From canonical babbling to early singing and its relation to the beginnings of speech.

Stadler Elmer, Stefanie

Stadler Elmer, S. (2020). From canonical babbling to early singing and its relation to the beginnings of speech. In F. Russo, B. Ilari, & A.J. Cohen (eds.), The Routledge Companion to interdisciplinary studies in singing. Vol. I. (pp. 25-38). New York: Routledge.

2

FROM CANONICAL BABBLING TO EARLY SINGING AND ITS RELATION TO THE BEGINNINGS OF SPEECH

Stefanie Stadler Elmer

Introduction

How does early song emerge, and how is it organized? It is remarkable that the ontogeny and phylogeny of human vocalization have been studied mostly with a focus on speech and language (e.g., Ackermann, Hage, & Ziegler, 2014; Francis, 2012; Oller, 2014). Cross et al. (2013) explored the relationships between language and music in evolutionary and cultural context. They considered six possible ways of conceiving of the evolutionary relationships between language and music in terms of common or different origins and of emergent pathways, and they suggested the capacity for vocal learning to be the most compelling candidate for the origin of language and music. They argue, however, that it is very difficult to reconstruct evolutionary developments because of a lack of knowledge of what was considered language and music in earlier cultures.

In this chapter, I address these issues from the perspective of the ontogenetic emergence of song in infancy. In so doing, it is crucial to elaborate criteria for distinguishing speaking-like and singing-like utterances, because many vocal phenomena – apart from affective-physiological expressions such as laughing and crying – exhibit both linguistic and musical features, as in poetry, songs with lyrics, and intuitive communication with infants (e.g., Papoušek, 1996). Conceptions regarding vocal modes and their social functions – speech, song, and intermediate forms – vary among cultures and researchers (List, 1963). Not only production but also perception of vocalization can be ambivalent as demonstrated by the speech-to-song illusion, that is, the perception of an artificially repeated spoken sentence as a sung song (Deutsch, Lapidis, & Henthorn, 2008).

Positions with respect to the relationship of song and speech in ontogeny – as subsets of music and language – can be summarized as follows: (1) treatment of song and speech as two separate and independent domains; (2) *linguistic primacy* or the assumption that singing develops out of speaking, for instance, that children would first reproduce the lyrics (for references see Stadler Elmer, 2011); (3) that speaking develops out of singing; and (4) that both develop from a common "proto-faculty" (3 and 4 may be combined, see Levman, 1992). Building on previous theoretical and empirical research bearing on the third and fourth positions (Stadler Elmer, 2012,

2015), this chapter deals with conceptual issues related to the following questions: How does song begin? How can we understand the process by which infants start to differentiate between early forms of speaking and singing?

Posing the question in this manner involves assumptions about development in general and about the nature of human vocalization that need explication, including:

- Song and speech have a common history and common biological roots in vocalization (e.g., Merker, 2012).
- Singing and speaking are acts with different functions in social life (e.g., Lomax, 1977), and their differentiation begins during the first year of life (Stadler Elmer, 2002, 2015).
- The content of both are products of intergenerational cultural transmission without which there is neither language nor music (Merker, Morley, & Zuidema, 2015).
- Structurally, song and speech are generative cultural systems (Merker, 2002), that, in short, are based on the *particulate principle*: A finite set of distinct elements selected from a continuum (e.g., vocal expression potential) can be combined to ever newer and richer patterns (Abler, 1989; Merker, 2002).
- Both song and speech employ the syllable (vowel with optional consonants) as a basic unit, but they differ in the way syllables are used to form larger patterns. In speech, the phoneme is the ultimate element of pattern formation, whereas in singing the syllable remains the basic unit, subject to constraints of pitch and duration, remaining either unsemanticized or combined into words (Stadler Elmer, 2002, 2015).
- Primordially, singing expresses positive affective states such as well-being, ease, or playfulness. Song can also induce affective states; it can signal cultural belonging as well as exclusion, and it forms a mainstay of ritual culture world-wide. In contrast, speech acts typically make a statement, to command or to request, and express an individual speaker's intentions. They cover a wide range of functions that lie beyond the scope of this chapter.

These characteristics and assumptions have a bearing on how vocal development and early singing are conceptualized and studied. The issues are inextricably linked to questions of the origins of music, language, and even of humans. Adopting the structure-genetic perspective introduced by Piaget, I focus on the genesis of vocal structures and emergent cultural forms, feelings, and consciousness in human development (Stadler Elmer, 2002, 2015). Postulating canonical babbling to be the precursor not only of speech but of song as well, I argue that syllables are the building blocks of both, yet they are differently formed and organized. The aim of this chapter is, accordingly, to outline and empirically illustrate the role of syllables in song and speech, and to discuss how the infant and toddler starts applying the contrasting rules for the two domains.

Vocal Learning and Assumptions about Development

Human song singing as well as speaking is a result of vocal development that literally begins with crying as the first vocal expression at birth. Any identification of the earliest forms of singing or speaking must be based on norms assigned to the phenomena, and scrutinized in the light of biological, functional, and structural evidence. Furthermore, developmental changes in the structural organization of behavior – in this case vocalizing – are directional processes with goals and interim states of competence within a specific culture. Vocal development can be said to aim at becoming at least a speaker of the native language and a singer of some traditional children's songs.

