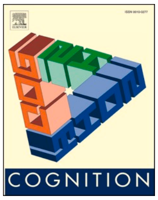
Cognition 243 (2024) 105688

Contents lists available at ScienceDirect

Cognition

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



Linking vestibular, tactile, and somatosensory rhythm perception to language development in infancy

Sofia Russo [a,*], Filippo Carnovalini [b], Giulia Calignano [a], Barbara Arfé [a], Antonio Rodà [b], Eloisa Valenza [a]

[a] Department of Developmental Psychology and Socialization, University of Padua, Padova, Italy
[b] Department of Department of Information Engineering, University of Padua, Padova, Italy

ARTICLE INFO

Keywords:
Rhythm processing
Language acquisition
Cross-domain transfer
Cross-sensory transfer
Pupillometry
Mixed-effects models

ABSTRACT

First experiences with rhythm occur in the womb, with different rhythmic sources being available to the human fetus. Among sensory modalities, vestibular, tactile, and somatosensory perception plays a crucial role in early processing. However, a limited number of studies so far have specifically focused on VTS rhythms in language development. The present work investigated VTS rhythmic abilities and their role in language acquisition through two experiments with 45 infants (21 females, sex assigned at birth; M age = 661.6 days, SD = 192.6) with middle/high socioeconomic status. Specifically, 37 infants from the original sample completed Experiment 1, assessing VTS rhythmic abilities through a vibrotactile tool for music perception. In Experiment 2, linguistic abilities were evaluated in 40 participants from the same cohort, specifically testing phonological and prosodic processing. Discrimination abilities for rhythmic and linguistic stimuli were inferred from changes in pupil diameter to contingent visual stimuli over time, through a Tobii X-60 eye-tracker. The predictive effect of VTS rhythmic abilities on linguistic processing and the developmental changes occurring across ages were explored in the 32 infants who completed both Experiments 1 and 2 by means of generalized, additive and linear, mixed-effect models. Results are discussed in terms of cross-sensory (i.e., haptic to hearing) and cross-domain (i.e., music to language) effects of rhythm on language acquisition, with implications for typical and atypical development.

1. Introduction

Rhythm perception, that is, the ability to encode cyclic or periodic repetition of events over time, is deeply rooted in the human brain (Nobre & Van Ede, 2018). It originates from our need for social interaction and bonding and has evolved hand-to-hand with communication and language (Patel, 2021). In the ontogenetic evolution, early experiences with rhythm occur already in the womb and encompass multiple sources of sensory stimulation (Lecanuet & Schaal, 2002). From the third trimester of gestation, fetuses can perceive rhythmic signals from the intrauterine and extra-uterine environment through bone conduction (Sohmer, Perez, Sichel, Priner, & Freeman, 2001). For instance, the maternal voice is transmitted directly to the amniotic fluid via body tissues and bones, with F0 and the first overtones being fully conducted through the spine and the pelvic arch (Lecanuet & Granier-Deferre, 1993). Similarly, external low frequencies (up to 500 Hz) are well perceived in utero. However, these multiple sources of rhythmic stimulation also include motor patterns produced by the mother's body movements (Lecanuet & Schaal, 2002); indeed, maternal heartbeat, breathing, and walking generate movements together with sound (Kisilevsky, Hains, Jacquet, Granier-Deferre, & Lecanuet, 2004). In sum, several kinds of rhythmic patterns are available to the developing brain since the very beginning (Granier-Deferre, Ribeiro, Jacquet, & Bassereau, 2011), conveyed by the temporal synchrony generated across auditory and vestibular-tactile-somatosensory (VTS) modalities (Provasi, Anderson, & Barbu-Roth, 2014).

Interestingly, this early imprinting with rhythm continues to affect newborn's behaviors after birth (Ullal-Gupta et al., 2013). For instance, early exposure to the isochronous rhythm of maternal heartbeat is hypothesized to direct children's listening preferences toward regular, binary meters during infancy (DeCasper & Sigafoos, 1983; Doheny, Hurwitz, Insoft, Ringer, & Lahav, 2012; Lahav, Saltzman, & Schlaug,

---

* Corresponding author at: Department of Developmental Psychology and Socialization, University of Padua, via Venezia 8, 35131 Padova, Italy.

E-mail addresses: sofia.russo@unipd.it (S. Russo), filippo.carnovalini@dei.unipd.it (F. Carnovalini), giulia.calignano@unipd.it (G. Calignano), barbara.arfe@unipd.it (B. Arfé), roda@dei.unipd.it (A. Rodà), eloisa.valenza@unipd.it (E. Valenza).

https://doi.org/10.1016/j.cognition.2023.105688
Received 14 June 2023; Received in revised form 30 November 2023; Accepted 6 December 2023

2007; Teie, 2016). Moreover, newborns seem also very proficient at discriminating rhythms, showing preference for listening to lullabies perceived prenatally versus novel ones (DeCasper & Spence, 1986). Afterbirth effects of early rhythmic experiences are shown even in premature infants, who adapt their breathing rate to the rhythmic structure of the VTS stimulation they receive (Zimmerman & Barlow, 2012).

In summary, the prenatal, crossmodal experiences with rhythm seem to contribute in shaping the rhythmic perception and behavior of newborns after birth.

Remarkably, perceiving rhythmic information in VTS sensory modality has been found to influence the encoding of auditory rhythmic stimuli also later in development (Phillips-Silver & Trainor, 2005). Passive rhythmic movements eliciting VTS, scaffolding auditory perception, continue to be present after birth. For instance, caregivers across cultures use to rock their newborns back and forth while singing or walking to comfort or let them sleep (Provasi et al., 2014). This pre- and after-birth parents' walking rate has been found to influence the spontaneous motor tempo of young infants (Rocha, Southgate, & Mareschal, 2021). Furthermore, a series of classical studies by Phillips-Silver and Trainor (2005, 2007, 2008) has shown that passive movements generating VTS stimulation alone influence the perception of auditory rhythm in infancy and adulthood. Specifically, Phillips-Silver and Trainor (2005) showed that 7-months-old infants who were bounced in synchrony to a given rhythm preferred to listen to a matching auditory stimulus compared to a mismatching one. By contrast, watching the rhythmic movements being performed by someone else did not elicit the same effect, thus suggesting that a direct engagement of the body is crucial in influencing rhythmic encoding at this age (Phillips-Silver & Trainor, 2007). In a subsequent study with adult participants, the authors demonstrated that VTS alone was sufficient to guide auditory rhythm processing (Phillips-Silver & Trainor, 2008). In fact, both passive head movements as well as the direct stimulation of the vestibular nerve was found to bias the encoding of auditory rhythms (Phillips-Silver & Trainor, 2008; Trainor, Gao, Lei, Lehtovaara, & Harris, 2009). Overall, these findings show that movements eliciting VTS perception shape the early development of structural and functional mechanisms underlying rhythm processing, thus influencing rhythm perception right after birth as along the lifespan. An explanation of the concurrent and longitudinal association between VTS rhythm perception and auditory rhythm encoding is that, both in the adult's and in the infant's brain, the encoding of VTS input conveyed by physical proximity to a sound source consists of the same energy (i.e., vibratory) used to encode auditory stimuli (Ammirante, Patel, & Russo, 2016).

Moreover, mechanoreceptors and ear cells are similarly structured and comparable in response characteristics such as the loudness summation of tones closely spaced in frequency (Hollins & Roy, 1996). Consistently, VTS inputs have been found to activate the auditory cortex (Caetano & Jousmäki, 2006), with VTS and auditory inputs being confused when presented simultaneously or in alternation (Gescheider & Niblette, 1967). Therefore, vibrations might evoke comparable low-level responses between VTS and auditory modalities. Based on these findings on VTS-auditory coupling, Tichko, Kim, and Large (2021); Tichko, Kim, Large, and Loui (2022) postulated a nonlinear, dynamical system in which two oscillatory neural networks, representing the auditory and motor systems, interact through weak, non-specific coupling. The authors propose that ontogenetic changes in rhythm perception and action occur via the resonance and the attunement of coupled auditory-motor systems and rhythmic inputs across development. Therefore, the coupling between auditory and motor systems can account for the developmental VTS influences on auditory rhythm perception.

1.1. The current study

Moving from this literature, the present work aims at extending current evidence on the relationship between VTS and auditory rhythm experience, demonstrating that VTS perception not only has a role in shaping the auditory encoding of rhythm but it can also drive the encoding of linguistic stimuli, thus acting as a boost in early language acquisition. Consistently. caregivers spontaneously combine synchronous touches of the infant's body with word rhythm in infant-directed communications and this VTS-auditory redundancy is found to foster the acquisition of first words in infancy (Custode & Tamis-LeMonda, 2020; Lew-Williams, Ferguson, Abu-Zhaya, & Seidl, 2019; Tincoff, Seidl, Buckley, Wojcik, & Cristia, 2019). For instance, presenting tactile cues synchronous to words promotes word learning in 5-month-old infants (Abu-Zhaya, Seidl, & Cristia, 2017; Seidl & Cristià, 2008). Moreover, at 8 months of age, combined auditory and tactile stimulation was found to elicit increased event-related potential (ERPs) and electrophysiological (EEG) activity in the beta-band frequencies compared to auditory stimulation alone (Tanaka, Kanakogi, Kawasaki, & Myowa, 2018). During preschool years, toddlers' ability to synchronize their tapping rate with an external beat are found to predict phonological processing, auditory short-term memory, and rapid naming as well as neural encoding of the speech syllable envelope (Woodruff Carr, White-Schwoch, Tierney, Strait, & Kraus, 2014).

While there is consistent evidence that auditory rhythmic skills play a crucial role in language development (Fiveash, Bedoin, Gordon, & Tillmann, 2021; Ladányi, Persici, Fiveash, Tillmann, & Gordon, 2020; Lense, Ladányi, Rabinowitch, Trainor, & Gordon, 2021), early VTS rhythmic skills and the underlying processes mediating their interaction with the developing linguistic abilities in infancy remain unexplored to this point. Specifically, it is still to be clarified whether VTS inputs contribute to build internal representations for the temporal structure of perceived stimuli, be these musical or linguistic ones, thus being part of a set of general rhythmic abilities serving the processing of multiple signals across modalities and domains.

In the present study, we attempted to fill this gap by hypothesizing the regulatory effect of VTS abilities acting as a scaffold for the attentional system in guiding the encoding of complex rhythmic signals including rhythm and language. To test this hypothesis, we investigated the association between 45, 7 to 35-months infants' early VTS rhythmic abilities and their language processing. All children participated in two experiments: in the first one, the ability to discriminate the underlying meter of different musical rhythms under VTS sensory modality was assessed; in the second, we tested the low-level linguistic processing (i. e., phonological and prosodic perception) in a mispronunciation paradigm for novel object-label pairs presented in auditory modality. To explore the specific contribution of VTS rhythmic skills on children's phonological and prosodic processing, performances in the mispronunciation test were modeled adding the VTS rhythmic abilities as a predictor to the null model, then performing model comparison. Furthermore, to explore cross-sectional differences in rhythmic and linguistic abilities, age (in days) was added as a continuous predictor to the model. Discrimination abilities for rhythmic and linguistic stimuli were inferred from changes in pupil diameter over time (Hepach & Westermann, 2016; Mathôt, 2018) to contingent visual stimuli measured by a Tobii X-60 eye-tracker (Calignano, Dispaldro, Russo, & Valenza, 2021; Calignano, Valenza, Vespignani, Russo, & Sulpizio, 2021; Russo, Calignano, Dispaldro, & Valenza, 2021). Along with other traditional oculometer measures (e.g., looking times), pupillometry is considered a robust complementary measure for attention deployment in young children (Hepach & Westermann, 2016). Pupil dilation is indeed associated with greater attentional effort or cognitive load both in children and adults (Mathôt, 2018) and is a robust physiological index of attention from early on in infancy (Jackson & Sirois, 2009).

