indicates the group mean, and bars depict standard errors of the mean. The dotted line in the middle depicts chance level.

4.4. Results

Infants fixated the center of the screen 76.1% of the total number of trials on average. The mean duration of their first fixation was 619 ± 190 ms. In the sequences with 3 repetitions of the syllable, infants' mean difference score was 0.15, which was not significantly greater than chance, [t(16) = 0.77; $p > .05$; Cohen d' = 0.19]. In the 4-syllables tests, mean difference score was −0.19. This value did not differ from chance [t(15) = −1.02; $p > .05$; Cohen's d = 0.25]. One infant did not provide data in this condition. Moreover, paired *t*-test comparing the infants' performance in the 3 and 4-syllables trials showed that the results did not differ [t(15) = 0.93; $p > .05$; Cohen's d = 0.23]. The results of this study are presented in Fig. 5A.

The results taking into consideration the proportion of correct looking time are presented in Fig. 5B. The analyses showed that infants' average performance in the 3-syllable sequence [Mean = 0.57; t(16) =0.92; $p > .05$; Cohen's d = 0.15] and the 4-syllable sequences [Mean = 0.55; t(15) = 0.63; $p > .05$; Cohen's d = 0.15] was not above chance.

Considering the longest look, the results showed that infants' mean difference score in the 3-syllable sequences (0.18) was not significantly above chance [t(16) = 1.91; $p > .05$; Cohen's d = 0.30]. Similarly, infants' mean difference score in the 4-syllable sequences (0.21) was not significantly above chance [t(15) = 1.11; $p > .05$; Cohen's d = 0.27]. The results are presented in Fig. 5C.

Finally, the results showed no modulation of the response latencies as a function of the sequence type (3 or 4 syllable-sequences). The mean response latencies were 830 ± 442 ms and 1083 ± 309 ms for the 3 and 4 syllable sequences respectively, and there was no statistical difference between them [t(9) = −1.75; $p > .05$; Cohen's d = 0.55, two-tails]. Seven infants did not show correct responses in at least one of the conditions.

Comparison between Experiments 1 and 3

As an additional analysis, we computed a combined repeated-measures ANOVA on the first look data with experiment (Experiment 1 vs. Experiment 3) and sequence type [smaller vs. bigger sequence] as main crossed factors. The results showed a main effect of sequence type [F(1,33) = 3.38, $p = .03$, $\eta_p^2 = 0.09$] as participants overall performed better in the smaller (mean estimate = 0.23) than in the larger sequence (mean estimate = −0.14). There were no significant effects of experiment [F(1,33) = 1.83, $p > .05$, $\eta_p^2 = 0.05$], and no interaction between factors [F(1,33) = 0.18, $p > .05$, $\eta_p^2 = 0.005$]. The results are depicted in Fig. 6.

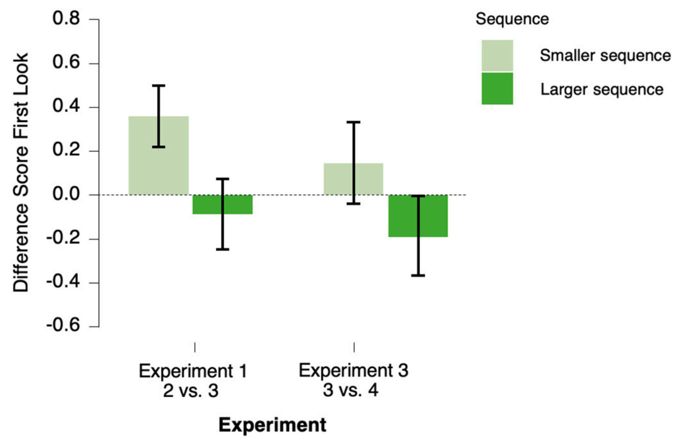

Fig. 6. Comparison of the infants' performance in Experiments 1 and 3. The y-axis shows the mean difference scores. Error bars depict standard errors of the mean.

4.5. Discussion

Experiment 3 provides no evidence that 9- to 10-month-old infants discriminate between auditory sequences of 3 versus 4 elements. This finding suggests a limit in the infants' ability to discriminate and represent speech units in working memory tests. It is important to note that infants successfully discriminated sequences that differed by only 1 unit (2 vs. 3 syllables) in Experiments 1 and 2, thus the absolute difference of 1 element used in Experiment 3 (3 vs. 4 syllables) was not the root of the difficulty.

5. Experiment 4: infants' ability to differentiate between sets of 2 and 3 tones

The results of experiments 1 and 2 suggest that infants at 9 to 10 months of age are able to accurately represent and distinguish small number of syllables, particularly 2 and 3 syllable sequences. The current experiment aims at directly testing the same distinction, using the same paradigm and with same aged infants, but with 2 and 3 tones (rather than syllables). If the infants' capacities to represent small number of auditory items are independent of the nature of the sounds, then infants should show a similar pattern of responses as in Experiment 1. Conversely, if infants posses better representational skills in relation to linguistic than to non-linguistic auditory stimuli, their performance might appear more limited when they are presented with tones.

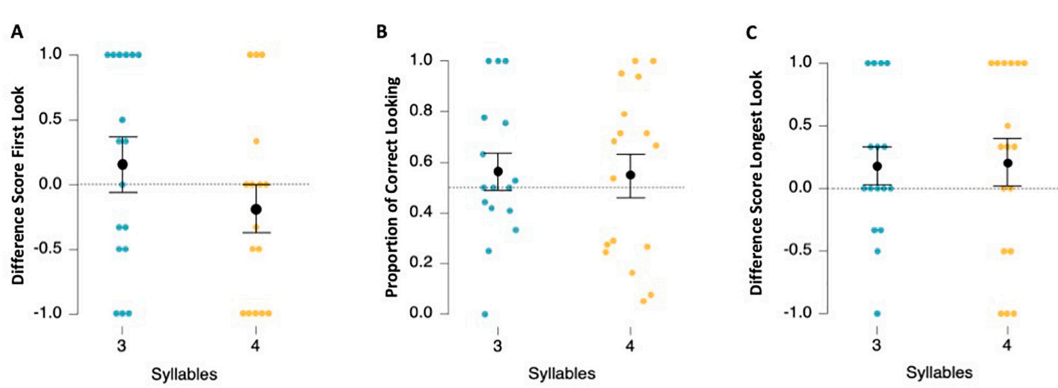

Fig. 5. Main findings of Experiment 3 – testing infants' ability to differentiate between sets of 3 and 4 syllables. A. Normalized difference scores computed over first look measures. B. Proportion of looking time towards the correct side of the screen. C. Normalized difference scores computed over the longest look. Colored dots represent individual participants' scores in each condition, the black dot indicates the group mean, and bars depict standard errors of the mean. The dotted line in the middle depicts chance level.

