

Vocalizations produced in the second year of life in response to speaking and singing

Psychology of Music

1–12

© The Author(s) 2017

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0305735617719335

journals.sagepub.com/home/pom



João Reigado¹ and Helena Rodrigues²

Abstract

We compared infants' vocalizations produced both in singing and speaking conditions. In this study we collected the vocalizations of 12 infants aged between 12 and 24 months. Each 30-minute session integrated two stimulus conditions – speaking and singing. An acoustic analysis was developed in order to measure both segmental properties and melodic and rhythmic features. The results showed significant differences in the vocalizations to which the two stimulus conditions gave rise: in the duration of the vocalizations and the duration of the phonation, the number of nuclei and the duration of each nucleus present in the vocalizations. In particular, the greater extension of vowels in vocalizations in response to the singing condition seems to be distinctive from those produced in response to the speaking condition. This may reveal that there is a precocious distinction between the singing voice and the speaking voice. The acoustic analysis that we used shows promise in monitoring children's sequential singing development from early infancy onwards.

Keywords

acoustics, early childhood, infants, singing, speech, voice

Voice and listening as tools for singing and language acquisition

The vocal tract is responsible for producing speech and singing. These two domains share cognitive and neural processes and their relationship has interested researchers in both areas for many years. Despite this interest, there has been little research devoted to the comparison of vocalizations produced within the context of speech and singing among infants and toddlers. The intricacy of such an assessment derives from the set of variables to be considered, including

¹Centro de Estudos de Sociologia e Estética Musical, Faculdade de Ciências Sociais e Humanas, Universidade NOVA de Lisboa, Portugal

²Centro de Estudos de Sociologia e Estética Musical, Departamento de Ciências Musicais, Faculdade de Ciências Sociais e Humanas, Universidade NOVA de Lisboa, Portugal

Corresponding author:

João Reigado, Centro de Estudos de Sociologia e Estética Musical, Faculdade de Ciências Sociais e Humanas, Universidade NOVA de Lisboa, Avenida de Berna, 26-C 1069-061 Lisbon, Portugal.

Email: joaoreigado@fcs.unl.pt

perceptual, cognitive, physical, cultural elements and even the intention of the vocal act (Welch, 1994).

In terms of perceptual variables it must be said that infants and children show great sensitivity to the acoustical parameters of sounds and perception of musical attributes, such as melody, tempo and rhythm (Stalinski & Schellenberg, 2010; Costa-Giomi & Ilari, 2014; Trehub & Gudmundsdottir, 2015; Trehub & Hannon, 2009; Trehub & Thorpe, 1989). The importance of the musical environment, specifically as far as infants' spontaneous responses to music and motor development is concerned, is well documented (Adachi, Nakata & Kotani, 2002) and a focus on culture as an important determinant of the musical engagement of young children was recently proposed by Ilari (2015). On account of its very specific characteristics, infant-directed singing was found to regulate infants' emotional states (de l'Etoile, 2006; Ghazban, 2013; Nakata & Trehub, 2004) and lullabies constitute an important part of infants' pre-linguistic communication.

Mothers and other caregivers interact with infants very closely and with much eye contact, emphasizing facial expressions and employing frequent repetition. These multimodal features attract the attention of young children who respond more easily to this communicative style (de l'Etoile, 2006; Lewkowicz, 1998; Trehub, Plantinga & Russo, 2016). In fact, in lullabies, the expressive and emotional aspects the caregiver uses in the presence of infants are highlighted more than the intrinsic elements (music or lyrics) that constitute them (Rodrigues, 2005). According to Longhi (2009), certain musical features highlighted in such songs establish regularities that help the infant to segment the flow of the song and elicit his or her first vocalizations. Thus, a mother's singing seems to emphasize the hierarchical structure of the song and provides a segmentation of the temporal structure of the interaction. This way of using the singing voice to address young children was termed "songese" by Longhi (2009). These early vocalizations were rapidly noticed as attracting the interest of mothers, thus reinforcing communication between them (Gratier & Magnier, 2012; H. Papousek, 1996; M. Papousek, 1996; Reigado, 2009; Van Puyvelde et al., 2010).