Recently, this investigative technique has been successfully applied to investigate the processing of tactile stimuli (van Hooijdonk et al., 2019), auditory musical rhythm (Bowling, Graf Ancochea, Hove, & Fitch, 2019; Marimon, Höhle, & Langus, 2022), and subtle lexical variations (Fritzsche & Höhle, 2015; Tamási, McKean, Gafos, Fritzsche, & Höhle, 2017) in adult and infant participants. Lastly, the continuous nature of pupillary data well suits with the application of strong statistical approaches including generalized (Baayen, Davidson, & Bates, 2008) and additive (Baayen, Vasishth, Kliegl, & Bates, 2017) mixed-effect models. Therefore, taking advantage of pupillometry and updated statistical approaches, this study explores the cross-domain (i.e., musical rhythm to language) and cross-sensory (i.e., VTS to auditory) effect of rhythm on language development.

2. Experiment 1 - developmental differences in rhythmic abilities

In this Experiment, rhythmic abilities in vestibular-tactile-somatosensory (VTS) modality were evaluated by means of a vibrotactile system for music perception and a metric discrimination paradigm. The aim of this Experiment was to test whether infants are able to discriminate between different meters only based on VTS input and how this ability changes across the first three years of life (from 7 to 35 months). Given the central role of VTS input in the perinatal period (Provasi et al., 2014; Ullal-Gupta et al., 2013), infants are expected to be able to tease apart VTS rhythms with novel versus familiar underlying meters. According to methodological paradigms of infant research (i.e., familiarization-test paradigm; Aslin, 2007), discrimination abilities can be inferred from different physiological (i.e., pupil dilation) measures indicating attention displayed toward novel vs familiar stimuli in the test phase. Discriminating among different meters in rhythm is essentially a cognitive process that implies the extraction of hierarchical patterns of strong and weak beats alternating in time (Fitch, 2013). This cognitive ability is crucial in processing rhythmic signals including music and speech (Kotz, Ravignani, & Fitch, 2018). Therefore, testing this cognitive ability across sensory modalities helps to shed light on the extent to which rhythmic abilities might be considered as a set of general, cognitive skills serving the processing of a vast range of signals (including music and language). Even though no study to our knowledge has investigated VTS rhythmic processing alone, evidence of efficient VTS rhythmic abilities are here expected, bringing new knowledge on the perceptual and cognitive abilities of young infants.

Lastly, given the considerable developmental changes occurring along with maturation in infancy, VTS rhythmic abilities are expected to change considering age as a continuous predictor. In the interim discussions of this first Experiment, the nature and direction of these changes will be discussed in terms of maturation and enculturation (i.e., the process by which infants acquire culture-specific knowledge about the structure of the music they are exposed to through everyday experiences; (Hannon & Trainor, 2007). We have complied with APA ethical standards in the treatment of our sample.

2.1. Methods

2.1.1. Participants

Forty-five infants with middle/high socioeconomic status were recruited from kindergartens (age and sex distribution for the total cohort of participants are displayed in Fig. 1). The inclusion criteria for all participants were (i) to be in good health, (ii) to have no known sensory or neurological disorders or (iii) family risks for language disorders, (iv) to be native Italian speakers, and (v) to perform at least 1 valid test trial per condition. Eight participants were excluded from the analysis because of non-compliance (e.g., fussiness or excessive irritability, $n = 5$) or because we could not collect a sufficient amount of data ($n = 3$). The final sample included 37 participants (17 females, sex assigned at birth; M age = 650 days, SD = 190). The sample size was fixed to at least 30 participants according to the indication that, in a regression analysis, increasing 5–10 observations per variable is likely to give at least an acceptable estimation of regression coefficients, standard errors, and confidence intervals (Hanley, 2016; Knofczynski & Mundfrom, 2008; Wolf, Harrington, Clark, & Miller, 2013).

2.1.2. Stimuli

Eight musical rhythms were originally synthesized as audio tracks with the library SoCal drum sound in GarageBand. Rhythms were all different except for their underlying meter, which could be a quadruple or a triple meter (Hannon & Johnson, 2005). The quadruple meter is a version of a double meter characterized by four primary beats (**1**234–**1**234–**1**234–**1**234). By contrast, the triple meter is characterized by three primary beats (**1**23−**1**23−**1**23−**1**23). Four different rhythms were generated for each meter, varying in the distribution of events and perceived accents across the units: for rhythms characterized by a quadruple meter, events and emerging accents occur more frequently every two or four units; whereas, for rhythms characterized by a triple meter, events and accents emerge more frequently every three units (Hannon & Johnson, 2005). All rhythms were played at 120 bpm

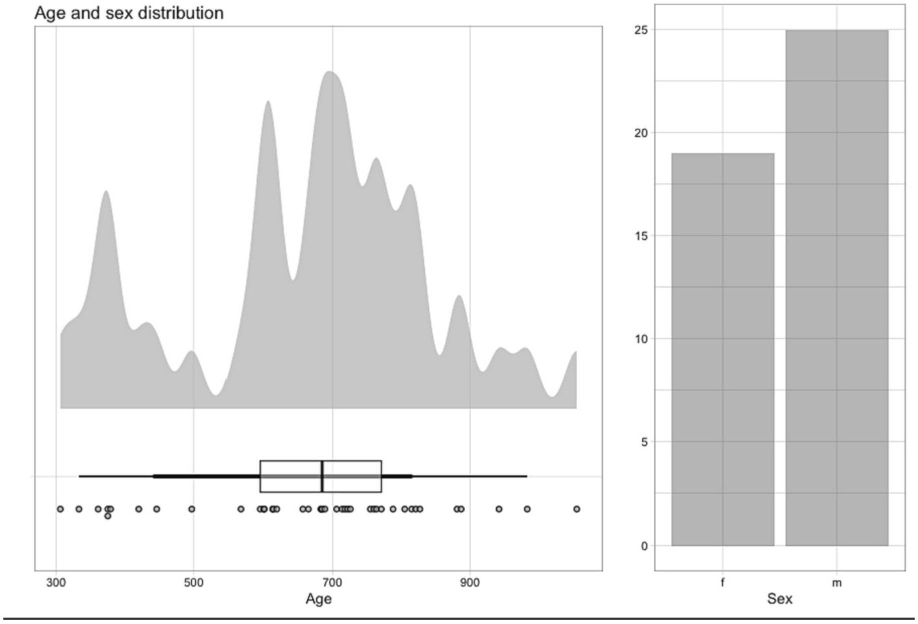


Fig. 1. Age and sex distribution. Density distribution of subjects per age in days (left panel) and discrete number of subjects per sex assigned at birth (right panel).

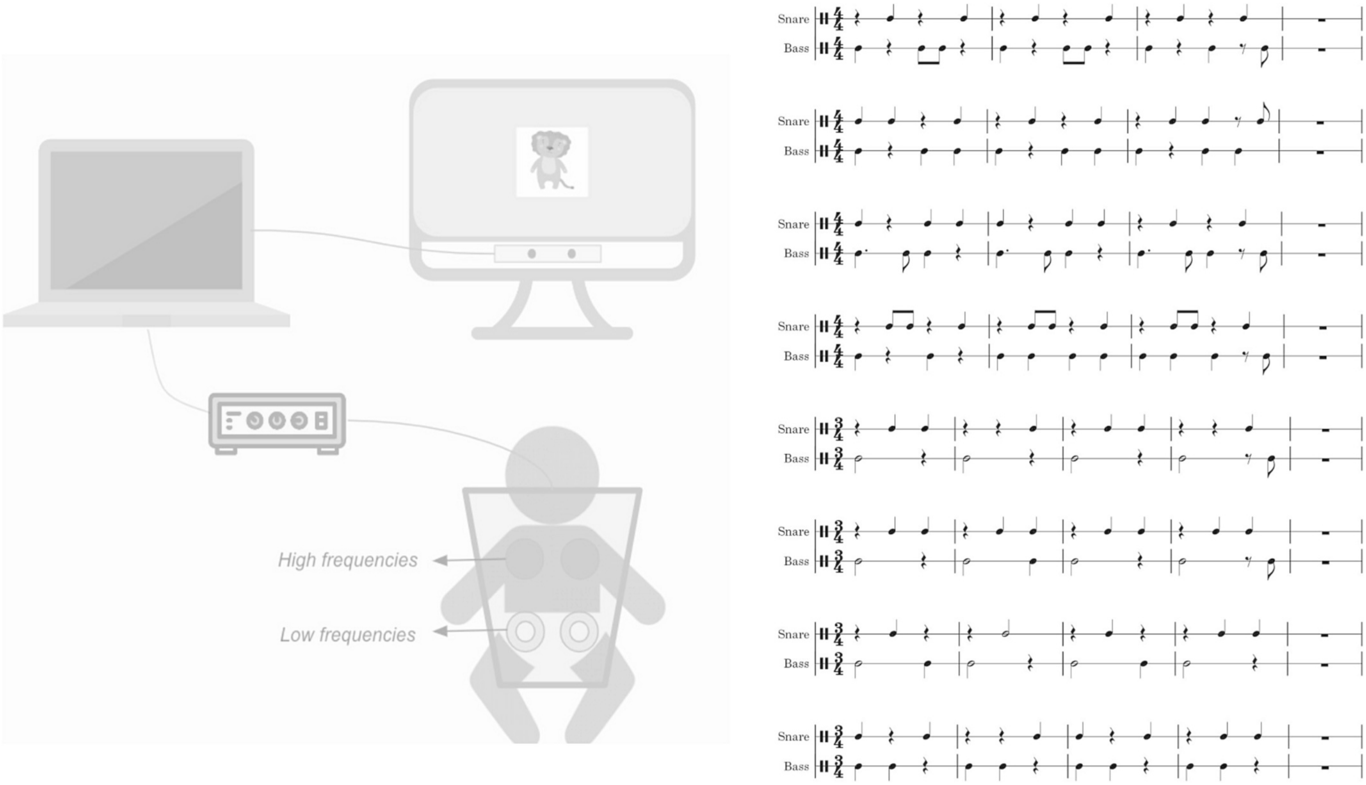


Fig. 2. The Infant vibrotactile system and the displayed rhythms. The experimental set-up included: the laptop running the experiment, the monitor displaying the visual stimuli, the eye-tracker collecting data, the amplifier, and the Infant Vibrotactile System transmitting VTS stimuli in low and high frequencies through actuator channels to the upper and low part of the infant's seat-back.

(Fig. 2). Each rhythm had a duration of 6 s and was displayed as a VTS stimulus by a vibrotactile system, described below, within a single trial together with a randomized image presented on a monitor.

Images consisted of ten, static and colored cartoon figures presented on a neutral background to attract the visual attention of the infants thus allowing the eye-tracker to collect the data.

The image areas corresponded precisely to the areas of interest (AOI), which measured 10 × 10 cm (9.554 deg) and remained visible throughout the trial. The visual stimuli were equated for luminance to reduce luminance-induced variability in the pupillometry measurements (Mathôt & Vilotijević, 2022).

