CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

How a young child sings a well-known song before she can speak

Stefanie Stadler Elmer [iD]

University of Zurich, Switzerland
Email: stefanie.stadlerelmer@uzh.ch

Abstract
Through micro-genetic analysis of early singing, I describe and explain the complexity of song as an elementary cultural expression. For educators, it is important to understand the key role of song with and by young children as a means to convey feelings and musico-linguistic rules. Song consists of melody and lyrics, both of which are connected by metrical rules to form a Gestalt. A song sung by 18-month-old Lynn exemplifies that she produces the melody with ease, but shows difficulty forming the words. By following rules, she forms and expresses feelings of belonging to those who shared their singing with her previously.

Keywords: Early singing; vocal development; song; micro-genetic analysis; grammar of songs for children

Introduction: Vocal development and origins of song and speech

Singing with and by young children is considered the prototypical cultural practice to induce and express feelings by generating vocal sounds in a playful and ritual manner. For music educators, it is important to understand the role of song in early childhood. No other musical expression begins so early and, due to the human-specific vocal learning capacity (Merker, 2012), infants adapt their vocalisation flexibly to the song and speech – or music and language – of their surrounding culture. As songs for young children are traditionally enriched with narrative, motor and material elements, song as a social practice can be seen as the ideal way of introducing children to their musico-linguistic lore and culture. Children have the specific human capacity to learn vocally, that is, to reproduce that which has been heard by ear (Merker, 2012). It evolves from the simplest and most elementary forms to ever more complex forms, starting with the infant's crying and exploring the voice towards singing and speaking.

In the research literature, the singing of young children is characterised by almost the same properties, but authors differ in how much weight they give to them: glissandi, unstable pitches, chants with undefinable sounds, neologisms, short phrases within a narrow vocal range, small and inaccurately tuned intervals (overview by Stadler Elmer, 2012a). Developmental progress is described by songs becoming longer, more stable with respect to tonality and rhythm, more complex, and linguistically more accurately pronounced. Some authors selectively take the accuracy of intervals or pitches as a key feature of singing and assume in-tune-singing to increase with age. Other authors claim that children primarily focus on and produce linguistic features both in song acquisition and development (overview by Stadler Elmer, 2011). Yet, when conceptualising song as a rule-governed whole – as presented below – pitch accuracy and linguistic features turn out to be just some properties among others that form a complex configuration.

Infants explore their voice. At the age of 5–10 months, they usually produce repetitive syllables, that is, canonical babbling (e.g., Nathani, Ertmer & Stark, 2006). Canonical babbling

is not semantised yet, and it can be seen as the undifferentiated beginning of singing- and speech-like expressions. Thus, canonical babbling can be considered as the precursor of song and speech and as their common root (Stadler Elmer, 2020). Feelings of ease and playfulness are likely to be related to singing-like expressions of song, and feelings of need to speaking-like ones. As the infant vocally expresses affective states, the ones related to positive or playful feelings are more likely to sound similar to singing. As can be evidenced acoustically, the syllables of singing-like expressions have prevailing musical features (Stadler Elmer, 2020). For the analysis of both song and speech, syllables are the building blocks, yet they are differently organised in the two modes. Although the development of word formation differs across languages (e.g., Elsen & Schlipphak, 2015), it seems that the production of melodies is generally – and probably in all languages – far easier than the production of words. The song-before-speech hypothesis holds that children – given musico-linguistic stimulation – sing melodies before they produce words to talk (Rousseau, 1781; Vaneechoutte & Skoyles, 1998; von Humboldt, 1836). Hence, melodies can function as a catalyst for word production. Although the aim of this study was not to test the song-before-speech hypothesis, this issue inevitably needs to be considered, as I will show below.

This study focuses on 18-month-old Lynn's spontaneous singing with the aim of answering the following research questions: What does the child vocally express whilst producing a song? How does the child generate the impression of creating a song? What does her singing reveal about her cognitive understanding of singing? In order to answer these questions, I first conceptualise the components and rules that constitute a song *per se*, hence, an abstract model or exemplar. Second, a recorded vocalisation must be analysed using a methodology that takes into account the triple dilemma posed by the facts that a) singing has universal properties, b) singing has culture- and language-specific articulations, and c) singing *in situ* is likely to deviate from song models or rules. To describe 18-month-old Lynn's early singing in detail, I introduce the context in which it occurred. I then conceptualise song as grammatically guided before providing an overview of the research on early singing and arguing for micro-genetic analysis. In the main body of the article, I present the empirical analysis of Lynn's singing, and finally, discuss the results and answer the research questions.