5.1. Participants

Sixteen infants, monolingual learners of Italian, participated in the study (9 males; M_{age} = 9 months and 25 days, range 8 months 14 days to 11 months 4 days). All participants were full term, with no birth complications, and no hearing or visual problems reported. Two additional infants were excluded from the analysis because of crying or fussiness ($N = 1$), or because the eye-movements were not clearly visible at the time of coding (N = 1).

Participants of this experiment were recruited online and lived in urban and rural areas of the Northern, Southern, and Central parts of Italy. The families were contacted by means of e-letters, on-line pamphlets, and social media advertisements. The testing took place between December 2020 and January 2021. Participants were all Caucasian, most of them were members of high to middle socioeconomic status families (15/16), and their parents hold a University degree (14/16). Infants' parents provided verbal and digital informed consent for the tests, demographical questions, and video recordings. The Institutional Ethics Committee of Psychology Research at Padua University approved the study (Protocol 3697).

5.2. Stimuli

Twenty different pure sine waves [250 Hz, 280 Hz, 310 Hz, 340 Hz, 370 Hz, 400 Hz, 430 Hz, 460 Hz, 490 Hz, 520 Hz, 550 Hz, 580 Hz, 610 Hz, 640 Hz, 670 Hz, 700 Hz, 850 Hz, 1000 Hz, 1150 Hz, 1300 Hz] were used in this experiment. The tones' pitch was selected from a range (250–1300 Hz) previously used in studies of tone discrimination in infants (VanMarle & Wynn, 2009). The sound sequences were created using Audacity® Cross-Platform Sound Editor, version 2.4.2. As in Experiment 1, the duration of each tone was 200 ms and there were pauses of 250 ms between tones. The intensity of the stimuli was 70 dB. For each participant, twelve different tones were randomly chosen for the familiarization trials (repeated either 2 or 3 times) and the remaining eight were exclusively used in the test phase, for generalization. The tones chosen for the familiarization and test varied randomly across participants. The visual stimuli were identical to those used in the previous experiments.

5.3. Procedure

The procedure was identical to that used in the previous experiments.

5.4. Data acquisition

The stimuli were presented online via the Labvanced platform (https://www.labvanced.com/). For the sake of methodological rigor, an instruction section for the parents (without the infant) preceded the actual test. Parents were guided to choose a luminous and quiet room in the house (where the participant faced the main source of light) to carry out the experiment. A personal computer (no tablets or cell phones) with a minimum screen size of 13″ was requested. A calibration procedure embedded in Labvanced was completed in order to guarantee that visual stimuli were identically displayed, despite screen size variations across computers. The quality of the Internet connection was verified by means of speedtest.net. Furthermore, the sound quality and intensity of the auditory stimuli were settled through an App (Niosh for Apple and Fonometro for Andriod) that parents installed on their cell phones and placed at the same distance to the computer were the infants were to be seated. Finally, parents were invited to close any App or program on the computer, to put the cell phone in silent mode, to turn off any other apparatus in the house (e.g. TV, radio), and to remove toys or other interesting objects from the infant's sight, which could interfere with the test.

During the test session, infants were seated on their parent's lap at 50 cm distance from the monitor. Infants' gaze, as well as the screen of the parents' personal computer was monitored and recorded with two computers via Zoom. The Zoom session was settled in such a way that infants could see only the stimuli displayed in full screen, while the experimenter recorded the image of the infant (in full screen) and the screen of the parents' personal computer (side-by-side floating mode) at the same time. The correct alignment between the stimulus presentation and the image of the infant was verified on a trial basis by means of short (20 ms) luminous flashes included at the end of each trial (the reflection of this light was quite evident on the baby's face). Two coders, independently of each other and naïve to the experimental conditions, performed an offline analysis of the videos by coding infants' eye movements frame-by-frame. The mean estimated reliability between observers was Pearson's $r = 0.87$, $p < .001$, computed across dependent variables on 13 out of 16 infants (81%).

5.5. Data analysis

The data analysis was identical to that used in Experiments 1–3.

5.6. Results

Infants fixated the center of the screen 88% of the total number of trials on average. The mean duration of their first fixation was 721 ± 231 ms. In the sequences with 2 repetitions of the tones, infants' mean difference score was −0.19, which was not significantly different from chance, [t(15) = −1.21; $p > .05$; Cohen d' = 0.30]. In the sequences with 3 tones, mean difference score was −0.04. This value did not differ from chance [t(15) = −0.32; $p > .05$; Cohen's d = 0.08]. Moreover, paired *t*-test comparing the infants' performance in the 2 and 3-tones trials showed that the results did not differ [t(15) = −0.64; $p > .05$; Cohen's d = 0.16]. The findings of this study are presented in Fig. 7A.

Considering the proportion of correct looking time, the results showed that infants' average performance in the sequences with 2 tones [Mean = 0.41; t(15) = −1.36; $p > .05$; Cohen's d = 0.36] and with 3 tones [Mean = 0.53; t(15) = 0.54; $p > .05$; Cohen's d = 0.14] was not above chance. The results are presented in Fig. 7B.

The results taking into consideration the longest look, the results showed that infants' mean difference score in the sequences with 2 tones (−0.22) was not significantly above chance [t(15) = −1.30; $p > .05$; Cohen's d = 0.33]. Similarly, infants' mean difference score in the 3-tones sequences (0.09) was not significantly above chance, [t(15) = 0.54; $p > .05$; Cohen's d = 0.13]. The results are presented in Fig. 7C.

The response latencies measures did not show modulation as a function of the sequence type (2 or 3 tones). The mean response latencies were 872 ± 440 ms and 821 ± 657 ms for the 2 and 3-tones respectively, and there were no statistical differences between them [t (11) =0.22; $p > .05$; Cohen's d = 0.06, two-tails]. Four infants did not show correct responses in at least one of the conditions, and were not included in this analysis.