2.1.3. *Apparatus*

The experiment was programmed and presented through Open Sesame software 3.1 (Mathôt, Schreij, & Theeuwes, 2012). Visual stimuli were displayed on a 27-in. monitor (Philips 300 × 300). Rhythms were presented in VTS modality through a custom-made, infant vibrotactile system: a music transduction device inspired by the Model Human Cochlea (MHC) from Karam, Nespoli, Russo, and Fels (2009); Karam, Russo, and Fels (2009) and specifically adapted to infant participants. Based on the sensory substitution field, music is displayed through actuators facilitating the direct translation of auditory information onto multiple discrete channels projected toward different portions of the back (Karam, Nespoli, et al., 2009; Karam, Russo, & Fels, 2009; Giordano, 2016; Fig. 2).

In this study, the Infant vibrotactile system was developed to adapt the MHC system to the infant body, proportionally reducing the number of channels transmitting the signal to four vibrotactile transducers embedded in a pillow and placed in the seat-back of an infant high-chair (Fig. 2). Original stereo sounds were processed by dividing kick timbre (bass) and snare timbre (high) sounds into two separate channels. Drum sounds were selected based on previous work with vibrotracks (Gunther & O'Modhrain, 2003; Holland, Bouwer, Dalgelish, & Hurtig, 2010) given the better performance obtained in processing tactile rhythms with strong rhythmical patterns (drums) transmitted via low frequencies (bass; Giordano, 2016). Each stimulus was processed with a parametric EQ filter to emphasize the vibration of the actuators and to remove resonances. The bass channel was processed with a Low-pass filter (500 Hz, Slope: 24 dB/Oct, Q = 0.75) and a Low-shelf filter (500 Hz, Gain: +24 dB, Q = 1.00). The high channel was processed with the same Low-pass filter, plus a High-pass filter (100hz, Slope: 24 dB/Oct, Q = 0.75) and a Peak filter (200 Hz, Gain: +24 dB, Q = 0.30). The left channel was displayed on the low part of the back and transmitted bass frequencies, whereas the right channel played high frequencies in snare drums to the upper part of the back. This frequency-based signal projection is inspired by the tonotopic organization of the human cochlea, where frequency-specific hair cell receptors are specialized in detecting individual band frequencies (Karam, Nespoli, et al., 2009; Karam, Russo, & Fels, 2009; Li et al., 2021). Each channel consisted of a set of two tactile transducers: the upper channel consisted of two voice coils (DAEX25VT-4 Vented 25 mm Exciter 20 W 4 Ohm), whereas the bottom channel consisted of two bass shakers (TT25–16 PUCK Tactile Transducer Mini Bass Shaker 4 Pk). All the tactile transducers received the signal from an amplifier connected through a jack cave to a laptop (Fig. 2). The remote, infrared eye-tracking camera (Tobii X2–60) placed directly below the screen 60 cm away from the participant recorded pupil dilation using bright-pupil technology at a sampling frequency of 60 Hz.

2.1.4. *Procedure*

This Experiment implemented a gaze-triggered, familiarization-test paradigm. At the beginning of each experimental section, a calibration was run. Participants were presented with small, cartoon images appearing on the screen at five locations (i.e., top-left, top-right, center, bottom-left, and bottom-right) together with a piece of cheerful music.

Once the calibration was successfully completed, the experiment started. The gaze-triggered procedure ensured that each experimental trial only began once the infant's gaze was detected as directed toward the AOI on the screen and was discontinued when infants looked away for >2 s. Between trials, an attentional getter was presented. During the familiarization phase, three different VTS rhythms with the same underlying meter were presented twice (6 trials) in random order. Each participant was randomly assigned to a familiarization group presenting double or triple meters (Hannon & Johnson, 2005). Therefore, half of the infants were assigned to a double meter familiarization group and the other half to a triple meter familiarization group. Immediately after the familiarization phase, the test phase started. Test trials consisted of two different rhythms presented twice (4 trials) and in random order, one of which was in double and the other one in triple meter. All infants were presented with the same test trials, regardless of the familiarization group. Infants were tested individually in a quiet room of their kindergarten; educators familiar with the infants assisted in the experimental session to comfort them in case of need. To ensure that the experimental setting elicited at its best a task-evoked response (Mathôt, 2018; *psychosensory pupil response*) rather than a mere luminance response to stimuli (Mathôt, 2018, *pupil light reflex*), the natural light of kindergarten rooms for testing was completely darkened with thick black curtains and then, semi-darkness constant luminance was obtained by placing the exactly same portable lamp 1 m behind the participant.

2.1.5. *Statistical analysis*

Based on inclusion criteria set prior to data collection (see *Participants section*), only data from participants who reached at least one valid test trial per condition (i.e., novel and familiar) were analyzed. Changes in pupil size (pupil dilation) under constant luminance were continuously collected and taken as a measure of cognitive processing during stimuli presentation (Beatty & Lucero-Wagoner, 2000; Calignano, Dispaldro, et al., 2021; Calignano, Valenza, et al., 2021; Mathôt & Van der Stigchel, 2015). Pre-processing steps were performed following the Hepach and Westermann (2016) procedure and the Mathôt and Vilotijević (2022) and Calignano, Girardi, and Altoè (2023)guidelines. Pupil data were then analyzed with generalized linear mixed models (GLMM) accounting for both random and fixed effects, specifying the distribution family. The gaussian distribution was selected to model the pupillary data (van Rij, Hendriks, van Rijn, Baayen, & Wood, 2019). A null model with a random intercept for participants was performed first. Age (in days) as a continuous predictor, test trial type (i.e., novel vs familiar), and familiarization group (i.e., double or ternary) as categorical predictors were subsequently inserted in the model. To find the best approximation to the true model, a hierarchical stepwise forward model comparison (Heinze, Wallisch, & Dunkler, 2018) was followed, based on AIC (Akaike Information Criterion) and AIC weight as indexes of the goodness of fit. Specifically, the model with the lowest AIC and the highest AIC weight was preferred (Wagenmakers & Farrell, 2004).

All data, materials, and code behind this analysis have been made publicly available at the *dataset*, *stimuli*, and *code* repositories of Experiment 1 and can be accessed at https://osf.io/rxwk8/?view_only. Data were analyzed using R, version 4.1.0 (Changes in R, 2018) and the packages *lme4* (Bates, Mächler, Bolker, & Walker, 2014) and *AICmodavg* (Mazerolle & Mazerolle, 2017).

2.2. *Results*

As shown in Table 1, data were best explained by Model 5. Specifically, Model 5 accounted for multiple interactions across age, trial type, and familiarization group effects on pupillary data (Fig. 3). The three-way interaction between predictors (i.e., trial type, age, and familiarization group) resulted as significant (b = 0.0001, SE = 0.00004, t = 4.007, $p < 0.001$).

Table 1
GLMM comparison for pupillary data in Experiment 1.

Models	Deviance	dAIC	AICw
M.0 Pupil diameter ~ (1|id)	16,580	346.69	0
M.1 Pupil diameter ~ age + (1|id)	16,598	346.29	0
M.2 Pupil diameter ~ age + trial type + (1|id)	16,595	366.03	0
M.3 Pupil diameter ~ age + trial type + familiarization group (1|id)	16,323	95.67	0
M.4 Pupil diameter ~ age * trial type + familiarization group (1|id)	16,323	98.38	0
M.5 Pupil diameter ~ age * trial type * familiarization group (1|id)	16,219	0.00	1

2.3. *Interim discussion*

This Experiment investigated rhythm discrimination under vestibular-tactile-somatosensory (VTS) modality in infancy. To this end, an Infant vibrotactile device for music transduction was designed based on previous works on sensory substitution systems in music technology (Branje & Fels, 2014; Giordano, 2016; Karam, Nespoli, et al., 2009; Karam, Russo, & Fels, 2009). This system was then combined with an eye-tracker to implement a gaze triggered, familiarization-test paradigm. Based on the statistics performed, the complex interaction between age as a continuous predictor and trial type and familiarization group as categorical ones, net of individual variability, best explained data distribution. Overall, the three-way interaction tells us that pupil diameter is influenced by the cumulative effects of trial type, age, and familiarization group. Specifically, the trial type effect's interaction with age is inferred by the direction of pupil increases in the test phase: in fact, for both familiarization groups (double and triple meter), younger infants display a larger pupil diameter and thus attentional investment when processing familiar stimuli (i.e., double for the double familiarization group and triple for the triple familiarization group) while older infants display a larger pupil diameter and thus greater attention toward novel stimuli (i.e., triple for the double familiarization group and double for the triple familiarization group).

Furthermore, the familiarization group itself plays a role in the three-way interaction too. In fact, Fig. 3 shows that, when presented with the exact same stimuli in the test phase (all infants were presented with the same test double and triple stimuli), older infants displayed a reduced response when familiarized with the triple meter compared to those familiarized with the double meter. This familiarization group effect can be explained by considering that, for the infants from the double familiarization group, the novel class of stimuli consisted of rhythms with a triple meter. According to the musical enculturation processes (Trainor & Hannon, 2013), following the principles of perceptual narrowing, processing abilities in young infants are expected to be broader and to tune toward culture-specific signals along development; at the same time, processing abilities for stimuli uncommon to native environments are expected to decrease (Lewkowicz, 2014; Maurer & Werker, 2014). Therefore, given that all infants participating in this Study were born and grew in a Western country, they were primarily exposed to rhythms characterized by a double meter (Trainor & Hannon, 2013).

Therefore, the interpretation here is that infants reacted more to the introduction of novel stimuli in the test phase when these stimuli were characterized by a meter that was unfamiliar to them not only for the task-related manipulation but also for cultural-background factors. In fact, older infants from the double familiarization showed more detailed discrimination abilities in the test phase compared to the triple familiarization group, thus being more sensitive to the novel stimulus when it consisted of a triple meter. Lastly, this difference emerged for older infants only, whereas younger subjects reacted mostly similarly to the test stimuli (by investing more attentional resources toward familiar ones) in both groups. This further stands in favor of an enculturation effect since younger infants are expected to show broader preference while cultural-specific discrimination abilities are expected to appear later in development due to the increased experience with the infant environment (Lewkowicz, 2014; Maurer & Werker, 2014; Trainor & Hannon, 2013).