Why should the elaboration of questions on early singing contribute to the state of the art of research in music education? First, description is one of the basic research activities in all sciences (Hoyningen-Huene, 2013). A reliable and valid description of the phenomenon we study provides the basis for scientific explanation and understanding. Moreover, the quality of communication about phenomena depends on the concepts we use and on the extent to which they are clarified and agreed upon. Often, we only perceive the things for which we already have a concept. Thus, our observations and descriptions of (musical) events depend largely on the conceptual knowledge available to us, and even more on the means we use to describe them by symbols (language, numbers, notation etc.) and icons. Figure 1 schematises activities involved in observation for gaining a systematic description of a phenomenon as a cycle comprised of observation, distinction, denotation and representation.

Second, in the literature, researchers only rarely observe and describe early singing in detail. Yet, describing a complex phenomenon with technical terms makes it possible to achieve a detailed conceptual framework and a thorough understanding. This claim is related to the self-understanding of music education as being a social practice and at the same time a scientific discipline. The emphasis on describing and understanding human actions and social practices is also known as phenomenological. However, for studying human actions, the intentions and the context are essential components that need to be taken into account.

The context of Lynn singing Twinkle, twinkle, little star

Lynn (18 months) had been exposed to *Twinkle, twinkle little star* in English, although she grew up in a predominantly German-speaking environment. The video recording of her spontaneous singing was given to me by her father with permission to examine it. He was in the same room
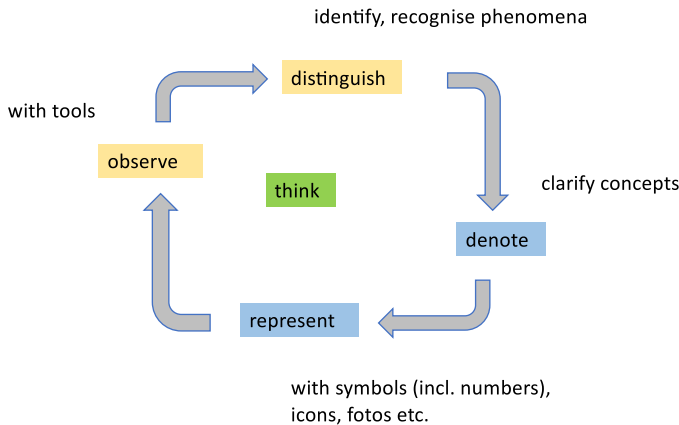
identify, recognise phenomena

with tools

distinguish

observe

think

clarify concepts

denote

represent

with symbols (incl. numbers),
icons, fotos etc.

Figure 1. Activities involved in the process of describing systematically a phenomenon.

with Lynn when he heard her spontaneous singing and reached for his camera to record 3 minutes and 31 seconds. Lynn and her older brother were sitting next to each other at a children's table. He was cutting paper with scissors while she was drawing with crayons and singing at the same time. She was singing to herself and obviously did not expect her father or brother to pay attention to her. However, when she noticed that her father was nearby – the video shows her looking briefly at the camera – she changed the syllables of her song from /ma – mi – ma/ to /ba – ba – ba/ and continued without interruption. To change her crayon, she twice climbed from her little chair onto the table to get another colour crayon from a box that was above the table, and then climbed down again. I mention this climbing because it affects the breathing pattern of her singing. Furthermore, she had difficulty climbing down from the table and finding her footing on the floor the second time. At that moment, she interrupted her singing and 'asked' her father for help by making unintelligible sounds. He told her to just slide down and let herself fall as she was already close to the ground. She did so and then continued singing and drawing.

By observing this scenario countless times, I always concentrated on the girl's singing and her movements. It was only when I was working on analysing this recording again after a long time that I noticed the following: on the one hand, she sang a clear melody identifiable as *Twinkle, twinkle, little star*, but on the other hand, when she had a problem climbing down, she could not form a single word in her communication with her father but used an unambiguous help-seeking prosody without recognisable words or protowords. Her parents told me that her language development was normal. Infants begin to learn phonetic categories of their language that provide the foundation of phonology, and word learning seems to be an enterprise that occurs during the second and third years, when children discover the meanings of words and other linguistic expressions (e.g., Swingley, 2009).

The present discrepancy between the clearly and identifiably sung melody and the yet unarticulated speech provides evidence of the singing-before-speaking hypothesis. The relationship between song and speech or music and language is a topic debated from evolutionary (e.g., Cross et al., 2013) and ontogenetic perspectives. The latter I will take up in the last section to discuss the relationship between melody and lyrics in songs in early childhood.