Singing acquisition starts early in infancy

Some researchers have stated that singing development starts early in infancy. According to Moog (1968/1976), 6-month-old infants make use of isolated, mainly vocalic sounds, just after bodily movements occur as an active response to music. Gordon (2000) stated that 6-month- to 1-year-old infants perform descending melodic patterns within a reduced pitch range. Two-year-olds can imitate rhythmic and melodic contours (Dowling, 1999), which reach clearer tonal relationships at around 3 years old (Hargreaves, 1996). Tafuri and Villa's research on musical development points to a stage of musical babbling, which occurs earlier among infants who have had previous exposure to musical experiences, both prenatal and neonatal (Tafuri & Villa, 2002). They also observed that the ability to sing in tune develops even before 3 years of age, which means that both the phonatory and perceptive-cognitive mechanisms are already developed by this age. Reigado, Rocha, and Rodrigues (2011) found that 9- to 11-month-olds produce shorter vocalizations in reaction to singing stimuli when compared with the duration of the vocalizations obtained after speaking stimuli. Moreover, both mean and maximum F0 (fundamental frequency) values were lower in the vocalizations produced in reaction to singing. Since this approach involved comparing acoustical features obtained in two different contexts of stimulation, the results obtained reflected processing specificity pertaining to the acquisition of singing.

It may be important to consider acoustical clues derived from studies with adult participants in the analysis of infant vocal development. In one of these examples, Scotto Di Carlo (2005; see also Scotto Di Carlo & Autesserre, 1992) ascertained that the duration of vowels increases when changing from speech to singing, while the consonant part of a syllabic nucleus tends to diminish. The researcher observed that this behavior illustrates the timbre as an important aspect of the specificity of a vocal context. In her opinion, this happens because the vowels are the ideal support for melody while the consonants affect the quality of *legato*, interrupting the melodic line.

Language acquisition research as an inspirational basis for the study of early attempts at singing

Research in the field of language acquisition has shown that infants already seem to distinguish changes in the syllables of utterances at 4 days old (Bertoncini, Floccia, Nazzi, & Mehler, 1995) suggesting that they may soon thereafter identify the elementary units of the rhythmic organization of a specific language. Since infants are sensitive to syllabic patterns – vowels or sequences of vowels, consonants and syllables – and these three units of length are particularly distinctive during the initial acquisition of speech (Bertoncini et al., 1995), it is expected that they would be able to highlight distinct features when conditioned by different vocal stimuli, such as singing and speech.

In the field of cross-linguistic variation in prosody, a number of investigations into the acquisition of rhythmic patterns in different languages have made use of variability indexes. Grabe, Post and Watson's (1999) pair-wise variability index (PVI) explores knowledge concerning vowel duration in both stress-timed and syllable-timed languages. Based on this research, stress-timed languages are expected to have relatively large PVI values because the durations of vowels in successive syllables vary. More recently, PVI indexes were used to characterize the rhythm of infant-directed speech and infant-directed singing, in two Portuguese variants – European and Brazilian; Salselas & Herrera, 2011). The researchers observed that, from a set of five characteristics that predict the linguistic variant of infant-directed speech, four of them concerned rhythm. The authors found that these indices were higher in European compared to Brazilian Portuguese, retaining the same characteristics of conventional adult speech. It was also observed that the rhythmic characteristics that were predictive of the Portuguese variant in infant-directed speech were also highly predictive in relation to songese. The PVI values found for songese were also higher in European than Brazilian Portuguese.

These two reported aspects – rhythmic patterns and vowel duration – have not yet been reliably analyzed in terms of infants' vocal production. Reliable and valid observations and analysis of infants' singing acquisition is missing. In fact, most studies on vocal development in infancy are concerned with speech production. In this sense, Stadler Elmer's methodology of analysis, based on acoustic measurement, is innovative (Stadler Elmer, 2008). Her theoretical proposal combines the analysis of children's singing by a computer-aided program that allows pitch and timing representation with a detailed notation viewer. The results that Stadler Elmer was able to obtain, from the observation of infants from 10–14 months old on, demonstrated the potential value of analysis focused on the structure of vocal behavior.

As may be seen, previous research seems to reveal the existence of an early vocalization that is specific for music. Nevertheless, very little is known about the acoustic rudiments of infants' singing. The psychoacoustic study we ran looked for distinctive features in the vocal production of infants in their second year of life, when presented with speaking and singing stimuli.

Method

Participants

Twelve infants (seven girls and five boys), aged 12 months ($M = 11.5$) participated in this study. They were recruited through an advertisement asking for volunteers to participate in biweekly sessions over the course of one year with the aim of investigating infants' behavior in response to speaking and singing. All the participants were monolingual (European Portuguese) with no developmental problems.

In return, we provided four music sessions for participants and their parents, which took place after the final phase of data collection. Participants were not involved in musical guidance sessions or in any specific music program for families before or during the study.

Materials

In accordance with the two conditions to be studied – speaking and singing – free conversation and three songs without words were used to encourage the infant to speak or sing. The songs contrasted with one another in terms of melodic and rhythmic contour, tonality, tempo and musical form. They were composed specifically for this study, with the intention of ensuring the unfamiliarity of the singing stimuli.