Comparison between Experiments 1 and 4

We computed a combined repeated-measures ANOVA on the first look data with experiment (Experiment 1 vs. Experiment 4) and sequence type [2 vs. 3] as main crossed factors. The results showed a significant main effect of experiment [F(1,33) = 4.98, $p = .03$, $\eta_p^2 = 0.13$], no main effect of sequence type [F(1,33) = 0.66, $p > .05$, $\eta_p^2 = 0.02$] and no significant interaction between factors [F(1,33) = 2.82, $p > .05$, $\eta_p^2 = 0.08$]. The main effect of experiment was driven by the fact that participants generally performed significantly better in Experiment 1 (mean estimate = 0.14) than in Experiment 4 (mean estimate = −0.12). The results are depicted in Fig. 8.

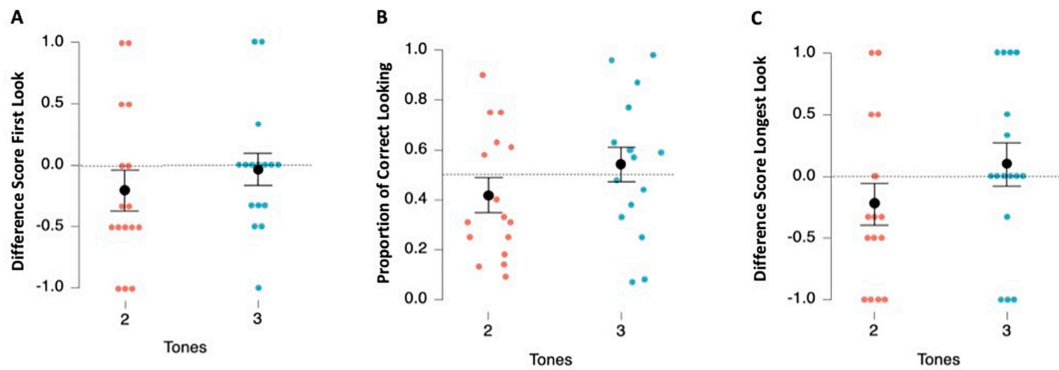

Fig. 7. Main findings of Experiment 4 – testing infants' ability to differentiate between sets of 2 and 3 tones. A. Normalized difference scores computed over first look measures. B. Proportion of looking time towards the correct side of the screen. C. Normalized difference scores computed over the longest look. Colored dots represent individual participants' scores in each condition, the black dot indicates the group mean, and bars depict standard errors of the mean. The dotted line in the middle depicts chance level.

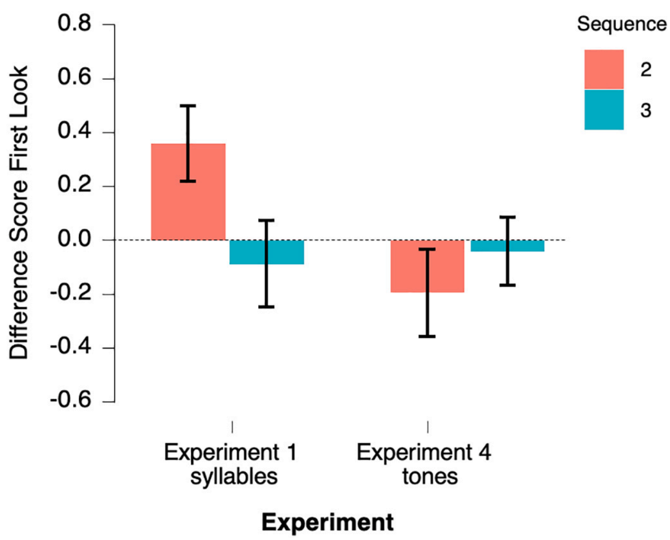

Fig. 8. Comparison of the infants' performance in Experiments 1 and 4. The y-axis shows the mean difference scores. Error bars depict standard errors of the mean.

5.7. Discussion

Experiment 4 provides no evidence that participants discriminate between auditory sequences of 2 versus 3 tones. These findings replicate previous works, which also reported that infants at 7 months of age (VanMarle & Wynn, 2009) and 9 months of age (Lipton & Spelke, 2004) were unable to make such a distinction with non-linguistic stimuli. Moreover, these findings show that infants fail at discriminating 2 and 3 tones also in a two-alternative looking procedure.

It should be noted that, for a better comparison with Experiments 1–3, we have presented this study in terms of tones being the unit on which numerosity can be computed. But one should keep in mind that participants naturally code relational properties, such as intervals (the frequency ratio of successive tones) or contours (the sequences of increments or decrements), which may yield different processing units (Trehub, 1987). It is thus conceivable that infants found the discrimination in Experiment 4 more difficult because of the ambiguity about which feature is relevant to compute numerosity (e.g. a tone's identity or the frequency interval between adjacent tones). It is as likely that participants did not discriminate the two conditions because they focused on the overall patterns of the two types of sequences (i.e. all the contours were flat and the intervals were zero within a given sequence), and not in the number of tones (2 vs. 3) or intervals (1 vs. 2) they contained.

Independently of what factor caused the infants' difficulty in this study, the present findings suggest that under similar conditions infants show different representational skills for small number of linguistic than non-linguistic auditory stimuli.

6. General discussion

The main finding of the present report is that the number of sounds that young infants can discriminate within the small numerosity range varies depending on the nature of the stimuli. Nine- to ten-month-olds successfully discriminate 2- and 3-syllable sequences in a two-alternative looking procedure (Experiments 1 and 2), just as 4-days-old infants do in a habituation paradigm, as reported by Bijeljac-Babic et al. (1993). On the other hand, infants failed to discriminate 2 vs. 3 non-linguistic sounds (Experiment 4) like infants did in other looking time habituation/familiarization tests (Lipton & Spelke, 2004; VanMarle & Wynn, 2009). Together, the available data suggest that more precise representations might operate over linguistic units than over non-linguistic sounds. In the following section, these findings are discussed in light of the proneness that infants display to speech sounds from birth onwards and the effects this might have over numerical representations of sounds.