In the present context, it makes sense to restrict the pivotal analysis of early singing to song as a traditional cultural practice in the form of children's songs. The latter are usually considered

simple as well as appropriate for introducing children to the musical culture, beginning with lullabies and actions songs. Having started as an infant to participate in songs, over the years, songs become automatized through practice, and – as with all other experiences in early childhood – this learning process is hardly ever available to recollection, let alone conscious scrutiny. We remember as little of our first song learning as of our learning to speak. This early phase must be reconstructed through empirical research. From that perspective songs are seen to be far from simple, but densely structured vocal patterns in which melody and meter are combined with lyrics. It goes without saying that the rules that govern this patterning do not need to be consciously known to the performer: even expert singers or musicians do not necessarily have explicit knowledge of the formal rules that govern their singing. Much early learning takes place implicitly, through procedural memory, and even adult performance relies largely on intuition.

Infants and toddlers are highly attentive to the vocal sounds of song and speech and eagerly learn to reproduce and recombine them by repeating, varying, and exploring, and by implicitly extracting the underlying rules. The overall trajectory of this gradual development can be characterized as the infant and toddler gaining increasing control over two domains: (a) the fine-tuning of the match between heard and produced sounds, like phonation and articulators (e.g., tongue, lips, mandible, velum, pharyngeal constrictors, etc.); and (b) the expression of affect by adapting to its socially mediated forms. The acculturation of vocal expression is underpinned by affect regulation that begins by co-regulation through parental care with meaning-making dialogues that transform co-constructed vocal signs into internalized affective self-regulation. Vocal sounds become shaped and coded into meaning fields of normative signs toward speech and song, or in general, language and music. These two normative and generative systems make it possible to create and maintain the relationship among human beings. Presumably, the human vocal expressive potential is the most powerful and primordial means for the regulation of affective states and for creating signs and systems. The first few years of life are most formative for this acculturative process of vocal learning and becoming a member of the culture through a process that is not accessible to conscious recollection.

To reconstruct this implicit process, researchers need explicit principles and general knowledge regarding the structures and functions it involves. From a structure-genetic view, the developmental pathways are expected to be neither arbitrary nor uniform, but to follow some structuring regularities that researchers are only slowly starting to uncover. Though regularities are likely to be cast in terms of developmental stages (Lourenço, 2016), the core issue of early vocal development is the adaptation of the individual's vocal expression to the social surrounding – primarily to parents, caregivers, and siblings – and thereby the acquisition of the cultural rules or conventions on how to use one's voice for communication and for participating in the cultural practices related to song and speech.

From an evolutionary perspective, singing is likely to be the first form of human music making – a truism as well for the ontogeny. In our evolutionary history, song might have preceded language and combined with dance as a shared social activity amounting to the first form of the human arts (Merker et al., 2015). Likewise, for ontogeny, many scholars adhere to the song-before-speech-hypotheses (see Stadler Elmer, 2012). Humans are the only primates in possession of the plastic vocal ontogeny technically known as vocal production learning (Janik & Slater, 1997), that is, the capacity to use feedback from one's own voice to achieve a match between heard sounds and vocal production (Konishi, 2004; Nottebohm, 1972). This evolutionary novelty is shared with certain songbirds and with a few mammals such as cetaceans. Being an absolute requirement for both human song and speech (and for little else in our behavior), this unique capacity must underpin any phylogenetic and ontogenetic theorizing on the origins of language and music.

Historic writings on this topic by Herder (1772), Rousseau (1781), Darwin (1871), von Humboldt (1836), and others document the above-mentioned third and fourth positions on the singing origin of speech and language, whereas other historic sources also state the second position, namely speech as the origin of music (Spencer, 1857). Unrelated and contemporary, the *linguistic primacy* position (see above) holds that in both song acquisition and singing development, words would appear first and speech-like chanting would precede the singing of songs proper.

Recently an attempt has been made to develop a framework of basic facts and principles bearing on the phylogeny and ontogeny of language and music (Merker et al., 2015). In this framework, generativity (Merker, 2002), cultural transmission, and vocal learning are shared by song and speech, while the motivation to learn songs (Merker, 2005) and the capacity to entrain to an isochronous pulse – another human capacity rarely met with in the animal world – is specific to music. The latter capacity is the basis for the time-keeping movements of dance and song. Some of these principles, particularly as they apply to the intergenerational transmission of ritual culture, can fruitfully be applied to the reconstruction of the vocal ontogeny leading to speech and song (e.g., Eckerdal & Merker, 2009).

Language and Music as Generative Systems

To explore the proposal that singing is the earliest musical expression and even the candidate precursor of speaking, we must scrutinize the distinct and common features of language and music (for details see Stadler Elmer, 2015). Having restricted our scope to spoken language, sung music, and to children's song as the primary introductory cultural practice combining melody and lyrics, I assume the syllable to be the common unit of both singing and speaking. A syllable consists of a vowel (V), or a combination of a vowel and consonants (C), for example, CVC, VC, CV, CVCC. Prolonging a vowel easily creates the impression of singing because pitch becomes salient and easy to modulate.