All in all, the video clearly shows a toddler singing a familiar melody with simple syllables (ma, mi, ba, ba). It is also visible that she sometimes synchronised her drawing movements with her singing. Remarkably, when she asked for help, she was not yet able to communicate her needs by articulated words. At the age of 18 months, Lynn has hardly developed any word formation techniques of German as her mother tongue (Elsen & Schlipphak, 2015). However, her singing suggests that someone must have shared this song with her previously, and we can also speculate that she probably had strong impressions of *Twinkle, twinkle, little star* that led to the phenomenon 'earworm' or haunting melody in her mind that urged expression.

The rule-governed making-up of a song: a grammatical conception

What allows us to identify a child's intention to sing a certain song, such as *Twinkle, twinkle, little star*? For the analysis of vocalisation as song, we need a theoretical model of the components and rules that make it up. A song can be defined as the product of a vocal act, as an event following rules to vocally create a coherent whole or Gestalt, and as such, a song can also be a mental image that can be recollected and collectively shared. Song can be invented *ad hoc*, or reproduced on the basis of individual or collective memory or of written sources. Voiced sounds typically are vowels (V) accompanied by consonants (C), thus forming syllables (VC, CV, CVC) that are the basic units of song and speech. For syllables to be perceived as singing-like, the vowel has to be elongated, whereas the speaking-like sounds tend to be short (Stadler Elmer, 2012a; 2021). The syllables as the building blocks of song and speech are organised differently for the two modes. In song, the prolonged syllables of vowels allow the producer to modulate the pitches, and thus, to form melodies. In both modes, song and speech, the successive syllables are organised according to language-specific and genre-specific stress patterns. Stressed syllables can be longer, higher, or louder than unstressed ones, since accent or stress is a *relational* feature of a sung or spoken syllable. It can be defined by the contrast to its adjacent syllables in terms of contrasting intensity, duration, pitch, or a combination of these (Stadler Elmer, 2021; Hall, 2000; van der Hulst, 2014). Thus, accented or stressed syllables do not have a unique phonetic, musical, or acoustic property, but exist only in contrast to adjacent syllables with respect to intensity, duration, or pitch. This abstract view on syllables is not at all trivial because it provides the technical terms to differentiate early singing from early speaking, and it also explicates the syllable as the building block for song and speech. The properties of syllables for song concern the prolongation that brings out pitch as a modulable feature to make melodies. In addition, syllables in song are periodically organised by binary or ternary metre, whereas speech is organised according to language-specific stress patterns. The rules for organising the features of syllables in song can be summarised as a set of rules to construct melody, lyrics, and their timing. Thus, song can be conceptualised as rule-governed in terms of a grammar, with songs for children presenting the most elementary musico-linguistic genre (Stadler Elmer, 2021). This *Grammar of Songs for Children*[1] allows one to produce an unlimited number of songs or sentences. Likewise, it can be used to analyse songs to study how the tonal, linguistic, and timing rules are used to generate a new *ad hoc* song, intended as an emulated reproduction or as an invention, or a combination of the two, and how the producer expresses her or his understanding of song as a coherent and rule-governed musico-linguistic vocal form.

If we conceive of song as organised by grammar, it follows that children acquire the musico-linguistic rules of their culture, and they begin to do so by exploring the properties of syllables such as intensity, the pitch dimension, vowels, consonants, stress patterns, and their infinite combinations. In the phase of canonical babbling, beginning between 5 and 10 months, the reduplicated syllables do not yet allow observers to identify them as either singing- or speech-like utterances, since they are not yet semantised (Merker, 2006; Stadler Elmer, 2020). When the child's affect state signals ease and playfulness, the produced syllables might appear more as singing-like than as protowords. As mentioned earlier, a playful mood is a necessary condition for a child to explore the vocal expressive potential toward generating singing-like forms, whereas, on the contrary, vocalisations signalling basic needs are likely to be interpreted as precursors of the child's speech.

In the present case study of Lynn, the research question concerns her manner of generating the well-known song *Twinkle, twinkle, little star*. How does she construct syllables and organise them into a coherent whole? What allows us to identify her intention to sing this song? What difficulties does she face while producing this song? These questions require a detailed description of the child's vocal production with a musico-linguistic vocabulary to make explicit the elements and rules in her vocalisation in connection with the *Grammar of Songs for Children*.