6.1. Early sensitivity to language

Our findings, along with results of previous studies suggest that infants who heard 2- and 3-syllable sequences provide evidence of discrimination, while infants who heard the same number of tone sequences failed to make such a distinction. This might not be surprising considering that young infants are especially sensitive to linguistic sounds, and that this sensitivity is subserved by dedicated cortical structures since birth (e.g., Dehaene-Lambertz et al., 2010; Peña et al., 2003). Newborns adjust their sucking behavior to preferentially listen to speech (Vouloumanos & Werker, 2007) and maintain these biases in the first months of life, as substantiated by their listening preferentially and longer to speech compared with other non-linguistic sounds (Glenn, Cunningham, & Joyce, 1981; Vouloumanos & Werker, 2004).

Importantly, although auditory stimuli like tones and melodic sequences are also engaging to infants, these stimuli do not promote certain processes to the extent language does. For example, words facilitate object categorization in preverbal infants of 9-, 6-, and even 3-months of age, in a way that well-matched tone sequences do not (Balaban & Waxman, 1997; Ferry, Hespos, & Waxman, 2010; Fulkerson & Waxman, 2007). In the above-mentioned studies infants were familiarized to different exemplars of a category accompanied by either a labeling phrase or a tone sequence. During the test, infants viewed new within-category exemplars. Infants who heard labeling phrases provided evidence of categorization -as revealed by their reliable preference for the member of the novel category- while infants who heard tone sequences failed to form the category.

Young infants also treat differently words and other sounds, like instrumental music, *vis-á-vis* memory mechanisms. When they first start encoding word forms (devoid of meaning) newborns recognize the familiarized word after a silent pause of few minutes, as evidenced by differential hemodynamic responses to familiar and novel words in a subsequent test (Benavides-Varela et al., 2011; Benavides-Varela et al., 2017; Benavides-Varela, Gómez, & Mehler, 2011; Benavides-Varela, Hochmann, Macagno, Nespor, & Mehler, 2012). Notably, neural signatures of recognition are still observed in the test after newborns repeatedly heard instrumental excerpts but not speech sounds during the pause (Benavides-Varela, Gómez, Macagno, et al., 2011). Thus, not only newborns discriminate between speech and non-speech stimuli, but also their primitive memory and forgetting mechanisms are sensitive to this distinction, channeling speech and other auditory inputs separately. Similarly, older infants at 16 months of age can leverage their early sensitivities to linguistic distinctions to hierarchically reorganize their memory representations, thereby overcoming working memory limits (Stahl & Feigenson, 2018).

Altogether the literature suggests that linguistic sounds, even in preverbal infants, influence the functioning of various perceptual and cognitive mechanisms including attention, categorization, and memory. The current study adds to this body of literature by suggesting that linguistic sounds also become especially effective in highlighting certain properties, such as the number of items, which may otherwise be difficult to detect or encode in other auditory stimuli.

6.2. Speech units on which numerosity can be computed

This work probed the infants' ability to discriminate speech components, specifically the number of syllables that are often considered to be central to speech perception (e.g., Nazzi, Bertoncini, & Mehler, 1998; Räsänen, Doyle, & Frank, 2018). However, speech in particular, and the auditory stimuli in general can be equivocal at signaling the units that can be used to extract numerosity. For example, three sounds might represent three separate units, assuming that each of them is segmented as a word, but they could also represent more units considering the number of syllables or phonemes that constitute each of these words. The three words might also be considered a single unit if, taken together, they constitute a single phrase or if only one person produces the words (assuming that the task concerns identifying the number of voices).

Numerous studies have demonstrated that, for young pre-verbal infants, there might be some perceptual and language specific cues available in the rich speech signal that could facilitate grouping, segmenting and breaking into the speech code (Jusczyk, 2000; Werker, 2018). Changes in prosody, namely pitch, intensity, duration, and rhythmic and stress patterns assist infants in defining speech units such as words and phrases, and other units (e.g., Abboub, Nazzi, & Gervain, 2016; Benavides-Varela & Gervain, 2017; Bion et al., 2011; Langus, Mehler, & Nespor, 2017; Morgan, 1996) over which numerosity could be computed.

Whether these units -as individuals or as groups- are specifically relevant for a given task or more generally for the infants' native language, might have a strong influence in determining the expression and evaluation of the capacity to extract numerical information. For example, Mehler and colleagues showed that newborns detect changes in the number of 2 and 3 syllables, but do not respond to comparable changes in the number of sub-syllabic units (Bertoncini, Floccia, Nazzi, & Mehler, 1995) or individual phonetic segments (Bijeljac-Babic et al., 1993). These findings suggest that infants identify the key elements for the organization of their native language, and that these elements are spontaneously used for various computations, such as the representation of number.

Conversely, under some circumstances infants might undermine and even ignore numerical information and pay more attention to other properties of the signal that are relevant in their own language. For example, Jusczyk, Cutler, and Redanz (1993) investigated whether English-learning infants display sensitivity to the predominant stress pattern of their native-language words. Their results indicated that 9-month-olds are capable of discriminating separate lists of words provided that one of the lists followed (i.e., strong/weak) and the other one did not follow (i.e., weak/strong) the predominant stress patterns in English, and in spite of the fact that the two lists contained words with exactly the same number of syllables.

Thus, although discrimination of speech units in the small numerosity range seems part of the infants' competence, the generalizability of this ability to other speech units should be evaluated with caution. The infant picks up and organizes the information along a number of multidimensional properties. The maturational level of the infant, as well as the amount of exposure to the native language, and requirements of the specific language task the infant is facing might work together to differentially direct attention to one (or more) of these properties (Werker & Curtin, 2005). Further studies should be undertaken to better define whether and under which circumstances infants do use numerical distinctions in the context of fluent speech.