Following von Humboldt (1836) and his important work on the general nature of languages, and Abler's (1989) concept of self-diversifying systems, Merker (2002) proposed that human music, like language, is a generative system. A generative system combines discrete elements – "particulates" that "do not blend by averaging on combining" (Merker, 2002, p. 4) – into composite patterns of boundless variety. In music, the particulate elements are obtained by discretizing the continua of pitch and time, yielding notes of distinct pitches to which a melody can return (such as the seven notes of a diatonic scale) and distinct durations (Merker, 2015). When these durations are related by whole integer proportions and combined in cycles and stress patterns, they supply the elements for the rhythms of all metric music.

In languages, the discretized elements are the phonemes, a select sample of the phonatory capacities of the human vocal apparatus specific to each language. These elements are combined into syllables and words, and these in turn to sentences in boundless variety. In so doing, each language uses specific rules for word formation on the basis of metrically accentuating the syllables. An accent or stress is a relational feature of a sung or spoken syllable defined by the contrast to its adjacent syllables in terms of contrasting intensity, duration, pitch, or a combination of these (Hall, 2000). Historically, the Greek and Roman philosophers, while construing grammars, first outlined phonological rules, before syntax, and created notions such as syllable and prosody, whereby applying terms from music theory to conceptualize accents, their features and metric rules (Schreiner, 1954).

In song, the lyrics are typically bound to a metrically organized melody, while in recited poetry they are bound to metric patterns without melody. This led Aristotle to suggest that poetry is positioned between song and speech, a view that has prevailed. In children's and folk songs, and many art songs, poetic language provides not only a periodic meter but also rhymes

Table 2.1 Overview of the common features manifested in the syllables produced as singing, reciting, and speaking, but organized differently by the generative principle of discretizing a continuum and applying combinatory rules

	Syllables as units: common features – differently organized			
	Continuous dimension	*Discrete categories*	*Accents (relational)*	*Characteristic combinations*
Singing	Intensity	Pitch, duration	Periodic	Discrete pitches, proportioned durations, periodic accents, repetition, syllables may be non-semantic, prolonged vowels
Reciting, chanting	Intensity, pitch	Duration, phonemes	Periodic (e.g., trochee, iambic, dactyl)	Proportioned durations, periodic accents, continuous intonation, syllable is a word or part of a word – semantic
Speaking	Intensity, pitch	Phonemes (pitch in tonal languages)	Language-specific rules	Short vowels, continuous intonation, syllable is a word or part of a word, polysyllabic words follow language-specific accentuation rules, semantic

that mark the ending of a verse or phrase. Table 2.1 gives an overview of the common features of language and music – intensity, pitch, time, accents – and how the syllables, as the common basic unit, are differently organized in singing, reciting, and speaking due to different generative principles and combinatory rules applied to them.

These shared but differently organized features can be related to the biological roots connected with the human vocal learning capacity, and to generativity as an organizing principle (Merker, 2002). Speech, song, and poetry are seemingly autonomous cultural products of a developmental process which, at the outset, shows no such differentiation. How does a child start to differentiate them and to use the rules for singing or the ones for speaking? To situate early singing in this process, I refer to research on vocal development, on the affective underpinnings of vocalization, and also on rules that govern vocalization to become speech and song.

Vocal Development, Affects, and Ritual Culture

The neonate's first cry might be an expression of pain and shock. We hear the infant's cries mainly as an alarm signal. In the first few months, the infant differentiates crying and expresses negative, neutral, and positive affective states – also reflected in facial expressions – and starts to discover contingencies in the multimodal communication to which it is exposed, and to elicit reactions in others. For more than five decades, researchers have been modeling the stages in vocal development that infants and toddlers pass through while becoming talkers, but hardly ever singers (but see Papoušek & Papoušek, 1981). As already noted, vocalization was treated almost solely as a pre-linguistic utterance, for instance, by broadly categorizing vocalization as speech-like and non-speech-like (Oller, 2000). Rarely, researchers (e.g., Nathani, Ertmer, & Stark, 2006; Stark, 1980) include non-speech-like sounds as a potential source for foundational elements recruited later for speech. It is indeed still a key challenge to achieve consensual definitions of the various vocal types, improve understanding of their functional flexibility in infants, and to theorize about regularities in early vocal development leading to speech as well as to song.

Linguists typically view early vocal development as self-organized and as resulting in advances in vocal production, first, as emergent phonatory control allowing for "protophones" (Oller, 2014) – typically vowel-like, squeals (high-pitched sounds, often in falsetto), and growls (low-pitched or raucous sounds often in creaky voice) – followed by closant-vocant combinations (canonical babbling, see below) to mature adult-like syllables (see Vihman, 1996; Masataka, 2005 for overviews). Models of vocal development provide estimates of the ages at which vocalization types are expected to emerge. For the earliest phase, Stark (1980) proposed a model with five levels: Reflexive Sounds (0–2 months); Cooing and Laughter (2–4 months); Vocal Play (4–8 months); Reduplicated Babbling (8–10 months); and Nonreduplicated Babbling and First Words (10–14 months). Another early developing vocal feature is intonation or speech melody that refers to the patterning of pitch changes in utterances. Again, this feature tends to be treated exclusively as a component of language and as interacting with grammatical, pragmatic, and affective levels of language description (Snow & Balog, 2002; but see Wermke & Mende, 2009). The general belief that infants master most or all the ambient intonation patterns before producing first words can also be conceptualized as a singing vocalization.