The sessions were taped using three microphones (DPA 4060 Hi-Sens), suspended from the ceiling and an M-Audio ProFire™ 2626 interface (24 bit/192 kHz) with preamp (Octane™) and JetPLL technology to eliminate jitter, connected to an iMac with 24-inch display (2.8 GHz/8 GB SDRAM/640 GB HD). The software Audacity 1.3.9-beta (<http://audacity.sourceforge.net/>) was used to divide each period of an infant's vocal productions into isolated vocalizations and also for labeling the sound samples. The acoustic analysis was carried out using two related tools: first, Praat 4.4.04 (Boersma & Weenink, 2006) was used to establish the highest and lowest levels for the detection of the F0 curve; second, Prosogram¹ (Mertens, 2004) was used to perform the automatic segmentation of the acoustic signal into sequences of syllabic nuclei.

Procedure

Each infant was observed every two weeks, between 12 and 24 months, in 30-minute sessions, resulting in the collection of 26 recorded sessions for each participant.

We chose a longitudinal methodology in order to clarify the pathways of singing acquisition at an early age, in which speaking acquisition strategies are concomitant. Notwithstanding, in this article we focus on the description of the vocalizations that the infants produced in reaction to linguistic and music conditions.

In each session, the dyad formed by the participant and the mother/father sat on the floor, facing the researcher, who was responsible for the presentation of stimuli. The experiment consisted of two parts: linguistic, with which the sessions always began, and musical. It was expected that beginning the sessions with speech stimuli would be effective in breaking the ice and helping to minimize the strangeness of the context for the participant, helping his/her social engagement.

Under the first condition, the stimulus was based either on speaking interactions or on moments of silence after conversation. The topic was free and, when necessary, use was made of soft toys to elicit vocalizations.

Under the singing condition, the stimuli comprised three elements: songs, singing interactions based on each song, and silence. Adhering to a regular beat, the adult would perform the

three songs in the same order. The determination of metronomic time enabled each song always to last for 40 seconds. The moments of singing interaction lasted 40 seconds each and were intended for the execution of tonal and melodic patterns from the song by the researcher and any vocal response by the participants.

Under both conditions, the moments of silence were essential for obtaining infants' vocalizations with no overlap with the voice of the researcher.

The sessions were recorded and the vocal output was collected and coded for further analysis. A total of 662 vocalizations produced in response to speech ($N = 449$) and in response to singing ($N = 213$) were collected.

Data analysis

In order to operationalize the variables, the beginnings and ends of vowels were obtained and were used to calculate the duration of each vowel or vowel sequence, consonants, and syllable. The acoustic features were then observed on two levels: one of the approaches was based on the analysis of the characteristics of segmental properties of each participant's vocal output, observing the duration and number of nuclei and the speed of vocal articulation under two conditions; the other approach was intended to describe both melodic and rhythmic features of the vocalizations, which were obtained from the analysis of tonal and durational parameters.

For pitch description we considered the minimum, mean, and maximum FO of each vocalization. The interest in determining these three levels of FO resides in the fact that it can reveal differences in the vocal range used in both conditions. The study of the melodic contour was based on both the sum of absolute variations of FO within a nucleus (intraabsFO) and in the sequence of nuclei within a vocalization (interabsFO).

The melodic interval variability (MIV_v) present in vocalization was also calculated. This measure may show differences in the way the melodic intervals succeed each other in the same vocalization under the two stimulation conditions. In the equation below, MIV_v is equivalent to 100 times the coefficient of variation (CV) of interabsFO (as proposed by Patel, 2008).

$$MIV_v = 100CV_{interabsFO}$$

Since it is based on the analysis of the variation of successive melodic intervals, this index cannot be calculated in vocalizations in which there are fewer than three nuclei (i.e., $n_{nucleos} \geq 3$).

The contrast between successive durations for a set of three types of nuclei was also calculated – vocalic ($nPVI_v$), consonantic ($rPVI_c$) and syllabic ($rPVI_{sil}$). The PVI (pairwise variability index) can reveal changes in the length of those elements that are constituents of vocalizations. In the equations below, m is the number of nuclei per vocalization and d_k is the length of the core of k order (Patel, 2008).

$$nPVI = \frac{100}{m-1} \times \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{\frac{d_k + d_{k+1}}{2}} \right|$$

$$rPVI = \left[\sum_{k=1}^{m-1} \frac{|d_k - d_{k+1}|}{m-1} \right]$$

Table 1. Number of vocalizations by sex and experimental condition.