Early singing and micro-genetic analysis

In researching musical development, 'one of the problems first encountered is the distinction between music and speech. The earliest kind of communication is neither, but contains elements of both' (Nettl, 1956: 30). Micro-genetic methods are used to analyse a person's strategies to solve a problem or to create something new in minute details (e.g., Wagoner, 2009; Werner, 1956). In the domain of early singing development, micro-genetic methods are only rarely used to study vocal productions, and also include acoustic analysis (e.g., Papoušek & Papoušek, 1981; Stadler Elmer, 2012a). The literature addressing early singing development is mainly interested in evidencing beneficial effects of early songs or singing, or in factors that might foster singing development (e.g., Cohen, 2021). Most studies on singing use ordinal scales for rating pitch accuracy, and other related criteria (overview by Svec, 2020). Young children's singing is also verbally described (e.g., Young, 2006; Forrester, 2009), or musically transcribed (e.g., Werner, 1917; McKernon, 1979).

As mentioned in the introduction, songs by children under 36 months of age are characterised similarly, but the beginnings, development and weighting of the dimensions are described and explained differently. From a science-historical perspective, it is remarkable that the linguistic branch of vocal development not only has a more intensive research tradition, but also has available a detailed phonological and morphological vocabulary to conceptualise and describe vocal utterances and vocal development. Although child phonology carries a big blind spot as far as musical aspects of vocal development are concerned, I notice that research on the onset of speech is far more differentiated and elaborated than that on early singing. Consequently, there is much to gain from this closely related yet also interdisciplinary field. For instance, Menn (2013) describes the transition from canonical babbling into the onset of speech by three definable utterances, i.e., sound-play, protowords and modulated babble. Here, the overlaps in the transition into singing are obvious. Yet, it is interesting that by tradition, in the classical child phonology literature, the terms 'song' and 'singing' do not yet exist. This means an exclusion of the more musical aspects. On the other hand, the research literature on early song, by its exaggerated focus on pitch accuracy or in-tune singing, tends to neglect other criteria of song such as stress patterns.

In the literature on musical development, however, authors largely agree that the course of development is necessarily stimulated by social interaction and depends on neurological growth as well as on the children's actions, feelings and interests. In my opinion, progress in scientific research can be achieved by elaborating the technical terms used to describe and communicate the phenomena in question. Clear and appropriate communication also serves music educators to better understand the relation between the content they teach and the children's learning.

Using microanalysis is nothing else than aiming to describe human actions as temporal phenomena that are still only rarely studied. While microanalysis also includes dialogue analyses, detailed descriptions of task-related learning outcomes or sequences, the use of the term micro-genesis is more specifically linked to the tradition of Heinz Werner, Lev Vygotskij, Aleksandr Luria and others (see Wagoner, 2009). The micro-genetic method has a long tradition. According to Catan (1986), the original ideas go back to the concept of 'actual genesis' by Sander (born 1889), a German Gestalt psychologist of the Leipzig School. But it was mainly Heinz Werner and the socio-historically oriented Soviet School of Luria and Vygotskij (e.g., 1978, 'Problems of Method': 58ff.), who further developed and applied the method in connection with theoretical principles.

The aim of the micro-genetic method is to understand the processes of the emergence of higher mental forms. Therefore, how the terms microanalysis and micro-genesis are used is not irrelevant (cf. also Catan, 1986, who gives examples of inadequate usage). While microanalysis stands as the general umbrella term for detailed descriptions of events that take place over time, micro-genesis additionally involves questions about the direction of changes with regard to higher mental forms (Stadler Elmer, 2002: 218). In the present case of Lynn and her singing of *Twinkle, twinkle,*
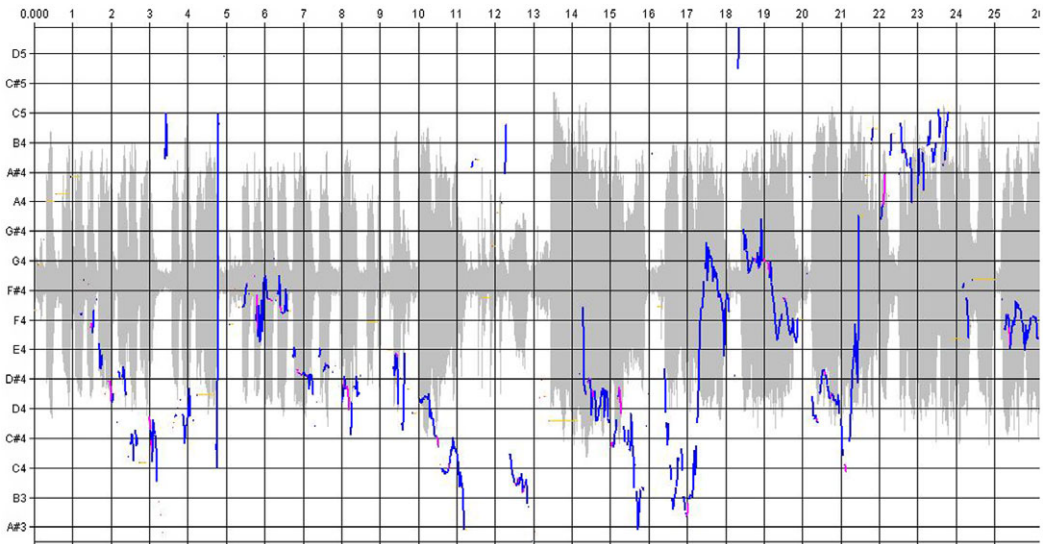
Figure 2. The acoustic analysis provided by the Pitch Analyzer yields this overview on *Twinkle, twinkle, little star* sung by Lynn (18 months). Here, the results calculated by the min-max algorithm are shown. The overview is the starting point for further analyses, first of all for identifying the segments, then by zooming in to analyse the syllables and then their properties such as pitch and time.