One way to expand the dominant linguistic view on vocal development by inclusion of the emergence of singing is provided by Merker's (2005, 2009) delineation of three cultural layers as summarized in Table 2.2. Humans share instrumental behavior (e.g., food washing) with their close primate relatives, but add two additional layers of cultural tradition, namely ritual culture and language. At the very heart of the ritual culture is obligatory adherence to arbitrary (in the sense of conventional) form. In instrumental culture the achievement of the desired instrumental outcome (a utility of some kind) is foremost, while in ritual culture adherence to the correct form is all-important. The temporal structuring of ritual performance, typically featuring repetition with variation, can serve inter-individual behavioral and emotional synchrony through the anticipation of upcoming events it makes possible (Merker, 2009).

Note that singing meets all the criteria for being a ritual. Parents, caregivers, and siblings intuitively teach and support the infant to conform vocalization to the ambient culture, communication, and rituals (intergenerational transmission). While cultivating vocal expression, human

Table 2.2 Overview of Merker's (2005, 2009) postulation of three cultural layers – one shared with animals – and the ritual foundation of human cultures

Cultural layers	*Features*	*Examples*	*Intergenerational transmission teaching – learning*
Instrumental behavior	Outcome is fixed, clear goals or products, shared with animals	Food washing, termite extraction	Observation, discovery that behavior is means for achieving goal = knowledge
Ritual behavior	Execution, **performance**, procedure is fixed (no immediate outcome), arbitrary patterns, **aim**: well-formed or "correct" performance, **correct form**	Music, song singing, ceremonies, decorations, arts, eating with etiquettes, mother–infant interaction	Imitation, participation (informal-formal), deliberate instructions, learning of skills and rules, practicing
Languages	Generative, most powerful means to create and share meaning and symbols	6,000–7,000 languages	Participation in meaning-making, face-to-face learning (imitation, construction) of symbols and rules

infants not only learn to differentiate vocal performance toward speaking and singing, but also simultaneously learn to control and shape their affect, and learn to become a participant member of ritual cultural traditions. Rituals – including the singing of songs – are a cultural technique for shaping and transforming emotional states (Vygotsky, 1976).

Speech qualifies as ritual culture in formal terms, but by mapping form into meaning it has acquired formidable instrumental uses. As such it embodies an intentional act (Searle, 1969) to achieve co-operative goals by making statements, referring to objects and events, requesting and commanding. These powers derive from language as a highly potent symbol system for forming and expressing thoughts. It is of course possible to tell stories by singing, for instance in ballads and other traditions of binding words into sung verses to orally transmit narratives over generations. The ritual frame of song integrates several functional and structural aspects: Deliberate intention is less important, since singing is originally acted out as a social event focused on performing synchronized sounds and body movements according to collectively shared patterns. The attunement and synchronization elicit affective states that are also shared. The shared feeling of being and acting together by adhering to rules is the essence of ritual. Adherence to the ritual form perpetuates its pattern across the generations.

The universal existence of song rituals might be due to the fact that singing does not require any other means than a person with a functional voice, that is, the preconditions for participation are minimal. The intention to sing appears to be less voluntary than speech acts, as can be inferred from the phenomenon of catchy melodies also known as "ear worms" – a tune continually repeating in one's mind, reminiscent of a real experience, triggering its vocal expression. This very phenomenon might play a crucial role in the emergence of early singing in that it promotes repetition and memorization. This involuntary process functions as if to interiorize the heard sounds by imposing a mental schema, and might be an intrinsic aspect of the vocal learning of melodies.

As a formally patterned vocal act, singing provides a temporal framework that can be – in contrast to speech – performed either collectively or alone, and can be both repeated and varied without violating convention. Each time a song is repeated, it creates the illusion of repeating the previous event. In this way, songs may keep the past alive in the present and might function as a cultural tool to reduce uncertainty about the future (Valsiner, 2003). To know a song means to own a mental tool for reproducing physiological and affective states that serve as reminders of previous events or symbolize something felt in the past that will be available in the future.

The affect-regulating functions of singing are partly due to their dense musico-linguistic organization creating coherent units that allow repetitions and variations in a way not comparable to spoken sentences. These very features – repetition and variation – and the melodic and metric ornamentation of syllables and words make the singing of songs distinct from speech acts and justify its classification as a distinctive ritual cultural practice (e.g., Levman, 1992; Lomax, 1967; Merker, 2009).