	Frequency (%)		Total
	Speaking	Singing	
Boys	116 (55.2%)	94 (44.8%)	210
Girls	333 (73.7%)	119 (26.3%)	452
N	449 (67.8%)	213 (32.2%)	662

As in the case of MIVv, this index requires a basic assumption for its calculation. Since it reflects the variation of the durations of the elements of a vocalization, it must be formed by at least two nuclei ($n_{nucleos} \geq 2$).

Results

In this section, we describe the analysis of both segmental properties and melodic and rhythmic features present in the infants' vocalizations produced under speaking and singing conditions. Table 1 shows the results of the sample of vocalizations.

The total of vocalizations is far superior under the speaking condition and much higher for girls than for boys. This result is mainly accounted for by a very large number of vocalizations that occurred under the speaking experimental condition produced by girls, as compared to the singing experimental condition, in which the difference in sex is not significant. The Chi-square test for independence with Yates' continuity correction revealed that the percentage of speaking vocalizations significantly differed by sex (boys/girls) ($\chi^2_{(1, N=662)} = 21.491, p < .0001$).

Segmental properties

Mean duration (in seconds) of the vocalization was longer under the speaking condition ($M = 1.52, SD = 1.46$) than under the singing condition. The phonation time was also longer under the speaking condition ($M = .98, SD = .77$).

The average number of nuclei per vocalization was greater under the speaking condition ($M = 5.48, SD = 4.47$).

Despite the fact that the overall duration of the vocalization and the phonation time were higher under the speaking condition, the average duration (in seconds) of each nucleus was found to be higher under the singing condition ($M = .087, SD = .036$). The vocalizations produced in reaction to the singing condition were also characterized by a greater speed of vocal production ($M = 6.13, SD = 4.89$) than under the speaking condition. Figure 1 represents graphically the variables in question.

Figure 1 shows greater dispersion of the durations of vocalizations, phonation time, and number of nuclei under the speaking condition. In Figure 2, both duration of nucleus and the speed of vocal productions are more dispersed under the singing condition.

An ANOVA was run in order to verify whether the differences between the values of the segmental properties found for each condition were significant. The ANOVA showed that the effect of the stimulus condition was significant for duration of vocalization ($F(1, 660) = 5.54, p = .019$), phonation time ($F(1, 660) = 8.19, p = .004$), and number of nuclei ($F(1, 660) = 12.20, p = .001$) and $n_{nucleos}$ ($F(1, 660) = 8.77, p = .003$).

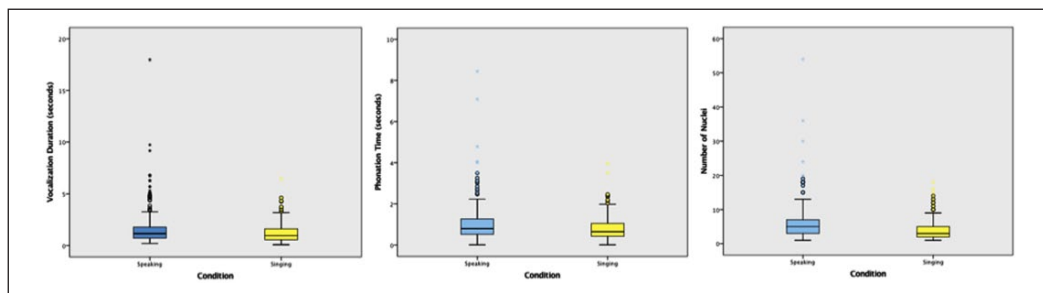


Figure 1. Duration of the vocalization (left), phonation time (center) and number of nuclei (right) between infants' vocal productions under speaking and singing conditions.

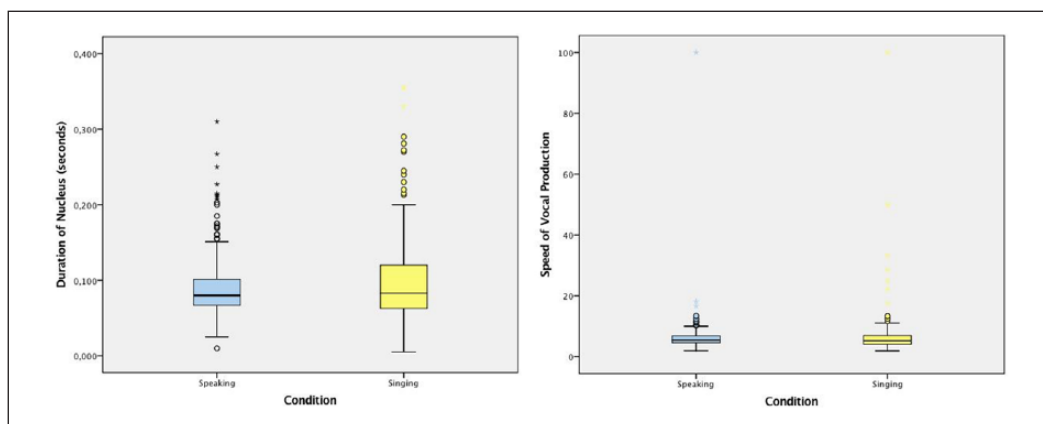


Figure 2. Mean duration of nucleus (left) and speed of vocal production (right) between infants' vocal productions under speaking and singing conditions.