6.3. The properties of the task and the cognitive capacities involved

This investigation implemented an adaptation of the two-alternative looking paradigm, successfully used in previous infant research to study a variety of cognitive functions such as categorization, generalization, and memory (e.g. Addyman & Mareschal, 2010; Albareda-Castellot et al., 2011; Benavides-Varela & Mehler, 2015; Hochmann et al., 2011; Hochmann et al., 2018; Kovács & Mehler, 2009a, 2009b; McMurray & Aslin, 2004; Shukla, Wen, White, & Aslin, 2011). Notably, this paradigm places more difficult demands than traditional habituation/familiarization or preference tests (McMurray & Aslin, 2004; Shukla et al., 2011). In the task implemented in this study, in order to succeed, infants needed to discriminate between two large sets of variable syllable sequences by forming abstract categories (like in habituation or preference tests, to some extent). Additionally, infants were required to rapidly switch attention between the two classes of intermixed stimuli, and to assign opposite predictions (left-right) to these two classes. Moreover, in the test –in the absence of the visual stimuli– participants also needed to remember the correct association with the corresponding side of the screen and to generalize to new sets of variable syllables. Thus, a success or a failure in this task should be interpreted with caution, taking into consideration the various cognitive functions involved. The results of experiments 1 and 2, for instance, indicated that infants discriminated between 2 and 3 syllable sequences, but were able to learn and generalize only one of the sequences. On the one hand, this shows that participants were able to extract an organizational property (i.e. number of syllables) and to sort out large sets of variable speech stimuli. On the other hand, these findings suggest that infants remembered and performed additional computations only over a portion of information, possibly due to limitations in executive functions and working memory. Indeed, following discrimination, infants seemed to focus on the type of stimuli that engaged less memory resources (2-syllables as opposed to 3-syllables-long sequences). Hence, this task approaches the highly complex nature of the online processing and may provide additional hints regarding the extent to which infants are able to apply discrimination capacities while processing multiple stimuli online.

It is also possible that the complexity of this task might have masked some of the infants' pure discrimination capacities. Empirically, one may attempt to lower the difficulty of the test for example by increasing the number of trials during the familiarization and in this way give infants more opportunities to process the stimuli. However, this might also increase the attrition rate as the experiment could turn out being too long for some of the participants. It would also be possible to intermingle novel stimuli and familiar stimuli (with puppet presentations) during the test in order to reduce memory demands, or to adapt the length of the familiarization phase according to each individual's performance in order to ensure that infants proceed to the test when they show evidence of learning both types of stimuli (see also Shukla et al., 2011).

Despite being a highly challenging task, additional modifications could also be implemented in the future in order to ask whether infants succeed if even more complex stimuli were included. Of particular interest for language acquisition studies could be the incorporation of properties that naturally vary in speech (e.g. irregular prosodic contours, syllable types within sequences, voices, etc.). For numerical cognition studies, a relevant manipulation concerns the variation of temporal properties that co-vary with number (e.g. duration and rate) within a single study. Although the results of Experiments 1 and 2 separately showed that infants do not used these cues for discrimination, we cannot exclude the possibility that participants employed different strategies to discriminate elements across experiments. Nevertheless, whatever the strategy to discriminate linguistic stimuli participants used, it was unavailable to infants in Experiment 4, for discriminating non-speech sounds.

The consistency of the results across different dependent measures also deserves some discussion. The results of the first look behavior (our primary dependent variable) were highly consistent with additional measures considering longest look, and proportion of correct looking time in the test. Latency, differently from these three measures, did not provide differential responses in the two conditions. The latter finding may reflect infants learning the fixed time structure of the events. During the familiarization, puppets always appeared 1000 ms after the auditory stimuli and in the test, the first fixations of the infants started after about that timing, on average. Thus, it is possible that response times measures might be more informative in other circumstances, for instance in studies varying the delay between the offset of the auditory stimulus and the onset of the reinforcer in the familiarization phase.

Finally, it is necessary to bear in mind that the results of the familiarization (Supplementary materials-2), although in some cases similar, are not directly comparable to those of the test phase for at least two related reasons. First, as described in the Methods section, during familiarization infants could make anticipatory looks during 1 s after the offset of the linguistic stimulus and before the onset of the reward. In the test, by contrast, the time window for anticipatory looks was much longer (2 s). Moreover, as discussed above, most infants appear to have learned the fix timing of the puppet's appearance during the familiarization. As a consequence, a limited amount of anticipatory looks was observed throughout the familiarization trials (35%–56%). Thus, the analysis performed on the anticipatory looks in the familiarization phase might be uninformative due to the limited amount of data. This high rate of missing values is not unusual, previous studies report anticipations from 25% to 45% in infant studies (e.g. Johnson, Amso, & Slemmer, 2003; Kovács, 2008).

To summarize, whether the robustness of these findings changes or generalizes to other tasks with different requirements remains to be attested in future studies. From the current studies it seems clear, however, that under identical conditions and task demands, the infants' discrimination and processing capacities of linguistic stimuli appear greater than those of non-linguistic sounds.

6.4. The nature of the numerical representation elicited by speech units

Another important issue raised by the current set of experiments concerns the properties and the limits of the cognitive system that consent encoding small number of elements in the speech signal.

Infants successfully discriminated 2- and 3-syllabic sequences (Experiments 1 and 2) but in Experiment 3 there was no evidence that infants discriminated 3 and 4-syllable sequences, suggesting that the representational system approached its upper limit capacity.

Two alternative explanations could be compatible with these results. First, infants may have encoded the numerosity of speech sounds in terms of the Approximate Number System (ANS). Under this view, the capacity and limits of discrimination observed could be adduced to ratio limits described by the Weber's Law. Indeed, the limits observed for linguistic stimuli in the present study are similar to those reported in the visual domain. Infants of this age typically fail at 3:4 ratios but not at 2:3 ratios with large numbers of visual items (see Section 1.1). Moreover, the discrimination function obtained here resembles that obtained for duration discrimination of audio-visual events in similar aged infants (Brannon et al., 2007). This alternative seems also consistent with a previous proposal arguing that infants use the ANS to represent tone sequences (VanMarle & Wynn, 2009) and, at first sight, with the general idea that the precision of this system increases with age. Indeed, in the study of VanMarle and Wynn (2009) 7-month-old infants discriminated tones differing by 1:2 ratio but not by 2:3 ratios, and in the current study 9- to 10-month-old infants succeeded in the 2:3 (syllables) but not in the 3:4 ratio. By accepting this interpretation, however, one should acknowledge a unique developmental path of the ANS: infants at birth succeed in the 2:3 ratio (Bijeljac-Babic et al., 1993), fail with the same ratio at 7 months of age (VanMarle & Wynn, 2009), and succeed again at 9 to 10 months of age (Experiments 1 and 2). It is also peculiar, that the computations necessary for discriminating 2 vs. 3 by means of the ANS, whose features are considered to apply cross-modally (Brannon et al., 2007), were available to 9–10 month-old infants when processing auditory linguistic elements, but apparently undefined or very poor for processing small-number of non-linguistic sounds (Experiment 4 and Lipton & Spelke, 2004).