How to Make a Song: The Grammar of Children's Songs

The term *song* is nowadays often used for a piece of any kind of music. Here I use the traditional sense of *song* to mean humanly voiced musical sounds. This covers a wide range of pieces from children's and folk songs to highly artistic and sophisticated ones. The earliest of these, addressed to infants and children, can be analyzed from two different perspectives: First, in terms of a song as an abstract and rule-governed model, and second, in terms of the act, that is, using the voice to produce sound patterns that conform to some minimal cultural rules. Whereas the model song typically represents a general and even abstract device collectively shared as a ritual in a culture, the act of singing is the epitome of an ephemeral event, yet complexly structured and resembling the general model. The rules governing the structures of both, the abstract song model as well as the act of

singing, need to be made explicit in order to reconstruct the child's acquisition. The rationale behind this paradigm is the assumption that singing develops essentially by learning the rules about how to generate songs in a well-formed manner and make meaning by acts of singing.

What are the principles and rules of such songs? Ethnomusicologist Nettl (2000, p. 469) describes songs as being "the world's simplest style", consisting of short phrases, repeated with small variations, and covering three or four different pitch categories in the range of a fifth. Yet, children's songs are not only melodic, but always include musico-linguistic features that are inherent in syllables. Although there is much research on commonalities and differences between music and language, they are only rarely studied in connection to song or their ontogenetic origins.

Inspired by Merker's (2002) outline of core principles of music, I have endeavored to formalize the rules of children's song as a grammar (for a full account, see Stadler Elmer, 2015, pp. 73–84). Here, I present the seven principles and only some examples of the 21 rules in order to illustrate how the musico-linguistic components are integrated into a hierarchical and coherent framework. These principles are:

1. Children's songs consist of lyrics (a verse or a poem) and a melody.
2. The two *generative systems* – music and language – are both involved, thus, inherently, yield a potentially infinite variety of songs.
3. The building block of both systems for making songs is the s*yllable* since it also contains musically relevant features such as pitch, duration, and accents.
4. Songs are *hierarchically organized*.
5. Lyrics and melody are relatively autonomous (see 2), forming parallel hierarchies. Ideally, lyrics and melody are well matched. If this is not possible, one or the other is subordinated.
6. The verse meter of the lyrics and the musical meter of the melody simultaneously rule the timing of the syllables. This may create tension.
7. Symmetries of temporal and sonorous forms aim at creating *well-formed wholes* or a gestalt.

The 21 rules concern the timing, pitch, and lyrics of children's songs (Stadler Elmer, 2015, pp. 73–84). Below are some examples of each:

Timing rules – examples

- Once the meter (measure – a pattern of periodic accents) is set, it is valid for the entire song. Children's songs do not change the measure or meter.
- Children's songs have an even number of measures, typically 4, 8, 10, 12, 14, and 16.

Tonal rules – examples

- Children's songs are in *major keys*, rarely in pentatonic scale or minor keys.
- The major key once chosen is maintained through the entire song. There is *no change of key*.
- Children's songs *begin* with one of the three tones of the tonic accord (*do, mi, sol*).
- The first measure of a children's song marks the key.

Rules concerning lyrics – examples

- The lyrics are formed in poetic language. That means it has a *meter*, and *verse lines* that end with *rhymes* (pair and cross rhymes).
- The meter defines the periodic accents, and the verse lines are defined by the number of syllables.
- The verse meter (trochee, iambic, dactyl, anapest) matches the musical meter (measure) of the melody to yield a well-formed entity or gestalt.

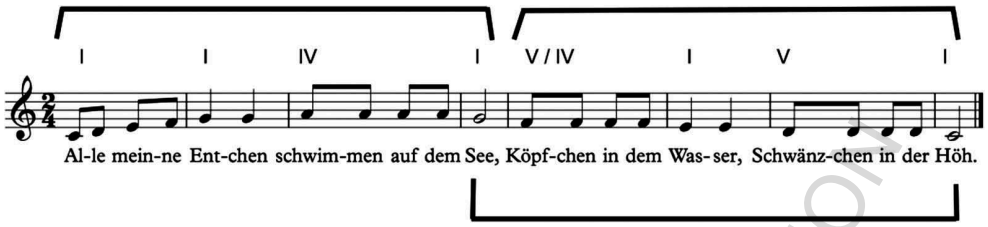

Figure 2.1 An example of a traditional German children's song. The top line shows the two phrases, the symbols I, IV, and V indicate the underlying harmonic structure of the melody, the last line points to the rhymes in the lyrics that mark the boundaries of the phrases. Note, for instance, the syllable–note correspondence, symmetries, and the same pitch of the beginning and ending notes.

Source: © Stadler Elmer, 2019.

Figure 2.1 exemplifies the structure of a traditional children's song in German. With each song acquired, children implicitly internalize the abstract song grammar and gain an understanding of the ritual. We have been studying implicit learning of the song grammar quasi-experimentally by systematically teaching children aged 2½ to 9 years songs that deliberately violate this grammar (Stadler Elmer, 2002, 2015). Micro-genetic analyses of coping with the unexpected rule violations show that children acquired the new songs on the basis of the previously acquired rules, and they corrected or circumvented the deviations in various ways. To conclude, this grammar serves as a reference system for the analysis of song models as well as of singing acts.