Melodic and rhythmic analysis

Using an acoustic analysis, we also tried to understand the nature of melodic and rhythmic vocalizations, by analyzing tonal and durational parameters. All of the three levels of F0 measurement (Hz) were higher in the speaking condition (minimum F0: $M = 258$, $SD = 151$; mean F0: $M = 406$, $SD = 115$; maximum F0: $M = 563$, $SD = 152$).

The absolute variation of F0 both in a nucleus (*intraabsF0*) and between successive nuclei (*interabsF0*) obtained higher values for the speaking condition ($M = 10.4$, $SD = 9.34$; $M = 33.8$, $SD = 48.3$, respectively). The vocalizations produced under the speaking condition presented higher variability in the durations of nuclei, not only for vowels ($nPVIv$: $M = 66.2$, $SD = 24.3$) but also for consonants ($rPVIc$: $M = .18$, $SD = .26$) and for syllables ($rPVIsl$: $M = .21$, $SD = .26$).

The variability of the amplitude of melodic intervals (*MIVv*) was also higher under the speaking condition ($M = 94.3$, $SD = 40.1$) than under the singing condition. Figure 3 accounts for the dispersion of these variables.

To test the significance of the above differences, an ANOVA was computed. The ANOVA confirmed the effect of the stimuli condition on mean F0 ($F(1, 516) = 4.8$, $p = .029$) and maximum F0 ($F(1, 516) = 4.36$, $p = .037$).

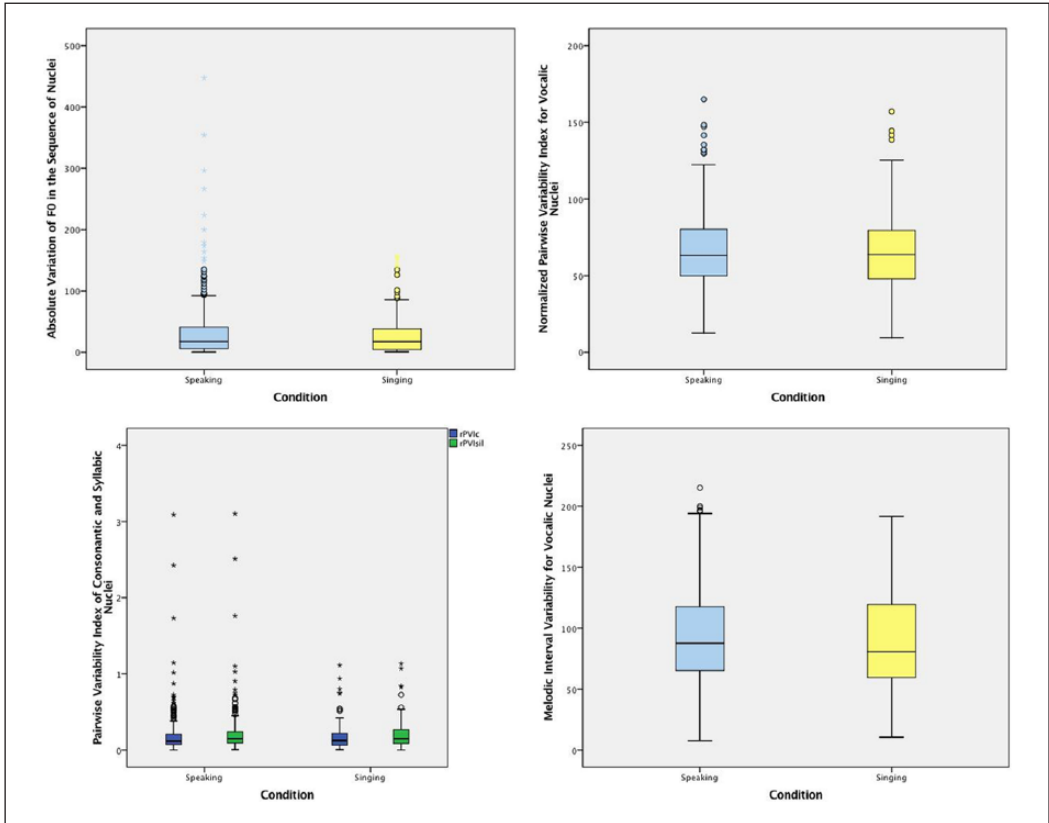


Figure 3. *Top:* Minimum, mean, and maximum F0 values in infants' vocalizations and absolute variation of F0 in a nucleus; *Middle:* absolute variation of F0 in the sequence of nuclei and normalized pairwise variability index for vocalic nuclei (*nPVIv*); *Bottom:* pairwise variability index of both consonantic and syllabic nuclei and melodic interval variability for vocalic nuclei (under speaking and singing conditions).