little star, I am interested in how she vocally expresses her way of following rules and what insights we can gain from this about early forms of shaping songs as a rule-governed whole.

Analyses of song and singing

In this section, I present an extract of Lynn's singing *Twinkle, twinkle, little star* in comparison to the song model. To achieve a micro-genetic analysis, I used the method with two computer programs we developed to analyse singing that does not have vibrato (Stadler Elmer & Elmer, 2000). The Pitch Analyzer yields acoustic features, whereas the Notation Viewer yields graphics on the basis of various measures that are given by using the Pitch Analyzer. Figure 2 gives the overview of Lynn's singing provided by the Pitch Analyzer. For the detailed analysis, the phrases have to be identified, and thereafter, each syllable needs to be specified with numbers representing properties of pitch and time, respectively. This dataset is then used by the Notation Viewer to provide graphic representations such as given in Figure 3.

Figure 3 represents the most relevant results of the micro-genetic analysis of Lynn's singing. Although we see melodic deviations from the song model by comparing the thin line of the model with the dots representing Lynn's sung version, we can clearly identify her intention to produce *Twinkle, twinkle, little star.* Yet, and this is most important, she only produces the melody or the overall melodic contour, and by no means any syllable of the lyrics. Instead of singing the target lyrics or some of its words or syllables, she entirely ignored the original English lyrics, and she substituted them by producing simple syllables such as /ba/, /da/, /me/, /mae/, /ma/, /mi/ in a regular trochaic pattern. We see that she varied the initial consonants and the vowels, but does not yet form bi-syllabic words, except /mami/ with the proper stressed first syllable, the German colloquial word for 'mother'. Lynn's identifiable emulation of the target melody is remarkable, since this confirms previous findings on early singing showing the same phenomenon, that is, to sing primarily the melody while neglecting the lyrics (Stadler Elmer, 2012b) or even ignoring them completely. In a nutshell, of all the grammatical rules, in her singing, Lynn bypassed the ones
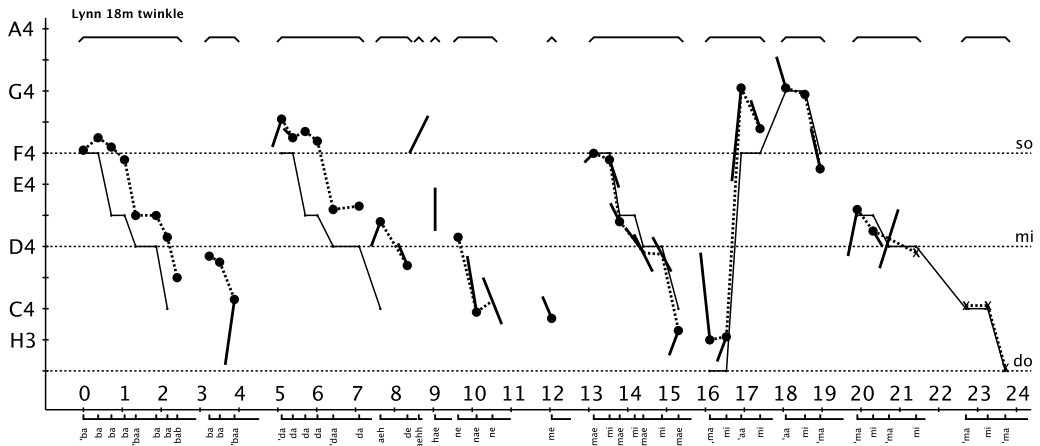
Figure 3. This comparative analysis starts with phrase B1 of *Twinkle, twinkle, little star*, as marked in Figure 4, followed by B2, then by another B (to become B3), yet distorted due to her climbing up, and again by a B (B4), and then she sung two phrases corresponding to A2. The Notation Viewer yields this graphic representation with the x-axis representing time in seconds, and the y-axis representing pitch as continuum that is subdivided into the pitch categories of our Western tonal system. The model song *Twinkle, twinkle, little star* is shown as a thin, continuous line, and Lynn's sung version is shown by the syllables and duration at the bottom, and by the dots representing the quality and tonal position of the pitches which are connected by the dotted line. The symbol x means pitch estimation at the basis of unclear signal.
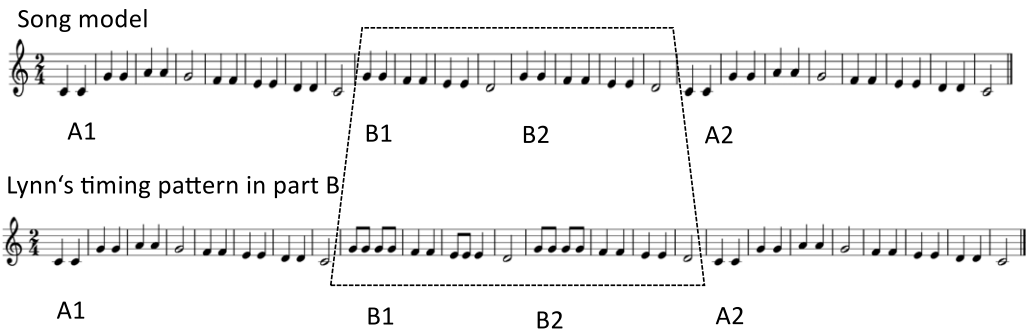