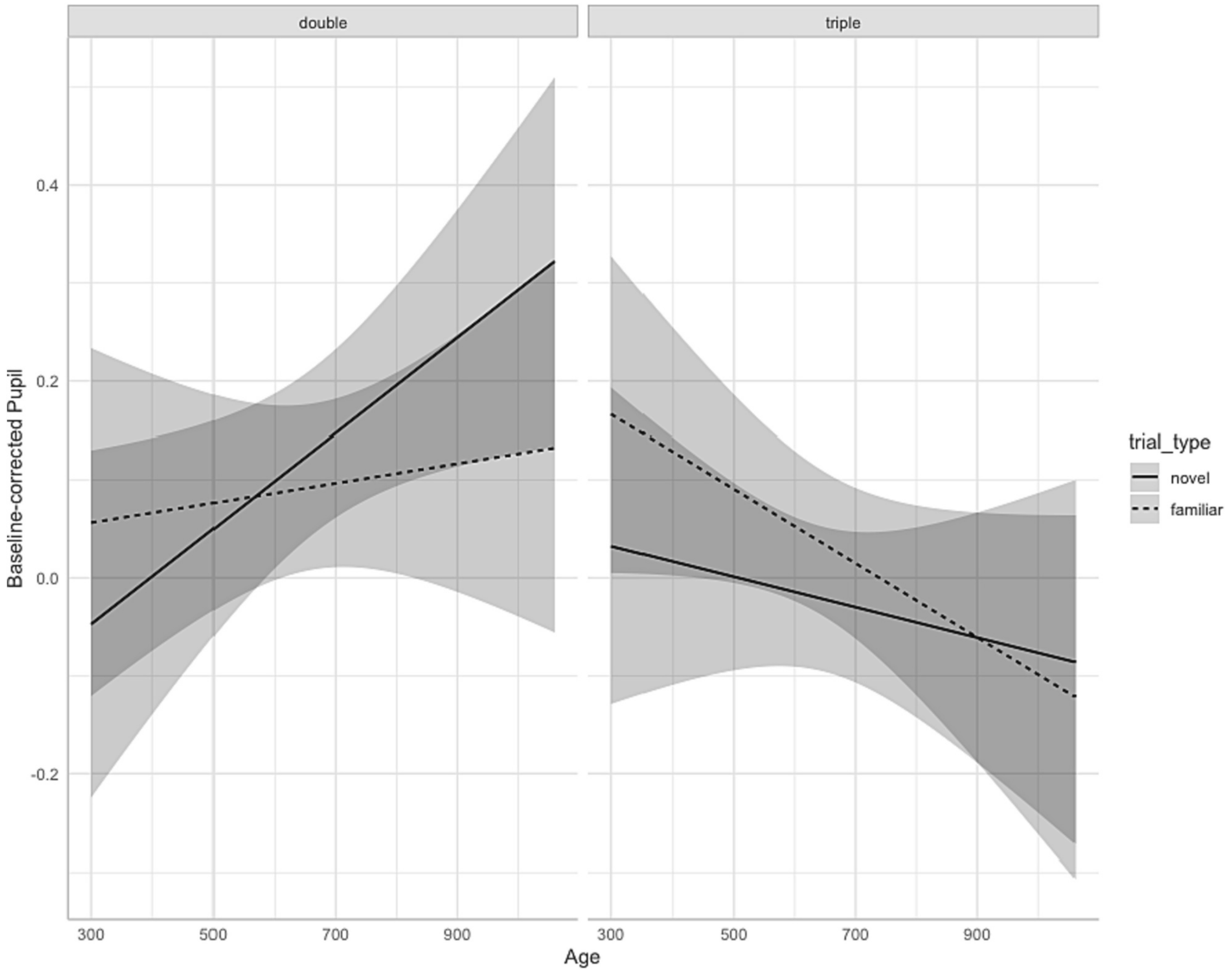


Fig. 3. Effect plots of the best model (M5) predicting pupil size in the test phase. Predicted changes in pupil diameter for novel (continue line) vs familiar (dotted line) trial type, for infants familiarized with double (left) and triple (right) meters, along with age (in days).

Therefore, results from this Experiment suggest that age-related and cultural-related changes in VTS rhythmic abilities might occur across infancy.

Starting from these results, the next step of this study was to investigate whether the VTS rhythmic abilities were related to early phonological and prosodic skills. Thus, the same cohort of infants participating in Experiment 1 was tested with a linguistic task in Experiment 2; then, an exploratory analysis investigated the relationship between individual performances across the two experiments.

3. Experiment 2 – Association between VTS rhythmic abilities and linguistic skills

In this Experiment, the emerging linguistic abilities of infants from 7 to 35 months were tested through a discrimination task for novel object-label pairs. Specifically, prosodic and phonological abilities were measured by investigating how infants responded to mispronunciations, occurring at stress or phonological level, in recently familiarized labels. Discrimination abilities were inferred by physiological (i.e., pupillometry) measures collected through the eye-tracker for screen images. Infants were presented with a visual stimulus (i.e., a cartoon character) and a corresponding auditory label (i.e., a disyllabic unit like /bosa/).

Labels were presented in their original form (i.e., familiar; F condition) or altered in their prosodic (i.e., novel stress; NS condition) or phonological (i.e., novel phonological; NP condition) structure. Changes occurring in pupil diameter were expected across conditions and taken as measures of prosodic and phonological discrimination abilities. Specifically, the rationale for this Experiment comes from the literature on early language showing that lexical encoding is sensitive to feature alterations (Delle Luche, Durrant, Poltrock, & Floccia, 2015; Swingley & Aslin, 2000; Yoshida, Fennell, Swingley, & Werker, 2009). Indeed, early word representations seem to be characterized by a high degree of phonological specificity (Mani, Durrant, & Floccia, 2012; Swingley & Aslin, 2000). Consistently, infants show fine phonological abilities since the first stages of language acquisition (Curtin & Archer, 2015; Velleman & Vihman, 2007) and even finer later in development, as their experience with the native language grows (Kuhl et al., 2006). Similarly, basic prosodic skills are already in place at birth (Abboub, Nazzi, & Gervain, 2016; Gervain, 2018) while fine discrimination abilities emerge by experience with native prosodic patterns later in development (Johnson & Jusczyk, 2001; Jusczyk, Houston, & Newsome, 1999).

Crucially, most of the studies investigating word representation in infancy rely on target looks in looking while listening tasks (Delle Luche et al., 2015; Mani et al., 2012; White & Morgan, 2008). In this paradigm, an auditory label is presented while two images (i.e., a target and a distractor) appear on the screen. Looks toward the target are expected to decrease in case of mispronunciations. Even though this paradigm offered useful insights into early word acquisition, it also presents some limitations. For instance, the use of a behavioral measure (i.e., looking times) can be sensitive to variables other than the ones of interest (e.g., motivation, familiarity, maturational level; DePaolis, Keren-Portnoy, & Vihman, 2016; Oakes, 2017). Furthermore, the presence of two simultaneous objects might affect the results due to familiarity and/or spontaneous preference (White & Morgan, 2008). Therefore, the present Experiment investigated object-label encoding by analyzing variations in pupil diameter while presenting an auditory label and a unique visual referent. As a direct measure of cognitive load, changes in pupil diameter can indeed accurately reflect the effort of encoding mispronounced labels better than looking time or preferential looking tasks (Hepach & Westermann, 2016; Sirois & Brisson, 2014).

Therefore, pupillometry might provide a detailed index of early object-label pairs representation in infancy, and in fact, prior studies have demonstrated that pupil diameter is sensitive to mispronunciations, with an increase in deviant labels compared to corrected or unrelated ones (Fritzsche & Höhle, 2015). Even with infants (35 months of age), pupil responses detect with relative precision sensitivity to different degrees of mispronunciations (Tamási et al., 2017; Tamási, McKean, Gafos, & Höhle, 2019). In the present Experiment, variations in pupil diameter to a visual referent while listening to corrected or altered (for phonological or prosodic features) labels are collected to replicate previous findings on pupil sensitivity to mispronunciations; furthermore, previous findings are extended by testing infants from 7 thus clarifying the developmental trajectory leading to the fine word representations reported in previous studies (at 35 months; Tamási et al., 2019, 2017). We have complied with APA ethical standards in the treatment of our sample.

3.1. *Methods*

3.1.1. *Participants*

The same cohort of forty-five infants participating in Experiment 1 also participated in this second Experiment. Five participants were excluded from the analysis because of non-compliance ($n = 4$) or because they did not meet the inclusion criteria established prior to data collection of at least 1 valid test trial per condition ($n = 1$). The final sample included 40 participants (18 females, sex assigned at birth; M age = 694 days, SD = 181). Lastly, a sample of thirty-two participants (15 females, sex assigned at birth; M age = 666 days, SD = 189) successfully completed Experiment 1 and 2 and was therefore included in the final explorative analysis linking the two experiments. The sample size was fixed to at least 30 participants according to the indication that, in a regression analysis, increasing 5–10 observations per variable is likely to give at least an acceptable estimation of regression coefficients, standard errors, and confidence intervals (Hanley, 2016; Knofczynski & Mundfrom, 2008; Wolf et al., 2013).

3.1.2. *Stimuli*

Three different labels were selected from the Novel Object and Unusual Name database (Horst & Hout, 2016). The three items selected for this Study were three disyllabic pseudo-words with a CVCV sequence (Table 2). The three pseudo-words were presented in their original form (familiar; F condition) or altered for phonological (novel phonological; NP condition) or prosodic (novel stress; NS condition) features. This resulted in 9 items (Table 2). Each label was paired with a unique visual reference. Images consisted of three cartoon characters presented on a neutral background. The image areas corresponded precisely to the areas of interest of the eye-tracker (AOI), which measured 10 × 10 cm (9.554 deg) and remained visible throughout the entire trial. Visual stimuli were selected from an online free database (freepik.com) and were equated to reduce luminance-induced variability in the pupillometry measurements (Hepach & Westermann, 2016; Mathôt & Vilotijević, 2022).

Table 2
Linguistic stimuli from Experiment 2.

Label	Familiar (F)	Novel Phonological (NP)	Novel Stress (NS)
Bosa	Bosa	Mosa	Bosa
Loche	Loche	Lome	Loche
Nare	Nare	Pare	Nare

3.1.3. *Apparatus*

The experiment was programmed and presented through the Open Sesame software 3.1 (Mathôt et al., 2012) running on a computer laptop (Acer travel mate 5772 g). Visual stimuli were displayed on a 27-in. monitor (Philips 300 × 300). Labels were auditory presented through two loudspeakers placed on both sides of the monitor. The remote, infrared eye-tracking camera (Tobii X2–60) placed directly below the screen, 60 cm away from the participant, recorded the eye movements using bright-pupil technology at a sampling frequency of 60 Hz.

3.1.4. *Procedure*

A gaze-triggered, familiarization-test paradigm was implemented in this Experiment. Variations in pupil dilation were continuously recorded during the experiment, eliciting the trial presentation. At the beginning of each experimental section, a calibration was run (see Experiment 1). Each trial consisted of a familiarization and a test phase. Before the label presentation, a silent time window of 500 ms allowed the eye-tracker to collect baseline data on pupil diameter prior to stimuli presentation. Once the gaze of participants was detected as inside the AOI and registered continuously for 500 ms, the familiarization phase began. In the familiarization phase, a label was repeated two times. With an inter-stimulus interval of 1 s, the test phase followed and consisted of a label presented two times: the first time in its original form and the second time in the familiar, novel phonological or novel stress condition. Each of the three labels was paired with one of the three visual referents which remained visible for the entire trial duration (5 s). Therefore, each trial consisted of one label presented through the familiarization and test phase. Each label-object pair was presented twice in the familiar condition, once in the novel phonological, and once in the novel stress conditions. Therefore, the experimental section consisted of 12 trials: 6 familiar and 6 mispronounced, all presented in random order. This design allows for testing multiple labels while keeping intact the familiarization-test structure (Calignano, Dispaldro, et al., 2021; Calignano, Valenza, et al., 2021; Russo et al., 2021). Moreover, this design allows for appropriate pupil data collection since every test event is time-locked within a specific time window, facilitating time course analysis. During the experimental section, infants were seated on a highchair positioned 60 cm away from the monitor and were tested individually in a quiet room of their kindergarten. To ensure that the experimental setting best elicited a task-evoked response (Mathôt, 2018, *psychosensory pupil response*) rather than a mere luminance response to stimuli (Mathôt, 2018, *pupil light reflex*), the natural light of kindergarten rooms for testing was completely darkened with thick black curtains and then, semi-darkness constant luminance was obtained by placing the exactly same portable lamp, positioned 1 m behind the participant.