Discussion

Results showed that infant's vocalizations elicit differences in both segmental properties and melodic and rhythmic features when produced in reaction to singing and speaking conditions.

The fact that there were more nuclei under the speaking condition seems to show the greater familiarity of participants with the characteristics of verbal language, possibly because of greater daily exposure to speaking stimulation. In fact, several studies show that vocal productions at this age are clearer and that series of varied syllables become more numerous (M. Papousek, 1996). In our findings, the vocalizations produced under the speaking condition are longer than those produced in the context of a song.

However, it is interesting to verify that the average length of the nuclei is higher under the singing condition. That is, though the vocal productions in the context of a song are shorter, the lengths of the nuclei are greater, showing what appears to be a systematic prolongation of the vowels, when infants vocalize under the singing condition. These results confirm the observations of Dowling (1999), who notes that, at around 2 years old, there is greater extension of vowels in vocalizations in response to music compared to those produced in a speaking context.

They are also in line with the observations of Scotto Di Carlo (2005) on the intelligibility of the adult singing voice, in which she notes that during singing the syllable undergoes a substantial prolongation, which leads to an increase in the relative duration of vowels and a decrease in the relative duration of consonants. It is surprising that the same kind of results were found in infant vocal productions.

As observed with the vocalizations of 9- to 11-month-old infants (Reigado et al., 2011), F0 analysis confirms higher values in response to speaking stimulation, leading us to the conclusion that infants restrict their vocal range under the singing condition in order to bring it closer to the range in which the adult sings the song. Using the well-tempered tonal system, we concluded that the range of infants' singing vocalizations is consistent with Gordon's findings (2000). That is, the infants' initial singing voice extends from D above middle C to B above middle C. In the case of the speaking condition, the vocal range includes D# above middle C to C#, a minor seventh above. These results reinforce Gordon's observation concerning initial singing and speaking ranges, which seem to overlap.

We found that infants tend to stabilize the pitch on each tone in a vocalization, when responding to the singing condition. The greater extension of vowels in vocalizations in response to the singing context seems to be a distinctive feature in comparison with those produced in response to the speaking condition. This behavior shows similarities with adult performances, wherein each sung syllable matches, in most cases, a single tone with well-defined pitch. A possible explanation for the occurrence of this behavior is that when singing there is a need to hold a tone for longer than is usually necessary in speech, in order to make the rhythm of the song apparent.

It was also found that, under the singing condition, the vocalization progresses around a stable system of melodic intervals, i.e., through a melodic line, which makes use of tones that are close in pitch. These findings converge with Patel (2008), who notes that there is a higher occurrence of small intervals in musical melodies, which, according to the author, accounts for the listener's perception of melodic continuity. This seems to be in line with what we have observed in the musical guidance sessions for families that we have been leading for several years. In these, the occurrence of vocal productions including minor and major seconds is common. Also, when infants listen to melodies that are interrupted at points that sound inconclusive they tend to vocalize, filling the gap (Hornbach, 2005; Valerio & Reynolds, 2009) by singing sounds that are close in terms of pitch to the last one heard (Rodrigues & Rodrigues, 2003; Valerio, Seaman, Yap, Santucci, & Tu, 2006). It seems possible to affirm that children tend to complete the musical phrase, incorporating tones that are adjacent to each other in terms of pitch.

With regards to rhythm, results showed that infants produce more durational contrasts when vocalizing in response to the speaking condition. On the other hand, vocal production under the singing condition displays a certain regularity in the duration of the tones and a less variable rhythm overall. However, this difference was not significant. Other studies in which this index was also observed in the context of the Portuguese language do not allow us to make significant comparisons with the results obtained in this investigation. For example, using a set of infant-directed outputs in two Portuguese language variants, Salselas and Herrera (2001), verified that the nPVI of European Portuguese (EP) was greater than Brazilian Portuguese (BP). However, we can speculate that infants' pre-linguistic productions display, from an early age, the typical rhythmic organization of their native language (it should be noted that EP is considered a stress-timed language, with a high durational contrast between consecutive syllables). Rather, the rhythmic organization of the vocalizations that were produced in the singing condition presents more regular durations.