Alternatively, infants might have access to a capacity limited "auditory event file system" delimited by the number of items that can be tracked and held in working memory (Hauser, Dehaene, Dehaene-Lambertz, & Patalano, 2002), and provided that the input sounds are linguistic in nature. Non-linguistic sound sequences in the small range might not access this representational system, or might be represented as a single mental numerical magnitude with an approximate representation of numerosity, only under some circumstances (i.e. VanMarle & Wynn, 2009). This explanation fits the results of the current experiments and also conforms to the findings of previous investigations assessing discrimination of syllables in newborns and non-linguistic stimuli in infants. For these reasons, we tentatively favor the "auditory event file" hypothesis as an explanation of the findings. However, additional evidence should be gathered in order to determine whether infants represent auditory linguistic objects in such a system and whether the properties of this system are comparable to those of the OFS.

An intriguing question, for instance, regards set size limit, and to what extent it resembles that of the visual system (i.e. 3 individual items). Here, In Experiments 1 and 2, infants could have successfully discriminated 2 from 3 syllables because, indeed, both sets fall within the limit of three. Then, in Experiment 3, participants might have failed to tell apart 3 from 4 syllables, because 4 exceeds this boundary. However, the current set of studies provides weak support that infants can represent three auditory objects exactly. The data only showed direct evidence that infants performed above chance with 2-syllable sequences in the test. If we were to know whether infants could also represent 3 auditory elements, their ability to associate the 3-syllable sequences to the corresponding side of the screen and to remember this sequence in the test, must be attested with the same task. A study addressing the infants' behavior when, for instance, 3-syllable sequences are associated to one side, and sequences of random number of elements are associated to the other side of the screen, might contribute to clarify this fundamental aspect in the future.

7. Conclusion

The results presented here provide a test of small number discrimination in the auditory domain using syllable and tone sequences. In two experiments 9- to 10-month-old infants successfully discriminated 2- and 3-syllable sequences on the basis of their numerosity when continuous variables were controlled. In contrast, infants failed to discriminate 3- from 4-syllable sequences and 2- from 3-tone sequences under similar conditions. We argue that infants are able to organize the auditory information along discriminable units and that the access to a precise representational system might be facilitated by sounds that are ecologically relevant to the infant, namely those of linguistic nature.

Author contributions

Silvia Benavides-Varela: Conceptualization; Methodology; Software; Investigation; Data curation; Formal analysis; Writing - original draft; Funding acquisition.

Natalia Reoyo-Serrano: Software, Investigation, Data curation; Validation; Visualization; Writing - review & editing.

Funding

This work was supported by the University of Padua under the STARS Grants programme (MINI) and by the Italian Ministry of Instruction, University and Research "Progetti di Rilevante Interesse Nazionale (PRIN)" (Prot. 2017PSRHPZ).

Acknowledgments

We thank Aldo Antonio Sarubbi, Chiara Nascimben, and Antonela Pepaj for their assistance in testing and coding. We also thank the parents and infants who participated in this research, Richard Aslin and two anonymous reviewers for their valuable comments on earlier versions of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2021.104637.

References

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

Addyman, C., & Mareschal, D. (2010). The perceptual origins of the abstract same/different concept in human infants. *Animal Cognition, 13*(6), 817–833.

Agrillo, C., & Bisazza, A. (2018). Understanding the origin of number sense: A review of fish studies. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1740), 20160511.

Agrillo, C., Dadda, M., & Bisazza, A. (2007). Quantity discrimination in female mosquitofish. *Animal Cognition, 10*(1), 63–70.

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

Ansari, D., Lyons, I. M., van Eimeren, L., & Xu, F. (2007). Linking visual attention and number processing in the brain: The role of the temporo-parietal junction in small and large symbolic and nonsymbolic number comparison. *Journal of Cognitive Neuroscience, 19*(11), 1845–1853.

Antell, S. E., & Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Development*, 695–701.

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology, 64*(1), 3–26.

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition, 86*(3), 201–221.

Benavides-Varela, S., Basso Moro, S., Brigadoi, S., Meconi, F., Doro, M., Simion, F., … Dell'Acqua, R. (2018). N2pc reflects two modes for coding the number of visual targets. *Psychophysiology, 55*(11), Article e13219.

Benavides-Varela, S., & Gervain, J. (2017). Learning word order at birth: A NIRS study. *Developmental Cognitive Neuroscience, 25*, 198–208.

Benavides-Varela, S., Gómez, D. M., Macagno, F., Bion, R. A., Peretz, I., & Mehler, J. (2011). Memory in the neonate brain. *PLoS One, 6*(11), Article e27497.

Benavides-Varela, S., Gómez, D. M., & Mehler, J. (2011). Studying neonates' language and memory capacities with functional near-infrared spectroscopy. *Frontiers in Psychology, 2*, 64.

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.

Benavides-Varela, S., & Mehler, J. (2015). Verbal positional memory in 7-month-olds. *Child Development, 86*(1), 209–223.

Benavides-Varela, S., Siugzdaite, R., Gómez, D. M., Macagno, F., Cattarossi, L., & Mehler, J. (2017). Brain regions and functional interactions supporting early word recognition in the face of input variability. *Proceedings of the National Academy of Sciences, 114*(29), 7588–7593.

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.

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

Bion, R. A., Benavides-Varela, S., & Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech, 54*(1), 123–140.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341–345.

Brannon, E. M., Roussel, L. W., Meck, W. H., & Woldorff, M. (2004). Timing in the baby brain. *Cognitive Brain Research, 21*(2), 227–233.

Brannon, E. M., Suanda, S., & Libertus, K. (2007). Temporal discrimination increases in precision over development and parallels the development of numerosity discrimination. *Developmental Science, 10*(6), 770–777.

Cantrell, L., & Smith, L. B. (2013). Open questions and a proposal: A critical review of the evidence on infant numerical abilities. *Cognition, 128*(3), 331–352.

Carey, S. (2011). Précis of the origin of concepts. *Behavioral and Brain Sciences, 34*(3), 113.

Castaldi, E., Piazza, M., Dehaene, S., Vignaud, A., & Eger, E. (2019). Attentional amplification of neural codes for number independent of other quantities along the dorsal visual stream. *Elife, 8*, Article e45160.

Clearfield, M. W., & Mix, K. S. (1999). Number versus contour length in infants' discrimination of small visual sets. *Psychological Science, 10*(5), 408–411.