3.1.5. *Statistical analysis*

Only data from participants who reached at least 1 valid test trial per condition were analyzed. Changes in pupil size under constant luminance were taken as a measure of cognitive processing during stimuli presentation (Calignano, Dispaldro, et al., 2021; Calignano, Valenza, et al., 2021; Mathôt & Van der Stigchel, 2015). Pre-processing steps were performed following the Hepach and Westermann (2016) procedure and the Mathôt and Vilotijević (2022) and Calignano et al. (2023) guidelines.

To explore the changes occurring in pupil dilation across time, baseline-corrected pupil data were modeled with Generalized Additive Mixed-effect models (GAMM). GAM models address non-linear relationships between pupil size variation and time. Data were modeled by using a maximal random structure: the trial was considered as a minimum statistical unit and set a minimum of 20 knots as the maximum number of turning points to be used during the smoothing process (Baayen et al., 2017). To explore whether the experimental manipulations statistically influenced pupil size across time, the model's estimates of the differences between conditions were evaluated and visually inspected (van Rij et al., 2019). Specifically, the difference curve was plotted based on the model predictions, thus specifying the time windows where differences were expected to be significant as well as the estimated effect size. First, data were modeled over time (in ms), by setting conditions (i.e., F, NP, NS) as a categorical predictor of pupil variations, while random effects were included as random smooths for participants, trials, and age (Model 1). Secondly, data within the test time-window were modeled across the age continuum (in days), by setting conditions (i.e., F, NP, NS) as a categorical predictor. In this second model, random effects were therefore included for participants and trials only (Model 2).

$$\text{M1}: \textit{Pupil diameter} \sim \textit{trialtype} + s(\textit{time}, \textit{by trialtype}, k = 20) + s(\textit{time}, \textit{age}, \textit{bs} = \text{`}\textit{fs}\text{'}, m = 1) + s(\textit{time}, \textit{id}, \textit{bs} = \text{`}\textit{fs}\text{'}, m = 1) + s(\textit{time}, \textit{trial}, \textit{bs} = \text{`}\textit{fs}\text{'}, m = 1)$$

All data, materials, and code behind this analysis have been made publicly available at the *dataset*, *stimuli*, and *code* repositories of Experiment 2 and can be accessed at https://osf.io/rxwk8/?view_only. Data were analyzed using R, version 4.1.0 (R Core Team, 2020) and the packages *lme4* (Bates et al., 2014; Mazerolle & Mazerolle, 2017), *mgcv* (Wood, 2007), *itsadug* (van Rij, Wieling, Baayen, van Rijn, & van Rij, 2017), and *MuMIn* (Barton & Barton, 2015).

3.2. Results

Regarding the linguistic task, the model structure resulted as follows:

The estimated differences between familiar and novel trials across time are plotted with pointwise 95% confidence intervals in Fig. 4. Difference among the curves is significantly different from zero for both the novel phonological (b = −0.0271, SE = 0.0024, $t = -10.923$, $p < 0.001$) and the novel stress (b = −0.0328, SE = 0.0024, $t = -13.344$, $p < 0.001$) conditions. Specifically, the time window in which increases in pupil diameter toward novel labels were registered corresponds to the test phase of each trial (4 to 5 s). Lastly, the estimated difference is higher for novel phonological labels than novel stress ones (Est. difference NP: 0.15, Est. difference NS: 0.10).

Data from the test phase were then modeled by adding condition (i.e., trial type) per age and time as continuous predictors with participants and trials as random effects (M2). The resulting structure is described as follows and results are plotted in Fig. 4, right. A significant increase in pupil diameter along age is reported for novel phonological labels (b = −0.0873, SE = 0.0250, $t = -3.487$, $p < 0.001$) but not for novel stress labels (b = 0.0051, SE = 0.0243, $t = 0.209$, $p = 0.834$).

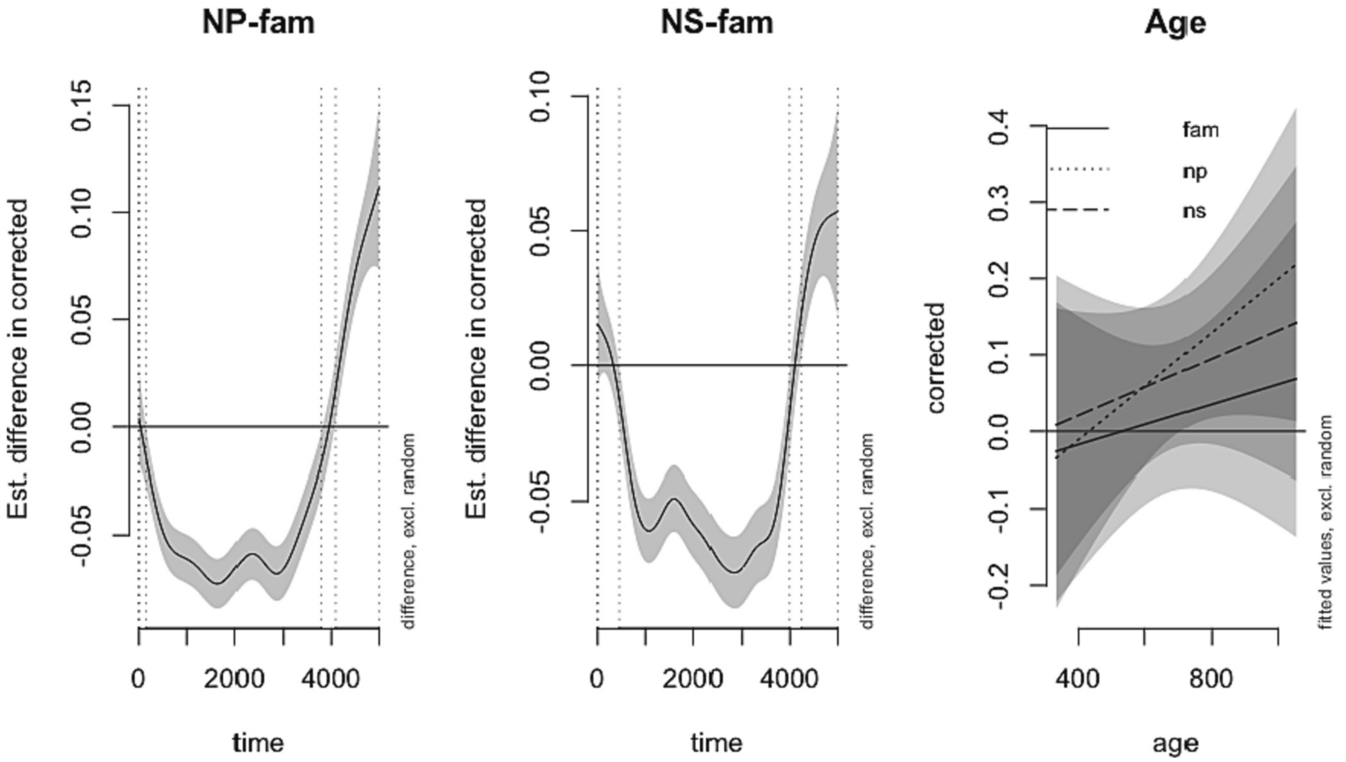


Fig. 4. Estimated differences with pointwise 95% confidence intervals on time (M1, left panels) and estimated effect on age (M2, right panel). Left panels show the estimated difference between variations in pupil diameter in novel phonological (NP) and novel stress (NS) labels compared to familiar labels (fam) across time, as predicted in Model 1. Right panel shows the estimated effect of changes in pupil diameter across conditions (NP, NS, fam) in the test phase time window (4 s to 5 s) across age, as predicted by Model 2.

$$M2 : Pupil\ diameter \sim triltype^{*}age + s(time, by = trialtype, k = 20) + s(time, id, bs=`fs`, m = 1) + s(time, trial, bs=`fs`, m = 1).$$

In summary, results from the linguistic task indicate that variations in pupil diameters signal infants discriminating against deviant phonological and prosodic features in recently familiarized label-object pairs; furthermore, phonological features seem to have a more detailed representation compared to prosodic ones and this trend seems to increase with age.

3.3. Interim discussion

Experiment 2 investigated the emerging linguistic abilities of 7- to 35-months-old infants. Specifically, phonological and prosodic abilities were assessed through a familiarization-test paradigm for object-label pairs. Results show a significant increase in the pupil diameter occurring in the test phase. Moreover, an increasing difference between familiar and novel phonological items was predicted with age. Results show also that pupil diameter increases for tests over familiar conditions in general and for changes occurring on phonological features in particular.

Overall, these results suggest a fine lexical representation across ages. Specifically, object-label pairs seem to be encoded with detailed phonological and prosodic features since infancy. This result is in line with previous findings reporting mispronunciation effects on word recognition in infants (Delle Luche et al., 2015; Swingley & Aslin, 2000). However, most of these studies consisted of a word recognition task (Delle Luche et al., 2015; Swingley & Aslin, 2000; White & Morgan, 2008). By contrast, in the present Study a new paradigm was developed allowing us to better specify the effect of segmental and suprasegmental features on object-label encoding, while minimizing the impacts of other incidental factors. Indeed, the presence of a distractor, and the use of real words as behavioral measures represent potential sources of variability that can be reduced by increasing the methodological strength (Klingner, 2010). Specifically, in the present Study infants were presented with novel object-label; furthermore, labels were pseudo-words and mispronunciations occurred right after the familiarization phase, minimizing possible external interference of time lag. Indeed, every trial consisted of a familiarization and a test phase (Calignano, Dispaldro, et al., 2021; Calignano, Valenza, et al., 2021; Russo et al., 2021). This allowed us to test up to three novel object-label pairs, improving the generalizability of our observations. Lastly, the dependent variable of this study was the change in pupil diameter occurring in correspondence to the test phase. This led to a narrowed response while applying a well-established physiological index of cognitive effort (Karatekin, 2007).

In sum, findings from the present Study can be interpreted as a sign of the infant shift from basic to fine linguistic abilities in lexical encoding. Specifically, prosodic cues are known to bootstrap the first steps of language acquisition in young infants; while fine phonological abilities are known to develop with age, with perceptual narrowing mechanisms aiding the processing of native phonological features as a function of experience (Gervain, 2018; Johnson & Jusczyk, 2001; Kuhl, 2004). Crucially, previous studies investigated the effect of mispronunciations on word recognition; by contrast, pupillometry is known to better reflect cognitive effort. This means that a higher cognitive effort was required to process altered labels. This result is more indicative of the mismatch between the familiarized object-label pair and the mispronounced one rather than word recognition.

Based on the results of this study, it can thus be inferred that labels varying in prosodic features require less cognitive effort to be processed than phonological ones and that phonological representations grow in specificity with age requiring more cognitive effort to be processed when a mismatch is detected.

This would be an adaptive strategy for infants since segmental information changes at a fast rate in speech and is highly informative in terms of conveyed meaning (e.g., /bat/ and /cat/; Fernald, Swingley, & Pinto, 2001). By contrast, suprasegmental features like prosodic characteristics are known to vary between different speakers and emotional intonation (Selkirk, 1995). Accordingly, while young infants strongly rely on prosodic cues when moving their first steps into language acquisition, the possibility to develop a greater degree of flexibility in prosodic processing could characterize later steps of lexical representations (Seidl & Cristià, 2008). This interpretation might explain: i) the larger effect in the novel phonological condition, and ii) the growth of this effect along with age. At the same time, even with a less robust effect and with no significant growth along with age, infants show an increase in pupil diameter in response to novel stress labels too. Once the differences between the two manipulations (i.e., phonological and prosodic) have been explained in terms of language-specific cue weighting, it can be clarified that changes in prosody remain informative even if the infant cognitive system treats with more flexibility those changes containing the impact on word recognition.