To our knowledge, there are no published studies using this measure when comparing the vocal output of infants, both spoken and sung. Nevertheless, the use of this index seems to be promising, as it allowed the separation of infant vocal productions under speaking and singing conditions, according to rhythm.

In sum, our findings seem to indicate that there are acoustic features that account for the distinction between the vocalizations produced in response to speaking and singing in the second year of life. In light of these results, it can be assumed that singing behavior emerges very early. It might reveal there is a precocious distinction between the singing voice and the speaking voice in terms of vocal production. It also indicates the existence of an early predisposition towards music.

This research also sought to establish a reliable methodology for analyzing singing development in infants and young children. From a survey of the literature, we conclude that no study on the development of the singing voice has crossed, objectively, both the segmental properties and melodic and rhythmic aspects of infant vocal productions. Thus, this methodology may offer some promise in monitoring sequential singing development from early infancy, allowing more extensive analysis of the vocal outputs of infants both younger and older than those who participated in this research. We propose that the use of such an analytical methodology may open up a space in which the current model of experimental research in musical development can be further investigated.

Acknowledgements

We thank Ana Isabel Pereira and Ekanem Ebinne for their review and advice.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: the Portuguese Fundação para a Ciência e Tecnologia (PTDC/EAT/6836/2006). Thanks are also due to the Portuguese Ministry of Education, which supported the first author of this article between 2009 and 2011.

Note

1. Prosogram is a Praat-integrated plug-in that provides a representation of intonation, considering that auditory perception of pitch variations depends on many factors other than F0 variation proper. Prosogram produces a representation that aims to capture the perceived pitch patterns of speech melody (a stylization based on perceptual principles).

References

- Adachi, M., Nakata, T., & Kotani, Y. (2002). Infants encountering music: An exploration of musical affordances. In C. Stevens, D. Burnham, G. McPherson, E. Schubert, & J. Renwick (Eds.), *Proceedings of the 7th International Conference on Music Perception and Cognition* [CD-ROM] (pp. 454–456). Adelaide, Australia: Causal Productions.
- 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.
- Boersma, P., & Weenink, D. (2006). Praat: Doing phonetics by computer (Version 4.4.04) [Computer software]. Retrieved from <http://www.praat.org/>
- Costa-Giomi, E., & Ilari, B. (2014). Infants' preferential attention to sung and spoken stimuli. *Journal of Research in Music Education*, 62(2), 188–194.
- de l'Etoile, S. K. (2006). Infant-directed singing: A theory for clinical intervention. *Music Therapy Perspectives*, 24(1), 22–29.