Clearfield, M. W., & Mix, K. S. (2001). Amount versus number: Infants' use of area and contour length to discriminate small sets. *Journal of Cognition & Development, 2*(3), 243–260.

Cordes, S., & Brannon, E. M. (2008). Quantitative competencies in infancy. *Developmental Science, 11*(6), 803–808.

Cordes, S., & Brannon, E. M. (2009a). Crossing the divide: Infants discriminate small from large numerosities. *Developmental Psychology, 45*(6), 1583.

Cordes, S., & Brannon, E. M. (2009b). The relative salience of discrete and continuous quantity in young infants. *Developmental Science, 12*, 453–463.

Coubart, A., Izard, V., Spelke, E. S., Marie, J., & Streri, A. (2014). Dissociation between small and large numerosities in newborn infants. *Developmental Science, 17*(1), 11–22.

Coubart, A., Streri, A., de Hevia, M. D., & Izard, V. (2015). Crossmodal discrimination of 2 vs. 4 objects across touch and vision in 5-month-old infants. *PLoS One, 10*(3), Article e0120868.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.

Dehaene, S. (2001). Précis of the number sense. *Mind & Language, 16*(1), 16–36.

Dehaene-Lambertz, G., Montavont, A., Jobert, A., Allirol, L., Dubois, J., Hertz-Pannier, L., & Dehaene, S. (2010). Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain and Language, 114*(2), 53–65.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van der Vrecken, O. (1996, October). The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. In *, Vol. 3. Proceeding of fourth international conference on spoken language processing. ICSLP'96* (pp. 1393–1396). IEEE.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*(3968), 303–306.

Elfenbein, J. L., Small, A. M., & Davis, J. M. (1993). Developmental patterns of duration discrimination. *Journal of Speech, Language, and Hearing Research, 36*(4), 842–849.

Feigenson, L. (2007). The equality of quantity. *Trends in Cognitive Sciences, 11*(5), 185–187.

Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science, 6*, 568–584.

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition, 97*(3), 295–313.

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science, 13*(2), 150–156.

Feigenson, L., Carey, S., & Spelke, E. (2002). Infants' discrimination of number vs. continuous extent. *Cognitive Psychology, 44*(1), 33–66.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314.

Féron, J., Gentaz, E., & Streri, A. (2006). Evidence of amodal representation of small numbers across visuo-tactile modalities in 5-month-old infants. *Cognitive Development, 21*(2), 81–92.

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3-and 4-month-old infants: An advantage of words over tones. *Child Development, 81*(2), 472–479.

Friedrich, M., & Friederici, A. D. (2017). The origins of word learning: Brain responses of 3-month-olds indicate their rapid association of objects and words. *Developmental Science, 20*(2), Article e12357.

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6-and 12-month-olds. *Cognition, 105*(1), 218–228.