Why Canonical Babbling Is the Precursor of Song

In vocal development, canonical babbling is generally considered to be an important landmark in developing speech. But at the same time, it can be heard and interpreted as a landmark in developing song. Canonical babbling, therefore, is a key developmental stage for scrutinizing the early beginnings of both song and speech.

After the initial stage of crying come cooing, screaming or "protophones" signaling increased phonatory control (Oller, 2014). This is followed by systematic supraglottal articulated sounds such as repeated syllables (/mama/, /dada/, /baba/). The key feature characterizing canonical babbling as a new stage is the emergence of well-formed syllables. Nathani et al. (2006) define canonical babbling as consisting of more than two consonant-vowel syllables (CV) in sequence, including reduplicated babbling (repeated productions of the same consonant-vowel sequence, e.g., /dadada/) and nonreduplicated babbling (sequence of different CV combinations, e.g., /mababa/). Others denote single syllables as "monosyllabic" (e.g., /da/) and repetitions as "polysyllabic" babbling (e.g., /dadada/) (Nathani et al., 2006). The well-formed syllables in canonical babbling may be mistaken for words, thus, Davis, MacNeilage, Matyear, and Powell (2000) suggest no intentionality for canonical babbling vocalizations. Such productions signal the infants' discovery of cyclically moving the mandibular joints that brings about the opening and closing phases of the mouth (Davis & MacNeilage, 1995). Together with other vocalization types, operational definitions of canonical babbling used to determine emergence vary across research teams, models, and diagnostic tools. For example, age-of-emergence estimates vary between 5–10 months of age, or 8 months, or around 8 months at the earliest (Nathani et al., 2006).

To sum up, the key feature newly emerging with canonical babbling is the well-formed syllable, apparently yet without meaning and intention. It is considered the basic timing unit of language patterning (Davis et al., 2000). As claimed above, the syllable is also the basic unit for

singing in that it inherently carries pitch that can be organized in discrete categories, can be periodically stressed, and timed with discrete proportional durations.

The very same unit, the syllable, is interpreted by linguists as the precursor of words, whereas from a song perspective, it is a singing unit. Moreover, as will be argued in more detail below, a playful state would point more to singing than to speech acts because of assumptions about the underlying and guiding affects that at times may distinguish the two. Be that as it may, canonical babbling is as yet too undifferentiated to assign it to either speaking or singing. It is worth noting in this connection that in the other major group of vocal learners, songbirds that learn their song, vocal learning traverses a stage analogous to human infant canonical babbling called subsong (Doupé & Kuhl, 1999).

I propose canonical babbling as the stage at which the infant's production of chains of syllables does not yet allow identifying proper and distinct criteria for singing or speaking. The discovery of syllable production might lead the infant to explore syllable formation in its unlimited potential, while reducing and organizing it with increasing conformity to the social environment. At this stage, he or she does not yet differentiate between the two modes as intentional vocal acts but starts to use syllables as means for meaning-making with the social partners that guide the vocal forms as a basis of shared attention and feelings.

Parents and caregivers "understand" the infant's vocalization by taking into account facial expressions along with pragmatic, contextual information habitually used in social communication. In this role, the older generations are completely and necessarily culturally biased, since without this there would be no culture. It is their role to constantly use cues from mutually shared contexts and established routines to guide the infant toward increasing mutual understanding and conformity. Along this line, parents and caregivers tend to interpret the infant's communicative means as speech-like, while almost paradoxically, their intuitive parenting style with exaggerated prosodic features can be perceived as musico-linguistic. It is generally agreed that this intuitive parenting style with its *musical* features has important attentional, affect-regulating, and bonding functions.

To conclude, the interpretation of canonical babbling solely as speech-like or singing-like is inaccurate, misguiding, and therefore obsolete, since syllables are the basic units of both modes. The infant discovers this unit, explores and channels it toward the cultural targets provided by the communicative context, be it singing or speech. In order to differentiate song and speech proper, the infant must gain full command of phonatory and articulatory capabilities including the language-specific accentuation rules for meaningfully sequencing syllables to create words and sentences, and songs, and to share cultural meaning. Along this path, social transmission of these cultural acts and rituals is a necessary condition for the infant to differentiate the two modes.

From Babbling to Early Song Singing

How can we determine an infant's vocalization to be singing or speaking? What makes sequences of syllables appear as singing-like or as speaking-like? The conceptual framework summarized in Table 2.1 provides general knowledge for the structural analysis of vocalizations, of which some features are acoustically verifiable. As a micro-genetic methodology, it facilitates identification and description of the temporal configuration of features that help in interpreting the producer's intention.