- Dowling, W. J. (1999). The development of music perception and cognition. In D. Deutsch (Ed.), *The psychology of music* (pp. 603–625). San Diego: Academic Press.
- Ghazban, N. (2013). *Emotion regulation in infants using maternal singing and speech* (Unpublished doctoral dissertation). Ryerson University, Toronto, Canada.
- Gordon, E. (2000). *Teoria de aprendizagem musical. Competências, conteúdos e padrões* (M. F. Albuquerque, Trans.). Lisbon: Fundação Calouste Gulbenkian.
- Grabe, E., Post, B., & Watson, I. (1999). The acquisition of rhythmic patterns in English and French. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco* (pp. 1201–1204). Retrieved from http://www.jimflege.com/files/McAllister_Flege_ICPhS_San_Francisco_1999.pdf
- Gratier, M., & Magnier, J. (2012). Sense and synchrony: Infant communication and musical improvisation. *Intermédialités*, 19, 45–64.
- Hargreaves, D. (1996). The development of artistic and musical competence. In I. Deliège & J. Sloboda (Eds.), *Musical beginnings: Origins and development of musical competence* (pp. 145–170). Oxford: Oxford University Press.
- Hornbach, C. M. (2005). *Ah-eee-ah-eee-yah-eee, bum and pop, pop, pop: Teacher initiatives, teacher silence, and children's vocal responses in early childhood music classes* (Unpublished doctoral dissertation). Michigan State University, East Lansing, MI.
- Ilari, B. (2015). Rhythmic engagement with music in early childhood: A replication and extension. *Journal of Research in Music Education*, 62(4), 332–343.
- Lewkowicz, D. J. (1998). Infants' response to the audible and visible properties of the human face: II. Discrimination of differences between singing and adult-directed speech. *Developmental Psychobiology*, 32, 261–274.
- Longhi, E. (2009). 'Songese': Maternal structuring of musical interaction with infants. *Psychology of Music*, 37, 195–213.
- Mertens, P. (2004). The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model. In B. Bel & I. Marlien (Eds.), *Proceedings of Speech Prosody 2004, Nara, Japan* (pp. 23–26). Groningen, The Netherlands: University of Groningen.
- Moog, H. (1968/1976). *The musical experience of the pre-school child* (C. Clark, Trans.). London: Schott.
- Nakata, T., & Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behavior & Development*, 27, 455–464.
- Papousek, H. (1996). Musicality in infancy research: Biological and cultural origins of early musicality. In I. Deliège & J. Sloboda (Eds.), *Musical beginnings: Origins and development of musical competence* (pp. 37–51). Oxford: Oxford University Press.
- Papousek, M. (1996). Intuitive parenting. In I. Deliège & J. Sloboda (Eds.), *Musical beginnings: Origins and development of musical competence* (pp. 88–112). Oxford: University Press.
- Patel, A. (2008). *Music, language, and the brain*. New York: Oxford University Press.
- Reigado, J. (2009). *Análise acústica das vocalizações de bebés de 9 a 11 meses face a estímulos musicais e linguísticos* [Acoustic analysis of infants aged 9–11 months towards musical and linguistic stimuli]. Lisbon: Edições Colibri.
- Reigado, J., Rocha, A., & Rodrigues, H. (2011). Vocalizations of infants (9–11 months old) in response to musical and linguistic stimuli. *International Journal of Music Education*, 29(3), 241–255.
- Rodrigues, H. (2005). A Festa da Música na iniciação à vida: Da musicalidade das primeiras interações humanas às canções de embalar [The festival of music at the beginning of life: From the musicality of first human interactions to human lullabies]. *Revista da Faculdade de Ciências Sociais e Humanas*, 17, 61–80.
- Rodrigues, H., & Rodrigues, P. M. (2003). *BebéBabá: da musicalidade dos afectos à música com bebés*. Porto: Campo das Letras.
- Salselas, I., & Herrera, P. (2011). Music and speech in early development: Automatic analysis and classification of prosodic features from two Portuguese variants. *Journal of Portuguese Linguistics*, 9(10), 11–36.
- Scotto Di Carlo, N. (2005). Contraintes de production et intelligibilité de la voix chantée. *Travaux Interdisciplinaires du Laboratoire Parole et Langage*, 24, 159–179.

- Scotto Di, Carlo, N., & Autesserre, D. (1992). L'organisation temporelle de la syllabe dans la parole et dans le chant. *Travaux de l'Institut de Phonétique d'Aix*, 14, 189–204.
- Stadler Elmer, S. (2008, June 29–July 4). A developmental approach to voice science. In Proceedings from *Acoustics'08 Paris*. Retrieved from <http://www.acoustics08-paris.org>.
- Stalinski, S., & Schellenberg, E. G. (2010). Shifting perceptions: Developmental changes in judgments of melodic similarity. *Developmental Psychology*, 46(6), 1799–1803.
- Tafari, J., & Villa, D. (2002). Musical elements in the vocalizations of infants aged 2–8 months. *British Journal of Music Education*, 19(1), 73–88.
- Trehub, S. E., & Gudmundsdottir, H. R. (2015). Mothers as singing mentors for infants. In G. F. Welch, D. M. Howard, & J. Nix (Eds.), *The Oxford handbook of singing* [Advance online publication]. Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780199660773.013.25
- Trehub, S., & Hannon, E. (2009). Conventional rhythms enhance infants' and adults' perception of musical patterns. *Cortex*, 45, 110–118.
- Trehub, S. E., Plantinga, J., & Russo, F. A. (2016). Maternal vocal interactions with infants: Reciprocal visual influences. *Social Development*, 25(3), 665–683.
- Trehub, S. E., & Thorpe, L. A. (1989). Infants' perception of rhythm: Categorization of auditory sequences by temporal structure. *Canadian Journal of Psychology*, 43, 217–229.
- Valerio, W. H., & Reynolds, A. M. (2009). Peekaboo and fill-in-the blank: Age-old games with a social-music-interaction twist. *Perspectives*, 4(3), 10–15.
- Valerio, W., Seaman, M. A., Yap, C. C., Santucci, P. M., & Tu, M. (2006). Vocal evidence of toddler music syntax acquisition: A case study. *Bulletin of the Council for Research in Music Education*, 170, 33–45.
- Van Puyvelde, M., Vanfleteren, P., Loots, G., Deschuyffeleer, S., Vinck, B., Jacquet, W., & Verhelst, W. (2010). Tonal synchrony in mother-infant interaction based on harmonic and pentatonic series. *Infant Behavior and Development*, 4(33), 387–400.
- Welch, G. (1994). The assessment of singing. *Psychology of Music*, 22(3), 3–19.