4. Relating VTS rhythm and language processing across age

The last step of this Study was to explore the relationships between linguistic skills (Experiment 2) and VTS rhythmic abilities (Experiment 1). According to the literature showing a core link between rhythm and language in infancy (Fiveash et al., 2021; Ladányi et al., 2020), a positive relationship between the two tests was expected. To this aim, individual scores in the rhythmic task were first computed from the ratio of averaged pupil diameter in the novel over total trials in the test phase [novel / (novel + familiar)]. Variations in pupil diameter in the linguistic task were then modeled by means of GLMM following a hierarchical stepwise model comparison.

Fixed effects were gradually added to the null model (with random intercept per participant), following the postulated hypothesis, and included i) scores in the rhythmic task and ii) age as continuous predictors, and iii) condition (i.e., trial type; fam, NP, NS) as a categorical predictor. Data were best explained by Model 4 (Table 3) accounting for the interaction between rhythmic scores and condition, and the additive effect of age. A significant interaction between the rhythmic scores and the pupil diameter in the novel phonological (b = 0.2382, SE = 0.0233, t = 10.208, $p < 0.001$) and novel stress (b = 0.1897, SE = 0.0217, t = 8.724, $p < 0.001$) conditions was predicted (Fig. 5).

Specifically, increases in pupil diameter toward deviant phonological and prosodic features were predicted by higher rhythmic scores. It is here important to specify that rhythmic scores were computed as the

Table 3
GLMM comparison for rhythmic performances on linguistic pupillary data.

Models	Deviance	dAIC	AICw
M.0 Pupil diameter ~ (1|id)	1077	6.75	0.03
M.1 Pupil diameter ~ rhythm + (1|id)	1161	92.05	0
M.2 Pupil diameter ~ rhythm + trial type + (1|id)	1161	96.12	0
M.3 Pupil diameter ~ rhythm + trial type + age (1|id)	1175	112.52	0
M.4 Pupil diameter ~ rhythm * trial type + age (1|id)	1058	0.00	0.97
M.5 Pupil diameter ~ rhythm * trial type * age (1|id)	1078	29.66	1

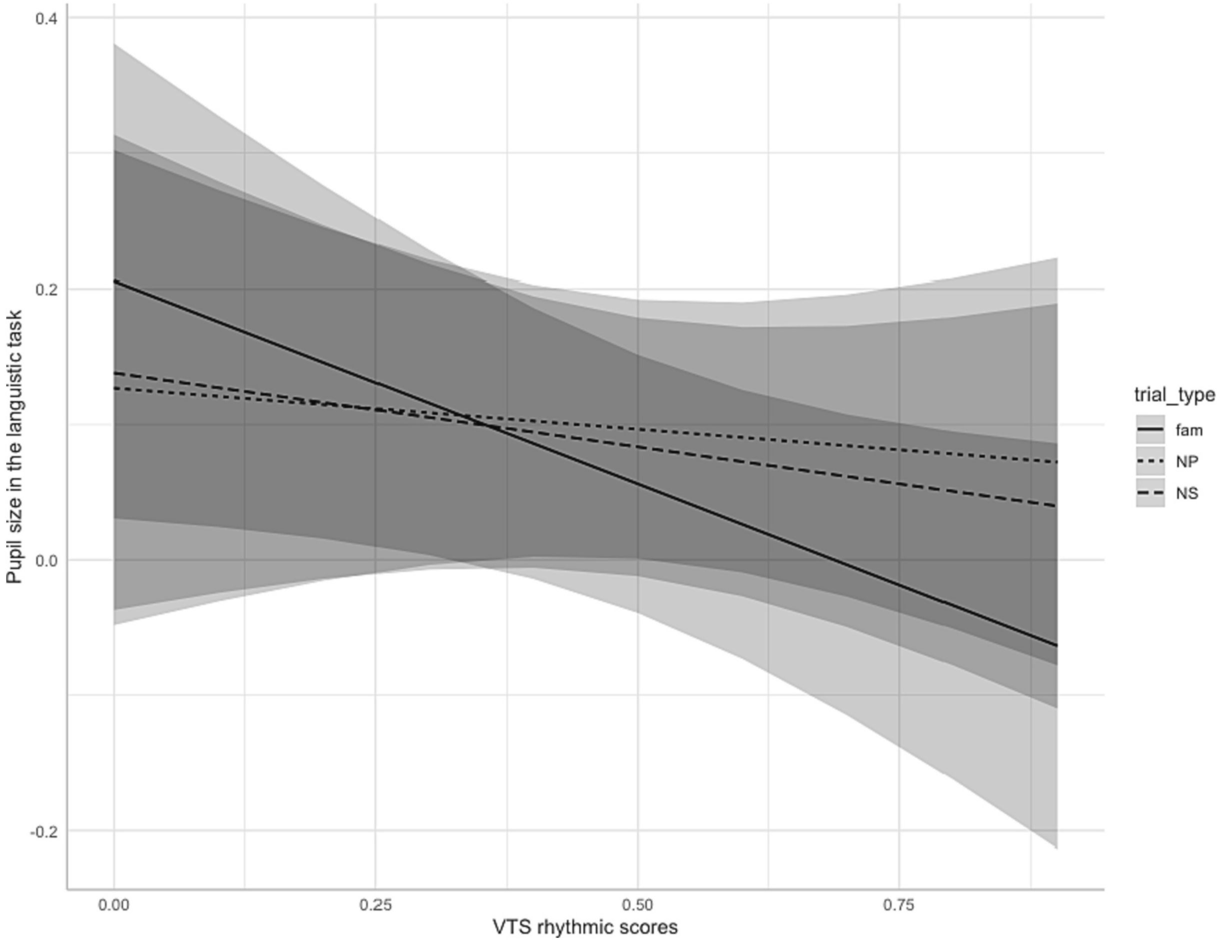


Fig. 5. Effect plots of the best model (M4). Predicted relationship between variations in pupil size (y-axis) across conditions (trial type: familiar, novel phonological, and novel stress) in the linguistic task and scores in the rhythmic task (x-axis).

ratio of the pupil increases toward novel test trials. Therefore, values above 0.5 indicate an increase in pupil diameter toward novel stimuli whereas values below 0.5 indicate an increase toward familiar ones. Therefore, infants showing an increase in pupil diameter toward familiar stimuli in the language task also did the same in the rhythmic task, and vice versa.

Since variations in pupil diameter were shown to be modulated by age in both tasks (Figs. 3 and 4), an exploratory GAM model was performed to investigate how the relationship between linguistic and rhythmic abilities change across age. Percentages of averaged increase in pupil diameter for novel stimuli were computed for the two deviant conditions in the linguistic task [novel / (novel + familiar)]; then, individual scores across tasks (i.e., rhythmic, phonological, and prosodic) were taken as dependent variables and the interaction between tasks and age was modeled as a continuous predictor (Fig. 6).

$$\mathrm{M3}: Scores \sim task + s(age, task, k = 20) + s(id, bs = `fs', m = 1).$$

Scores are predicted to grow with age, with no significant difference between tasks (NP: b = 0.0933, SE = 0.0918, $t = 1.016$, $p = 0.315$; NS: b = 0.1406, SE = 0.0917, $t = 1.533$, $p = 0.132$). Paralleling the findings from the linguistic task, phonological abilities show a rapid growth compared to prosodic and rhythmic ones, even if differences are not significant.

5. General discussion

The present work aimed at exploring the developmental association between VTS rhythmic abilities and language processing in infancy. To this aim, discrimination abilities for VTS rhythms (Experiment 1) and language (Experiment 2) were tested in a sample of 45 infants from 7 to 35 months. Specifically, 37 infants from the original sample completed Experiment 1, which showed that i) infants are able to discriminate between different rhythms based on their underlying meters perceived via VTS sensory modalities and that ii) increased rhythm discrimination abilities are found for infants familiarized with double meters, in line with previous findings on musical enculturation processes in auditory development (Trainor & Hannon, 2013). In Experiment 2, infant discrimination abilities for phonological and prosodic features of linguistic stimuli were successfully tested in 40 participants from the original cohort via auditory sensory modality, showing that: i) infants are sensitive to mispronunciations at the prosodic and, especially, phonological levels of novel object-label pairs, and that ii) sensitiveness to phonological alterations increases with age, in line with language acquisition and phonological specialization (Seidl & Cristià, 2008). Apart from being informative per se, the two experiments together shed light on the role that VTS rhythmic abilities can have in terms of predicting factors of intants' linguistic performances. In fact, the comparison approach to GLM models, performed on the data from 32 infants who completed both Experiments 1 and 2, allowed for overall identifying the best model as the one explaining the changes in pupil diameter during the linguistic task as the interaction of experimental manipulation (i.e., test conditions) with individual scores in the rhythmic task, considering the effect of age. It can be recalled that proportional pupil changes toward test stimuli are comparable between the rhythmic and the linguistic task, varying with age. This result tells us that: i) there might be a link between performances in the rhythmic and linguistic

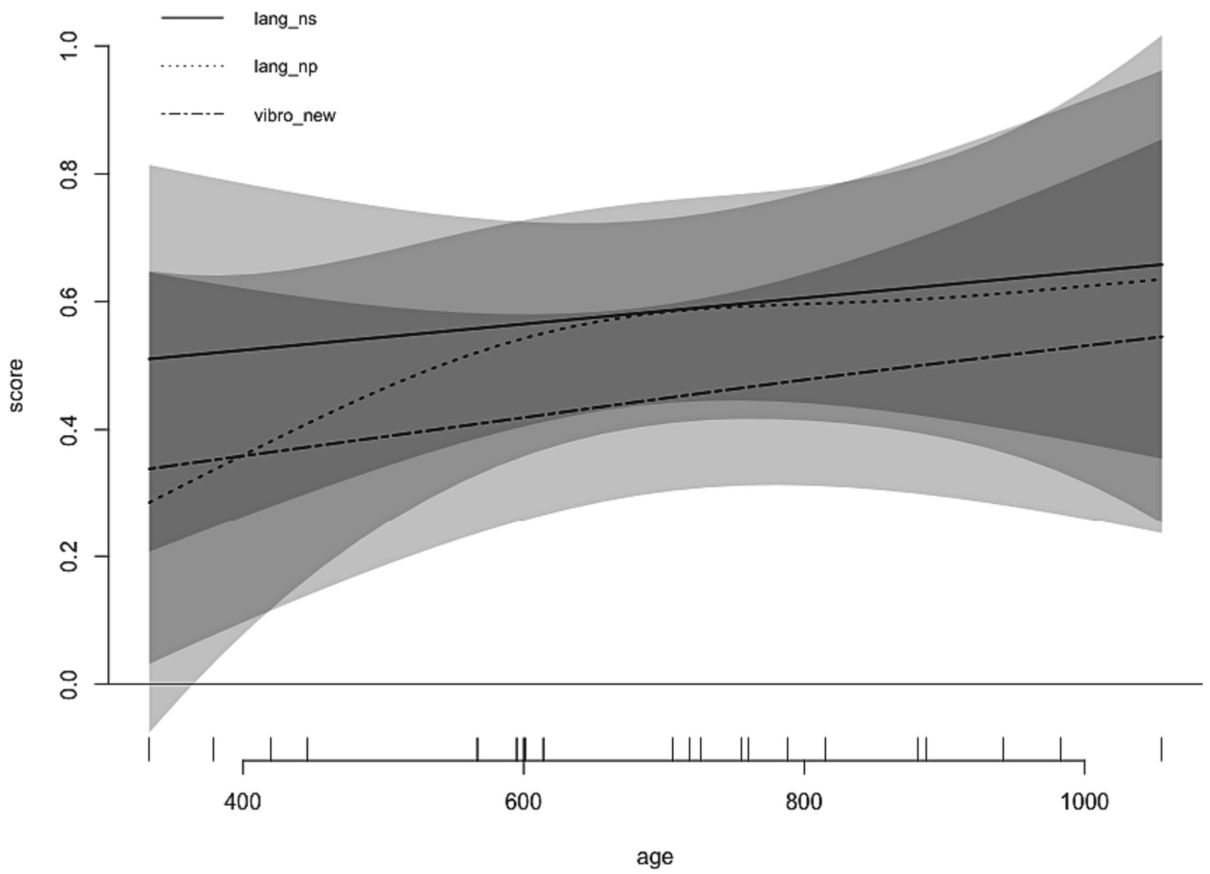


Fig. 6. Estimated effect (Model 3) task scores across age. Performance scores (y-axis) per task (i.e., rhythm as *vibro new*; phonological as *lang np*; and prosodic as *lang ns*) across age (x-axis).