Gelman, R., & Butterworth, B. (2005). Number and language: How are they related? *Trends in Cognitive Sciences, 9*(1), 6–10.
Glenn, S. M., Cunningham, C. C., & Joyce, P. F. (1981). A study of auditory preferences in nonhandicapped infants and infants with Down's syndrome. *Child Development*, 1303–1307.
Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience, 5*(11), 887–892.
Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*(5), 1457.
Hauser, M. D., Dehaene, S., Dehaene-Lambertz, G., & Patalano, A. L. (2002). Spontaneous number discrimination of multi-format auditory stimuli in cotton-top tamarins (Saguinus oedipus). *Cognition, 86*(2), B23–B32.
de Hevia, M. D. (2016). Core mathematical abilities in infants: Number and much more. In *, Vol. 227. Progress in brain research* (pp. 53–74). Elsevier.
de Hevia, M. D., Addabbo, M., Nava, E., Croci, E., Girelli, L., & Macchi-Cassia, V. (2017). Infants' detection of increasing numerical order comes before detection of decreasing number. *Cognition, 158*, 177–188.
Hochmann, J. R., Benavides-Varela, S., Fló, A., Nespor, M., & Mehler, J. (2018). Bias for vocalic over consonantal information in 6-month-olds. *Infancy, 23*(1), 136–151.
Hochmann, J. R., Benavides-Varela, S., Nespor, M., & Mehler, J. (2011). Consonants and vowels: Different roles in early language acquisition. *Developmental Science, 14*(6), 1445–1458.
Hyde, D. C. (2011). Two systems of non-symbolic numerical cognition. *Frontiers in Human Neuroscience, 5*, 150.
Hyde, D. C., & Spelke, E. S. (2009). All numbers are not equal: An electrophysiological investigation of small and large number representations. *Journal of Cognitive Neuroscience, 21*(6), 1039–1053.
Hyde, D. C., & Spelke, E. S. (2011). Neural signatures of number processing in human infants: Evidence for two core systems underlying numerical cognition. *Developmental Science, 14*(2), 360–371.
Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences, 106*(25), 10382–10385.
Jensen, J. K., & Neff, D. L. (1993). Development of basic auditory discrimination in preschool children. *Psychological Science, 4*(2), 104–107.
Johnson, S. P., Amso, D., & Slemmer, J. A. (2003). Development of object concepts in infancy: Evidence for early learning in an eye-tracking paradigm. *Proceedings of the National Academy of Sciences, 100*(18), 10568–10573.
Jordan, K. E., & Brannon, E. M. (2006). The multisensory representation of number in infancy. *Proceedings of the National Academy of Sciences, 103*(9), 3486–3489.
Jordan, K. E., Suanda, S. H., & Brannon, E. M. (2008). Intersensory redundancy accelerates preverbal numerical competence. *Cognition, 108*(1), 210–221.
Jusczyk, P. W. (2000). *The discovery of spoken language*. MIT Press.
Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64*(3), 675–687.
Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*(2), 175–219.
Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science (Wiley-Blackwell), 18*(12), 1084–1089.
Kobayashi, T., Hiraki, K., & Hasegawa, T. (2005). Auditory–visual intermodal matching of small numerosities in 6-month-old infants. *Developmental Science, 8*(5), 409–419.
Kovács, Á. M. (2008). *Learning two languages simultaneously: Mechanisms recruited for dealing with a mixed linguistic input*. Trieste: SISSA (PhD dissertation).
Kovács, Á. M., & Mehler, J. (2009a). Flexible learning of multiple speech structures in bilingual infants. *Science, 325*(5940), 611–612.
Kovács, Á. M., & Mehler, J. (2009b). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences, 106*(16), 6556–6560.
Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition, 80*(1–2), 97–126.
Langus, A., Mehler, J., & Nespor, M. (2017). Rhythm in language acquisition. *Neuroscience & Biobehavioral Reviews, 81*, 158–166.
Libertus, M. E., & Brannon, E. M. (2017). Stable individual differences in number discrimination in infancy. *Developmental science, 13*(6), 900–906.
Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense: Large-number discrimination in human infants. *Psychological Science, 14*(5), 396–401.
Lipton, J. S., & Spelke, E. S. (2004). Discrimination of large and small numerosities by human infants. *Infancy, 5*(3), 271–290.
McCrink, K., & Wynn, K. (2007). Ratio abstraction by 6-month-old infants. *Psychological Science, 18*(8), 740–745.
McMurray, B., & Aslin, R. N. (2004). Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy, 6*(2), 203–229.
Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). Multiple cues for quantification in infancy: Is number one of them? *Psychological Bulletin, 128*(2), 278.
Mix, K. S., Levine, S. C., & Huttenlocher, J. (1997). Numerical abstraction in infants: Another look. *Developmental Psychology, 33*(3), 423.
de la Mora, D., Nespor, M., & Toro, J. M. (2013). Do humans and nonhuman animals share the grouping principles of the iambic–trochaic law? *Attention, Perception, & Psychophysics, 75*(1), 92–100.
Morgan, J. L. (1996). Prosody and the roots of parsing. *Language and Cognitive Processes, 11*(1–2), 69–106.
Mou, Y., & vanMarle, K. (2014). Two core systems of numerical representation in infants. *Developmental Review, 34*(1), 1–25.
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.
Peña, M., Maki, A., Kovačić, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., & Mehler, J. (2003). Sounds and silence: An optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences of the United States of America, 100*(20), 11702–11705.
Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America, 106*(26), 10598–10602. https://doi.org/10.1073/pnas.0904134106.
Posid, T., & Cordes, S. (2015). The small–large divide: A case of incompatible numerical representations in infancy. In *, Vol. 1. Mathematical cognition and learning* (pp. 253–276). Elsevier.
Räsänen, O., Doyle, G., & Frank, M. C. (2018). Pre-linguistic segmentation of speech into syllable-like units. *Cognition, 171*, 130–150.
Rugani, R., Cavazzana, A., Vallortigara, G., & Regolin, L. (2013). One, two, three, four, or is there something more? Numerical discrimination in day-old domestic chicks. *Animal Cognition, 16*(4), 557–564.
Rugani, R., & De Hevia, M. D. (2017). Number-space associations without language: Evidence from preverbal human infants and non-human animal species. *Psychonomic Bulletin & Review, 24*(2), 352–369.
Rugani, R., Regolin, L., & Vallortigara, G. (2008). Discrimination of small numerosities in young chicks. *Journal of Experimental Psychology: Animal Behavior Processes, 34*(3), 388.
Rugani, R., Vallortigara, G., & Regolin, L. (2013). Numerical abstraction in young domestic chicks (Gallus gallus). *PLoS One, 8*(6), Article e65262.
Rugani, R., Vallortigara, G., & Regolin, L. (2014). From small to large: Numerical discrimination by young domestic chicks (Gallus gallus). *Journal of Comparative Psychology, 128*(2), 163.
Semenza, C., & Benavides-Varela, S. (2018). Reassessing lateralization in calculation. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1740), 20170044.
Shukla, M., Wen, J., White, K. S., & Aslin, R. N. (2011). SMART-T: A system for novel fully automated anticipatory eye-tracking paradigms. *Behavior Research Methods, 43*(2), 384–398.
Stahl, A. E., & Feigenson, L. (2018). Infants use linguistic group distinctions to chunk items in memory. *Journal of Experimental Child Psychology, 172*, 149–167.
Starkey, P., & Cooper, R. G. (1980). Perception of numbers by human infants. *Science, 210*(4473), 1033–1035.
Starkey, P., Spelke, E. S., & Gelman, R. (1983). Detection of intermodal numerical correspondences by human infants. *Science, 222*(4620), 179–181.
Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition, 36*(2), 97–127.
Strauss, M. S., & Curtis, L. E. (1981). Infant perception of numerosity. *Child Development*, 1146–1152.
Trehub, S. E. (1987). Infants' perception of musical patterns. *Perception & Psychophysics, 41*(6), 635–641.
Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review, 101*(1), 80.
Vallortigara, G., Chiandetti, C., Rugani, R., Sovrano, V. A., & Regolin, L. (2010). Animal cognition. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(6), 882–893.
Van Loosbroek, E., & Smitsman, A. W. (1990). Visual perception of numerosity in infancy. *Developmental Psychology, 26*(6), 916.
VanMarle, K., & Wynn, K. (2006). Six-month-old infants use analog magnitudes to represent duration. *Developmental Science, 9*(5), F41–F49.
VanMarle, K., & Wynn, K. (2009). Infants' auditory enumeration: Evidence for analog magnitudes in the small number range. *Cognition, 111*(3), 302–316.
Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science, 7*(3), 270–276.
Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science, 10*(2), 159–164. https://doi.org/10.1111/j.1467-7687.2007.00549.x.
Warren, J. D., & Griffiths, T. D. (2003). Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *Journal of Neuroscience, 23*(13), 5799–5804.
Werker, J. F. (2018). Perceptual beginnings to language acquisition. *Applied PsychoLinguistics, 39*(4), 703–728.
Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*(2), 197–234.
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.
Wood, J. N., & Spelke, E. S. (2005). Infants' enumeration of actions: Numerical discrimination and its signature limits. *Developmental Science, 8*(2), 173–181.
Wynn, K. (1992). Addition and subtraction by human infants. *Nature, 358*(6389), 749–750.
Wynn, K., Bloom, P., & Chiang, W. C. (2002). Enumeration of collective entities by 5-month-old infants. *Cognition, 83*(3), B55–B62.
Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition, 89*(1), B15–B25.
Xu, F., & Arriaga, R. I. (2007). Number discrimination in 10-month-old infants. *British Journal of Developmental Psychology, 25*(1), 103–108.
Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*(1), B1–B11.