The research on early song singing – under two years of age – is sparse (for reviews see Stadler Elmer, 2015). Researchers describe early singing – with varying emphasis – as consisting of glissandi, unstable pitches, chants with indefinable sounds, neologisms, short phrases within a narrow vocal range, and small and inaccurately tuned intervals. Further, progress is described in terms of increasing the duration and complexity of phrases, as getting more stabilized with respect to tonality and rhythm, and as linguistically more accurately pronounced. Some researchers reduce complexity by focusing on pitch accuracy – relative to the Western tonal system – and by

assessing progress in relation to increasing age only. Papoušek and Papoušek (1981) conducted pioneering research on early vocal communication from a psychobiological perspective and by using acoustic analysis. Our acoustically based micro-genetic analysis (Stadler Elmer & Elmer, 2000) of early song singing by a 20-month-old girl (Stadler Elmer, 2012, 2015) shows that she clearly identifiably produced a traditional melody and followed the grammar of children's song, but least of all with regard to proper pronunciation of the lyrics' syllables or words. At her level of language development, she was not yet able to articulate correct words or even lyrics, and so she formed onomatopoeic syllables resembling the target language. The syllable accentuation pattern followed the melody's meter (and not word formation rules), that was supported by the girl's regular rocking movements. Altogether, her early singing conforms to the singing-before-speaking or the musical-origins-of-language hypothesis.

An Infant's Intentional Transition from Speaking to Singing

Our longitudinal studies on early vocal development include systematic video recordings of dialogues, monologues, and musical activities with the aim of reconstructing how the infant begins to distinguish between singing and speaking. Here I present a scenario in which an infant changes from one mode to another and thus signals his intention.

Tom – age 14 months – grew up with song singing being a shared activity and with German as mother tongue. The selected video-excerpt lasts 41 seconds and shows the caregiver and child sitting on a couch and looking at a picture book. The caregiver points to an object and says /da/ meaning *here* or *this*. The infant repeats the word, explores it and unexpectedly starts to sing. As soon as the caregiver notices the transition to singing, she laughs briefly, and the child continues exploring his singing voice, and this scene ends with him singing a well-formed melody. What does he change while switching the mode? Which rules does he apply in either case?

When the infant began to sing, he regularly moved his upper body from side to side. In the following we have not analyzed his movements any further. Rather, we have made acoustically assisted microanalysis to investigate the vocal structures that emerge and culminate in a well-formed melody. The upper part of Figure 2.2 gives an overview of the acoustic analysis of the 41-second scene created with Pitch Analyzer (see http://mmatools.sourceforge.net, Stadler Elmer & Elmer, 2000). The *x*-axis shows the syllables' timing, especially onset, duration, and phrasing. Beneath each of the 38 syllables, which all appear grey in the picture, is the linguistic transcription with their accentuation or stress patterns. Numberings are indicated below. The *y*-axis represents the pitch continuum on which the Western tone system with its 12 semitones is indicated. The pitch of each syllable is calculated with an algorithm and shown in the illustration as black curves. This analysis contains the essential musical linguistic features that allow reconstruction of the rules that the child uses to move from speaking to singing: the syllables (CV combinations), their accentuations, their timing and pitches. With syllable 9 (Figure 2.2) Tom begins to move his body regularly, and at the same time he changes the quality of the pitch of the syllables: while he previously varied the pitch within the syllables, he now produces the pitch categorically, that is, relatively stable within syllables and varying between the syllables. The transition from speaking to singing – between syllables 8 and 9 – is shown enlarged at the bottom left of Figure 2.2, and at the bottom left as well the final well-formed melody (syllables 32–38) to demonstrate Tom's categorical use of the syllables' pitches as one of the major features of his singing in contrast to his speaking. With syllable 9 – at the same time as the beginning of regular body movements – he also periodically begins to accentuate the syllables, first trochaic, and from syllables 22 to 29 iambic, then trochaic again, and finally rhythmically complex (upbeat, etc.). The accents of the sung syllables are created by Tom through contrasts of loud–soft, high–low, long–short and combinations, while he varies the syllables only in the vowels (a, ä, ö), but not consonants.