tasks; ii) this link might have a composite pattern rather than an absolute value sensitive to developmental changes.

In sum, results suggest that the way in which infants react to unpredicted changes occurring in the underlying meter of musical rhythms is related to the way in which they respond to unpredicted changes in the phonological and prosodic features of linguistic stimuli.

Furthermore, stimuli in the rhythmic task were perceived as VTS stimuli while labels in the linguistic task were presented only as auditory. This suggests that an overlapping set of basic rhythmic abilities might serve the processing of a vast range of signals across domains (i.e., music to language) but also across sensory modalities (i.e., touch to hearing). This result is particularly informative within the growing literature on the link between basic rhythmic skills in language acquisition (e.g., Fiveash et al., 2021) and cross-sensory perception (e.g., Karam, Nespoli, et al., 2009; Karam, Russo, & Fels, 2009). Specifically, the first framework and theories on rhythm in language development called for a detailed investigation into the link between cross-domain and cross-sensory abilities underlying rhythm processing in musical and linguistic domains. Accordingly, the present Study brings significant insights into the extent to which rhythm processing suggests a general cognitive ability serving the processing of different temporal signals across modalities.

This evidence might contribute to both experimental and applied research. In fact, exploring how infants process rhythm can shed light on the perceptual abilities underlying complex cognitive processes as well as the extent to which these processes might be considered as general abilities serving multiple cognitive functions. Accordingly, basic mechanisms underlying rhythm processing are hypothesized to be crucial to language development (Fiveash et al., 2021). Furthermore, signs of rhythmic difficulties were found across neurodevelopmental disorders (Lense et al., 2021) pointing out atypical rhythm as a possible risk factor for neurodevelopmental disorders (Ladányi et al., 2020). Finally, given the crucial role of rhythm in cognitive development, exploring rhythmic abilities across different sensory modalities can be further informative for designing screening and intervention programs for early sensory deprived infants as well as children with sensory-related difficulties (Ghanizadeh, 2011; Tomchek, Huebner, & Dunn, 2014; Russo and Valenza, 2021; e.g., ASD and ADHD). Therefore, even if further research is needed to replicate and extend the present findings, this work is the first to show a link between the processing of musical rhythms perceived only via VTS inputs and phonological and prosodic abilities perceived through auditory input in infancy. In conclusion, the present findings might contribute to inform the future theoretical models of cognitive development as well as the next steps of cognitive research, with implications for health and education.

All data, materials, and code behind this analysis have been made publicly available at the *dataset*, *stimuli*, and *code* repositories respectively and can be accessed at https://osf.io/rxwk8/?view_only.

Author contribution

Conceptualization: S.R.; Methodology: S.R., F·C; Formal analysis: G. C., S.R.; Data curation: G.C, S.R.; Resources: F.C., A.R.; Software: F.C., A. R; Writing-original draft: S.R., G.C., E.V.; Project administration: S.R.; Supervision: E.V., A.R., B.A.; Funding acquisition: E.V., B.A.

Funding statement

This work was supported by the PRIN: Progetti di Ricerca di rilevante interesse Nazionale – Bando 2022 (Prot.2022748MSA) “The impact of rhythmic and predictable patterns on neuro-cognitive development “, CUP C53D23003970006.

Ethics approval statement

The research was conducted in accordance with the principles of the Declaration of Helsinki, and the research protocol was approved by the Ethics Committee of the University of Padua (2020_66R1). Caregivers signed a consent form before taking part in the experiment.

Author note

Sofia Russo. Mailing address: via Venezia 8 - 35131 Padova (IT). Tel: +39 049 827 6500.

Filippo Carnovalini. Mailing address: via Gradenigo 6/b 35131 - Padova (IT). Tel: +39 049 827 7600.

Giulia Calignano. Mailing address: via Venezia 8 - 35131 Padova (IT). Tel: +39 049 827 6500.

Barbara Arfé. Mailing address: via Venezia 8 - 35131 Padova (IT). Tel: +39 049 827 6575.

Antonio Rodà. Mailing address: via Gradenigo 6/b 35131 - Padova (IT). Tel: +39 049 827 7581. Fax: +39 049 827 7699.

Eloisa Valenza. Mailing address: via Venezia 8 - 35131 Padova (IT). Tel: +39 049 827 6583.

CRediT authorship contribution statement

Russo Sofia: Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Carnovalini Filippo:** Methodology, Software, Validation. **Calignano Giulia:** Data curation, Formal analysis, Methodology, Writing – review & editing. **Arfé Barbara:** Supervision, Writing – review & editing. **Rodà Antonio:** Methodology, Software, Supervision, Validation. **Valenza Eloisa:** Funding acquisition, Supervision.

Declaration of Competing Interest

The authors have no conflicts of interest to declare.

Data availability

I have shared the link to my data in the manuscript.

Acknowledgements

We would like to thank the Cooperative Now Project and Mr. Maria Pia Boellis for their support in data collection.

References

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

Abu-Zhaya, R., Seidl, A., & Cristia, A. (2017). Multimodal infant-directed communication: How caregivers combine tactile and linguistic cues. *Journal of Child Language, 44*(5), 1088–1116.

Ammirante, P., Patel, A. D., & Russo, F. A. (2016). Synchronizing to auditory and tactile metronomes: A test of the auditory-motor enhancement hypothesis. *Psychonomic Bulletin & Review, 23*, 1882–1890.

Aslin, R. N. (2007). What's in a look? *Developmental Science, 10*(1), 48–53.

Baayen, H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language, 94*, 206–234.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59* (4), 390–412.

Barton, K., & Barton, M. K. (2015). Package 'mumin. *Version, 1*(18), 439.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint*. arXiv:1406.5823.

Beatty, J., & Lucero-Wagoner, B. (2000). *The Pupillary System*.

Bowling, D. L., Graf Ancochea, P., Hove, M. J., & Fitch, W. T. (2019). Pupillometry of groove: Evidence for noradrenergic arousal in the link between music and movement. *Frontiers in Neuroscience, 12*, 1039.

Branje, C., & Fels, D. I. (2014). Playing vibrotactile music: A comparison between the Vibrochord and a piano keyboard. *International Journal of Human-Computer Studies, 72*(4), 431–439.

Caetano, G., & Jousmäki, V. (2006). Evidence of vibrotactile input to human auditory cortex. *Neuroimage, 29*(1), 15–28.

Calignano, G., Dispaldro, M., Russo, S., & Valenza, E. (2021). Attentional engagement during syllable discrimination: The role of salient prosodic cues in 6-to 8-month-old infants. *Infant Behavior and Development, 62*, Article 101504.

Calignano, G., Girardi, P., & Altoè, G. (2023). First steps into the pupillometry multiverse of developmental science. *Behavior Research Methods*, 1–20.

Calignano, G., Valenza, E., Vespignani, F., Russo, S., & Sulpizio, S. (2021). The unique role of novel linguistic labels on the disengagement of visual attention. *Quarterly Journal of Experimental Psychology, 74*(10), 1755–1772.

Changes in R. (2018). *The R Journal, 10*(1), 561–570.

Curtin, S., & Archer, S. L. (2015). Speech perception. In *The Cambridge Handbook of Child Language* (pp. 137–158).

Custode, S. A., & Tamis-LeMonda, C. (2020). Cracking the code: Social and contextual cues to language input in the home environment. *Infancy, 25*(6), 809–826.

DeCasper, A. J., & Sigafoos, A. D. (1983). *The Intrauterine Heartbeat: A Potent Reinforcer for Newborns*. Infant Behavior & Development.

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development, 9*(2), 133–150.

Delle Luche, C., Durrant, S., Poltrock, S., & Floccia, C. (2015). A methodological investigation of the intermodal preferential looking paradigm: Methods of analyses, picture selection and data rejection criteria. *Infant Behavior and Development, 40*, 151–172.

DePaolis, R. A., Keren-Portnoy, T., & Vihman, M. (2016). Making sense of infant familiarity and novelty responses to words at lexical onset. *Frontiers in Psychology, 7*, 715.

Doheny, L., Hurwitz, S., Insoft, R., Ringer, S., & Lahav, A. (2012). Exposure to biological maternal sounds improves cardiorespiratory regulation in extremely preterm infants. *The Journal of Maternal-Fetal & Neonatal Medicine, 25*(9), 1591–1594.

Fernald, A., Swingley, D., & Pinto, J. P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development, 72* (4), 1003–1015.

Fitch, W. T. (2013). Rhythmic cognition in humans and animals: Distinguishing meter and pulse perception. *Frontiers in Systems Neuroscience, 7*, 68.

Fiveash, A., Bedoin, N., Gordon, R. L., & Tillmann, B. (2021). Processing rhythm in speech and music: Shared mechanisms and implications for developmental speech and language disorders. *Neuropsychology, 35*(8), 771.

Fritzsche, T., & Höhle, B. (2015). Phonological and lexical mismatch detection in 30-month-olds and adults measured by pupillometry. In *ICPhS*.

Gervain, J. (2018). Gateway to language: The perception of prosody at birth. In *Boundaries Crossed, at the Interfaces of Morphosyntax, Phonology, Pragmatics and Semantics* (pp. 373–384).

Gescheider, G. A., & Niblette, R. K. (1967). Cross-modality masking for touch and hearing. *Journal of Experimental Psychology, 74*(3), 313.

Ghanizadeh, A. (2011). Sensory processing problems in children with ADHD, a systematic review. *Psychiatry Investigation, 8*(2), 89.

Giordano, M. (2016). *Vibrotactile Feedback and Stimulation in Music Performance*. Canada: McGill University.

Granier-Deferre, C., Ribeiro, A., Jacquet, A. Y., & Bassereau, S. (2011). Near-term fetuses process temporal features of speech. *Developmental Science, 14*(2), 336–352.