Figure 4. Lynn spontaneously varies the temporal rules at the beginning of part B1 and B2. She achieves this by subdividing the duration of two given syllables by reduplicating them to four syllables in a well-formed manner to keep the temporal framework.

concerning the text, yet, as we will see, she followed the temporal and tonal rules to produce *Twinkle, twinkle, little star*.

With regard to comparing the temporal rules of the song and of Lynn's version, a second occurrence is worth noting: her climbing up some furniture whilst singing explains her breathing pattern at secs. 9 to 12 (see Figure 3) when producing part B that became B3. It appears scattered and incoherent. In general, a song is segmented through breathing, and from a grammatical view, breathing creates phrases and hence, it should occur between phrases. Lynn's breathing or segmenting is represented by the brackets at the top of Figure 3 below the title. We see that her segments are smaller than the ones in the model. Yet, despite her smaller segments, their timing always follows the boundaries of the song's phrases, that is, she began and ended her segments according to the beginnings and endings of phrases of the model. Except for her part B3, her segmentation is far from random, as it does show signs of regular grouping. According to

the *Grammar of Songs for Children* (Stadler Elmer, 2015, 2021), phrases are units at a high hierarchical level whose boundaries have prominent melodic and linguistic markers.

Most remarkable is Lynn's subdivision of the duration of the syllables in the beginning of part B1 – at second 0 – and of part B2 – at second 5 – in comparison to the song model. In both cases, Lynn doubled the syllables by subdividing the duration of the one syllable. Hence, her adding syllables at the beginning of B1 and B2 did not prolong these phrases, but conserved the overall duration and thus the song's temporal symmetry. Thus, Lynn invented a variation regarding the temporal rules without violating them. Figure 3 represents this timing rule variation repeated twice, in B1 and B2. This reduplication of syllables while maintaining the following of the temporal rules can be interpreted as striving for formal symmetry and also as grasping rule-following to bring out well-formedness.

With regard to the tonal rules of the *Grammar of Songs for Children*, we observe the following. She produces the melodic contours of the target song. She repeats and varies the complex pitch pattern of part B. She produces a clearly identifiable melody, although often not at a stable or recurrent pitch level. An example for unstable pitch categories is visible in Figure 3 in her syllable duplications in B1 and B2. Presumably intended as repetitions, the pitches deviate within a range of approximately 40–50 cents.

Conclusions

The micro-genetic analysis of an extract of spontaneous singing by the 18-month-old Lynn shows how she produced the well-known song *Twinkle, twinkle, little star*. The detailed descriptions, the empirical contextualisation, the resemblance with previously reported analysis of early singing and the connections with theoretical considerations inspired me to sum up my conclusion with three points.