Gunther, E., & O'Modhrain, S. (2003). Cutaneous grooves: Composing for the sense of touch. *Journal of New Music Research, 32*(4), 369–381.

Hanley, J. A. (2016). Simple and multiple linear regression: Sample size considerations. *Journal of Clinical Epidemiology, 79*, 112–119.

Hannon, E. E., & Johnson, S. P. (2005). Infants use meter to categorize rhythms and melodies: Implications for musical structure learning. *Cognitive Psychology, 50*(4), 354–377.

Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: effects of enculturation and formal training on development. *Trends in cognitive sciences, 11*(11), 466–472.

Heinze, G., Wallisch, C., & Dunkler, D. (2018). Variable selection–A review and recommendations for the practicing statistician. *Biometrical Journal, 60*(3), 431–449.

Hepach, R., & Westermann, G. (2016). Pupillometry in infancy research. *Journal of Cognition and Development, 17*(3), 359–377.

Holland, S., Bouwer, A. J., Dalgelish, M., & Hurtig, T. M. (2010, January). Feeling the beat where it counts: Fostering multi-limb rhythm skills with the haptic drum kit. In *Proceedings of the Fourth International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 21–28).

Hollins, M., & Roy, E. A. (1996). Perceived intensity of vibrotactile stimuli: The role of mechanoreceptive channels. *Somatosensory & Motor Research, 13*(3–4), 273–286.

van Hooijdonk, R., Mathot, S., Schat, E., Spencer, H., van der Stigchel, S., & Dijkerman, H. C. (2019). Touch-induced pupil size reflects stimulus intensity, not subjective pleasantness. *Experimental Brain Research, 237*, 201–210.

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior research methods, 48*, 1393–1409.

Jackson, I., & Sirois, S. (2009). Infant cognition: Going full factorial with pupil dilation. *Developmental Science, 12*(4), 670–679.

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

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

Karam, M., Nespoli, G., Russo, F., & Fels, D. I. (2009, February). Modelling perceptual elements of music in a vibrotactile display for deaf users: A field study. In *2009 Second International Conferences on Advances in Computer-Human Interactions* (pp. 249–254). IEEE.

Karam, M., Russo, F. A., & Fels, D. I. (2009). Designing the model human cochlea: An ambient crossmodal audio-tactile display. *IEEE Transactions on Haptics, 2*(3), 160–169.

Karatekin, C. (2007). Eye tracking studies of normative and atypical development. *Developmental Review, 27*(3), 283–348.

Kisilevsky, B. S., Hains, S. M. J., Jacquet, A. Y., Granier-Deferre, C., & Lecanuet, J. P. (2004). *Maturation of Fetal Responses to Music*.

Klingner, J. (2010). *Measuring Cognitive Load During Visual Tasks by Combining Pupillometry and Eye Tracking*. Stanford University.

Knofczynski, G. T., & Mundfrom, D. (2008). Sample sizes when using multiple linear regression for prediction. *Educational and Psychological Measurement, 68*(3), 431–442.

Kotz, S. A., Ravignani, A., & Fitch, W. T. (2018). The evolution of rhythm processing. *Trends in Cognitive Sciences, 22*(10), 896–910.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5*(11), 831–843.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*(2), F13–F21.

Ladányi, E., Persici, V., Fiveash, A., Tillmann, B., & Gordon, R. L. (2020). Is atypical rhythm a risk factor for developmental speech and language disorders? *Wiley Interdisciplinary Reviews: Cognitive Science, 11*(5), Article e1528.

Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience, 27*(2), 308–314.

Lecanuet, J. P., & Granier-Deferre, C. (1993). Speech stimuli in the fetal environment. In *Developmental Neurocognition: Speech and Face Processing in the First Year of Life* (pp. 237–248).

Lecanuet, J. P., & Schaal, B. (2002). Sensory performances in the human foetus: A brief summary of research. *Intellectica, 34*(1), 29–56.

Lense, M. D., Ladányi, E., Rabinowitch, T. C., Trainor, L., & Gordon, R. (2021). Rhythm and timing as vulnerabilities in neurodevelopmental disorders. *Philosophical Transactions of the Royal Society B, 376*(1835), 20200327.

Lewkowicz, D. J. (2014). Early experience and multisensory perceptual narrowing. *Developmental Psychobiology, 56*(2), 292–315.

Lew-Williams, C., Ferguson, B., Abu-Zhaya, R., & Seidl, A. (2019). Social touch interacts with infants' learning of auditory patterns. *Developmental Cognitive Neuroscience, 35*, 66–74.

Li, H., Helpard, L., Ekeroot, J., Rohani, S. A., Zhu, N., Rask-Andersen, H., … Agrawal, S. (2021). Three-dimensional tonotopic mapping of the human cochlea based on synchrotron radiation phase-contrast imaging. *Scientific Reports, 11*(1), 4437.

Mani, N., Durrant, S., & Floccia, C. (2012). Activation of phonological and semantic codes in toddlers. *Journal of Memory and Language, 66*(4), 612–622.

Marimon, M., Höhle, B., & Langus, A. (2022). Pupillary entrainment reveals individual differences in cue weighting in 9-month-old German-learning infants. *Cognition, 224*, Article 105054.

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition, 1*(1).

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*, 314–324.

Mathôt, S., & Van der Stigchel, S. (2015). New light on the mind's eye: The pupillary light response as active vision. *Current Directions in Psychological Science, 24*(5), 374–378.

Mathôt, S., & Vilotijević, A. (2022). Methods in cognitive pupillometry: Design, preprocessing, and statistical analysis. *Behavior Research Methods*, 1–23.

Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology, 56*(2), 154–178.

Mazerolle, M. J., & Mazerolle, M. M. J. (2017). Package 'AICcmodavg'. In *R package* (p. 281).

Nobre, A. C., & Van Ede, F. (2018). Anticipated moments: Temporal structure in attention. *Nature Reviews Neuroscience, 19*(1), 34–48.

Oakes, L. M. (2017). Sample size, statistical power, and false conclusions in infant looking-time research. *Infancy, 22*(4), 436–469.

Patel, A. D. (2021). Vocal learning as a preadaptation for the evolution of human beat perception and synchronization. *Philosophical Transactions of the Royal Society B, 376*(1835), 20200326.

Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the beat: Movement influences infant rhythm perception. *Science, 308*(5727), 1430.

Phillips-Silver, J., & Trainor, L. J. (2007). Hearing what the body feels: Auditory encoding of rhythmic movement. *Cognition, 105*(3), 533–546.

Phillips-Silver, J., & Trainor, L. J. (2008). Vestibular influence on auditory metrical interpretation. *Brain and Cognition, 67*(1), 94–102.

Provasi, J., Anderson, D. I., & Barbu-Roth, M. (2014). Rhythm perception, production, and synchronization during the perinatal period. *Frontiers in Psychology, 5*, 1048.

van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the time course of pupillometric data. *Trends in Hearing, 23*, 2331216519832483.

Rocha, S., Southgate, V., & Mareschal, D. (2021). Infant spontaneous motor tempo. *Developmental Science, 24*(2), Article e13032.

Russo, S., Calignano, G., Dispaldro, M., & Valenza, E. (2021). An integrated perspective on spatio-temporal attention and infant language acquisition. *International Journal of Environmental Research and Public Health, 18*(4), 1592.

Russo, S., & Valenza, E. (2021). Apprendere attraverso il corpo, la musica e il ritmo: gli effetti positivi della Sincronizzazione Ritmica secondo i principi della Cognizione Incarnata. Nucleo tematico Musica, cognizione e apprendimento: tra neuroscienze ed educazione. In *Psicologia dell'Educazione, n2* (pp. 23–40).

Seidl, A., & Cristià, A. (2008). Developmental changes in the weighting of prosodic cues. *Developmental Science, 11*(4), 596–606.

Selkirk, E. (1995). Sentence prosody: Intonation, stress, and phrasing. In *, 1. The Handbook of Phonological Theory* (pp. 550–569).

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science, 5*(6), 679–692.

Sohmer, H., Perez, R., Sichel, J. Y., Priner, R., & Freeman, S. (2001). The pathway enabling external sounds to reach and excite the fetal inner ear. *Audiology and Neurotology, 6*(3), 109–116.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition, 76*(2), 147–166.

Tamási, K., McKean, C., Gafos, A., Fritzsche, T., & Höhle, B. (2017). Pupillometry registers toddlers' sensitivity to degrees of mispronunciation. *Journal of Experimental Child Psychology, 153*, 140–148.

Tamási, K., McKean, C., Gafos, A., & Höhle, B. (2019). Children's gradient sensitivity to phonological mismatch: Considering the dynamics of looking behavior and pupil dilation. *Journal of Child Language, 46*(1), 1–23.

Tanaka, Y., Kanakogi, Y., Kawasaki, M., & Myowa, M. (2018). The integration of audio–tactile information is modulated by multimodal social interaction with physical contact in infancy. *Developmental Cognitive Neuroscience, 30*, 31–40.

Teie, D. (2016). A comparative analysis of the universal elements of music and the fetal environment. *Frontiers in Psychology, 7*, 1158.

Tichko, P., Kim, J. C., Large, E., & Loui, P. (2022). Integrating music-based interventions with gamma-frequency stimulation: Implications for healthy ageing. *European Journal of Neuroscience, 55*(11–12), 3303–3323.

Tichko, P., Kim, J. C., & Large, E. W. (2021). Bouncing the network: A dynamical systems model of auditory–vestibular interactions underlying infants' perception of musical rhythm. *Developmental Science, 24*(5), Article e13103.

Tincoff, R., Seidl, A., Buckley, L., Wojcik, C., & Cristia, A. (2019). Feeling the way to words: Parents' speech and touch cues highlight word-to-world mappings of body parts. *Language Learning and Development, 15*(2), 103–125.

Tomchek, S. D., Huebner, R. A., & Dunn, W. (2014). Patterns of sensory processing in children with an autism spectrum disorder. *Research in Autism Spectrum Disorders, 8*(9), 1214–1224.

Trainor, L. J., Gao, X., Lei, J. J., Lehtovaara, K., & Harris, L. R. (2009). The primal role of the vestibular system in determining musical rhythm. *Cortex, 45*(1), 35–43.

Trainor, L. J., & Hannon, E. E. (2013). *Musical Development*.

Ullal-Gupta, S., Bosch, V., der Nederlanden, C. M., Tichko, P., Lahav, A., & Hannon, E. E. (2013). Linking prenatal experience to the emerging musical mind. *Frontiers in Systems Neuroscience, 7*, 48.

van Rij, J., Wieling, M., Baayen, R. H., van Rijn, H., & van Rij, M. J. (2017). *Package 'itsadug'*.

Velleman, S. L., & Vihman, M. M. (2007). Phonology in infancy and early childhood: Implications for theories of language learning. *Phonology in Context*, 25–50.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11*, 192–196.

White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language, 59*(1), 114–132.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 73*(6), 913–934.

Wood, S., & funded by EPSRC, P. (2007). *R mgcv-package*.

Woodruff Carr, K., White-Schwoch, T., Tierney, A. T., Strait, D. L., & Kraus, N. (2014). Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proceedings of the National Academy of Sciences, 111*(40), 14559–14564.

Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science, 12*(3), 412–418.

Zimmerman, E., & Barlow, S. M. (2012). The effects of vestibular stimulation rate and magnitude of acceleration on central pattern generation for chest wall kinematics in preterm infants. *Journal of Perinatology, 32*(8), 614–620.