Contrary to the common misconception about a developmental sequence of singing of lyrics, rhythm, melodic contour, intervals, in the present case study, Lynn primarily produced a recognisable melody and hardly any textual properties. Microanalyses of early singing do not support the speech-before-singing hypotheses, but the singing-before-speech one. It may be that the misconception of the linguistic dominance of song development is oriented towards children who grow up in environments with little musical stimulation, and who therefore find the linguistic aspect of songs easier to access. But in any case, this false generalisation can no longer be maintained.

The main argument in favour of the singing-before-speaking hypothesis is based on the cognitive requirements. In ontogeny, the formation of words is much more demanding than the formation of melodies. The present case with Lynn shows that she manages to produce a well-known melody in such a way that it is easily identifiable. She is not yet able to do the same with words, and this was evident when she was unable to articulate her climbing-down problem to her father in the same context.

This case study shows clearly how Lynn expresses her belonging to Western culture by reproducing a widely-known melody at a very early age. Her manner of following vocal rules can be interpreted as the expression of an 'earworm', a mental image that has been imprinted on her and repeatedly urges her to express it. Nevertheless, this is a sign of cultural belonging. The phenomenon of a mentally imprinted melody, as it can show early on, is as a yet unexplored phenomenon since it is difficult to observe systematically. This mental phenomenon might also be linked to the positive effect associated with songs, which allows people to memorise the shared social context and linked feelings of belonging, and to easily repeat the related heard patterns and the own vocal production.

The higher mental processes expressed in this early singing involve the capacity to learn rules vocally (Merker, 2012) and to begin to grasp their generativity. As Lynn's case shows, she changed

the timing rule of one phrase (B1, B2) in a well-formed way in her singing. She was handling the organisation of her vocal sounds in a well-formed manner, yet, she seemed not yet able to reflect on her actions. From a developmental viewpoint, she already produced complex vocal patterns that are cultural signs she acquired through hearing and sharing them with significant others. According to Vygotskij (2022), she uses a cultural sign system that she has already begun to integrate into her imagination.

Note

1 A detailed explication of *The Grammar of Songs for Children* was published by Stadler Elmer (2015, 2021).

References

CATAN, L. (1986). The dynamic display of process: Historical development and contemporary uses of the microgenetic method. *Human Development*, **29**, 252–263.

COHEN, A.J. (2021). Through singing to music across the life course. In A. Creech, D. Hodges & S. Hallam (eds.), *The Routledge International Handbook of Music Psychology in Education and the Community* (pp. 201–216). Abingdon: Routledge.

CROSS, I., et al. (2013). Culture and evolution. In M. A. Arbib (ed.), *Language, Music and the Brain* (pp. 540–562). Cambridge, MA: MIT Press.

ELSEN, H. & SCHLIPPHAK, K. (2015). Word formation in first-language acquisition. In P.O. Müller, I. Ohnheiser, S. Olsen & F. Rainer (eds.), *Word Formation. An International Handbook of the Languages of Europe*, (vol. **3**, pp. 2117–2137). Berlin: de Gruyter.

FORRESTER, M.A. (2009). Emerging musicality during the pre-school years: a case study of one child. *Psychology of Music*, **38**, 131–158. doi: 10.1177/0305735609339452

HALL, T.A. (2000). *Phonologie: Eine Einführung*. Berlin: de Gruyter.

HOYNINGEN-HUENE, P. (2013). *Systematicity: The Nature of Science*. Oxford: Oxford University Press.

MCKERNON, P.E. (1979). The development of first songs in young children. In D. Wolf (ed.), *New Directions for Child Development* (pp. 43–58). San Francisco, CA: Jossey-Bass Inc.

MENN, L. (2013). Development of articulatory, phonetic, and phonological capabilities. In M.M. Vihman & T. Keren-Portnoy (eds.), *The Emergence of Phonology: Whole-word Approaches and Cross-linguistic Evidence* (pp. 168–214). Cambridge, UK: Cambridge University Press.

MERKER, B. (2006). Layered constraints on the multiple creativities of music. In I. Deliège & G. Wiggins (eds.), *Musical Creativity: Multidisciplinary Research in Theory and Practice* (pp. 25–41). Hove: Psychology Press.

MERKER, B. (2012). The vocal learning constellation: imitation, ritual culture, encephalization. In N. Bannan (ed.), *Music, language, and human evolution* (pp. 215–260). Oxford: Oxford University Press.

NATHANI, S., ERTMER, D.J. & STARK, R.E. (2006). Assessing vocal development in infants and toddlers. *Clinical Linguistics & Phonetics*, **20**(5), 35–369.

NETTL, B. (1956). Notes on infant musical development. *The Musical Quarterly*, **42**(1), 28–34.

PAPOUŠEK, M. & PAPOUŠEK, H. (1981). Musical elements in the infant's vocalization: Their significance for communication, cognition, and creativity. *Advances in Infancy Research*, **1**, 163–224.

ROUSSEAU, J.-J. (1781). Essai sur l'origine des langues. Collection: "Les classiques des sciences sociales" [Essay on the origin of languages. Collection: "Les classiques des sciences sociales".]. Genève. http://classiques.uqac.ca/classiques/Rousseau_jj/essai_origine_des_langues/essai_origine_langues.html (accessed June 2022)

STADLER ELMER, S. & ELMER, F.-J. (2000). A new method for analyzing and representing singing. *Psychology of Music*, **28**(1), 23–42.

STADLER ELMER, S. (2002). *Kinder singen Lieder: Über den Prozess der Kultivierung des vokalen Ausdrucks [Children sing songs. On the process of cultivating the vocal expression.]*. Münster: Waxmann.

STADLER ELMER, S. (2011). Human singing: towards a developmental theory. *Psychomusicology: Music, Mind and Brain*, **21**(1–2), 13–30.

STADLER ELMER, S. (2012a). Characteristics of early productive musicality. *Problems of Music Pedagogy*, **10-11**, 9–23.

STADLER ELMER, S. (2012b). Structural aspects of early song singing. In A. Baldassarre (ed.), *Musik – Raum – Akkord – Bild [Music – Space – Chord – Image]. Festschrift für [Memorial Publication for] Dorothea Baumann* (pp. 765–782). Bern: Peter Lang Verlag.

STADLER ELMER, S. (2015). *Kind und Musik. [Child and Music].* Heidelberg: Springer.

STADLER ELMER, S. (2020). From canonical babbling to early song singing and its relations to speech. In F. Russo, B. Ilari & A. J. Cohen (eds.), *The Routledge Companion to Interdisciplinary Studies in Singing.* (vol. **1**, (pp. 25–38). New York: Routledge.

STADLER ELMER, S. (2021). Song transmission as a formal cultural practice. *Frontiers in Psychology*, **12**, 654282.

SVEC, C.L. (2020). A meta-analytic perspective on the development of singing in children. In F. Russo, B. Ilari & A. J. Cohen (eds.), *The Routledge Companion to Interdisciplinary Studies in Singing.* (vol. **1**, pp. 304–321). New York: Routledge.

SWINGLEY, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B*, **364**, 3617–3632.

VANEECHOUTTE, M. & SKOYLES, J.R. (1998). The Memetic Origin of Language: modern humans as musical primates. *Journal of Memetics*, **2**, 84–117.

VAN DER HULST, H. (2014). The study of word accent and stress: past, present, and future. In H van der Hulst (ed.), *Word Stress – Theoretical and Typological Issues* (pp. 3–55). Cambridge, UK: Cambridge University Press.

VON HUMBOLDT, W. (1836). Über die Verschiedenheit des menschlichen Sprachbaus und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts. Druckerei der Königlichen Akademie der Wissenschaften [On language: The diversity of human language structure and its influence on the mental development of mankind. Translated by P. Heath & H. Aarsleff (1988). Cambridge: Cambridge University Press.].

VYGOTSKIJ, L.S. (1978). *Mind in Society. The Development of Higher Psychological Processes.* Cambridge, MA: Harvard University Press.

VYGOTSKIJ, L.S. (2022). *Imagination: Textes choisis: avec des commentaires et des essais sur l'imagination dans l'oeuvre de Vygotskij [Imagination: Selected Texts: with commentaries and essays on imagination in the work of Vygotskij.].* Bern: Peter Lang.

WAGONER, B. (2009). The experimental methodology of constructive microgenesis. In J. Valsiner, P. Molenaar, M. Lyra & N. Chaudhary (eds.), *Dynamic Process Methodology in the Social and Developmental Sciences* (pp. 99–121). New York: Springer.

WERNER, H. (1917). Die melodische Erfindung im frühen Kindesalter [The melodic invention in early childhood.]. *Bericht der Kaiserlichen Akademie*, **182**, 1–100.

WERNER, H. (1956). Microgenesis and aphasia. *Journal of Abnormal Social Psychology*, **52**, 347–353.

YOUNG, S. (2006). Seen but not heard: young children, improvised singing and educational practice. *Contemporary Issues in Early Childhood*, **7**(3), 270–280.