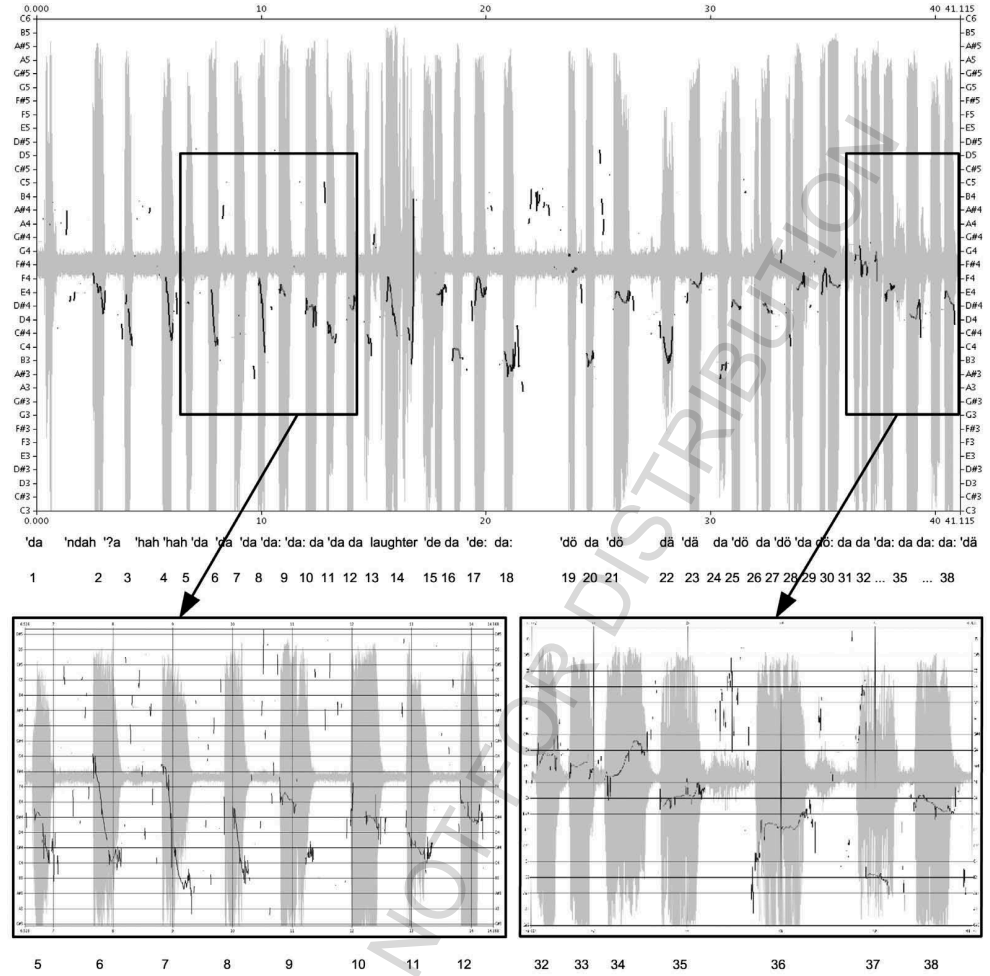

Figure 2.2 Analysis with Pitch Analyzer. The caregiver says /da/ (syllable 1), the infant says syllables 2 to 8, and starts singing with syllable 9 and ends with a well-formed melody (syllables 32 to 38). The enlarged section at the bottom left shows the spoken syllables 5 to 8 and the sung syllables 9 to 12, the latter are periodically accentuated and produced with categorical pitches. The enlarged section at the bottom right shows the final well-formed melody with rather stable pitches within the syllables and variations between them.

In summary, Tom expresses his intention to move from speaking to singing by moving regularly, repeating and varying syllables, adding a regular meter, and using categorical pitches. The vocal structures he produces reveal the manifestation of his implicit understanding of the rules of singing and those of speaking. At the center of these rules are the syllables that are differently shaped and organized for speaking and singing. For speaking, syllables are composed into words and sentences, while for singing – as does Tom – they may be repeated, varied – without linguistic semantics – and combined into melodies by meter and pitch categories. The joint viewing of a picture book and the short laughter of the caregiver (see syllable 14, Figure 2.2) suggest a playful and relaxed affective state – an important condition for the readiness and self-intentionality to sing.

Conclusion

My attempts to reconstruct the beginnings of singing have prompted me to postulate and consider canonical babbling as a milestone, both in speech and song development, and as inherently indecisive, since the producer expresses neither a discernible intention nor a discernible meaning. Canonical babbling is the precursor of both. As soon as the infant applies some cultural rules to form and organize the syllables, we no longer need to speculate about intentions. Vocalizations are guided by affect, and early speech acts primarily express needs, whereas early singing acts primordially express well-being or playfulness. The various acoustically supported micro-genetic studies propose to conclude that infants and toddlers find it much easier to adapt their vocalization to the rules of singing than to those of speaking. Implicitly, in regulating and expressing affective states such as playfulness or needs, they seem to discover features of syllables to combine and generate meaningful patterns. The minimal criteria for singing consist in prolonging vowels or syllables, and in modulating and periodically accentuating their pitches. Forming words with syllables is much more demanding and thus later in development than creating melodies with them.

References

Abler, W. L. (1989). On the particulate principle of self-diversifying systems. *Journal of Social and Biological Structures*, *12*, 1–13.

Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences*, *37*, 529–604. doi:10.1017/S0140525X13003099

Cross, I., Fitch, W. T., Aboitiz, F., Iriki, A., Jarvis, E. D., Lewis, J., … Trehub, S. E. (2013). Culture and evolution. In M. A. Arbib (Ed.), *Language, music and the brain* (pp. 540–562). Cambridge, MA: MIT Press.

Darwin, C. (1871). *The descent of man, and selection in relation to sex*. London: J. Murray.

Davis, B. L., & MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech and Hearing Research*, *38*(6), 1199–1211.

Davis, B. L., MacNeilage, P. F., Matyear, C. L., & Powell, J. K. (2000). Prosodic correlates of stress in babbling. *Child Development*, *71*(5), 1258–1270.

Deutsch, D., Lapidis, R., & Henthorn, T. (2008). The speech-to-song illusion. *Journal of the Acoustical Society of America*, *124*, 2471.

Doupé, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience*, *22*, 567–631.

Eckerdal, P., & Merker, B. (2009). 'Music' and the 'action song' in infant development: An interpretation. In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality: Exploring the basis of human companionship* (pp. 241–262). Oxford: Oxford University Press.

Francis, N. (2012). Poetry and narrative: An evolutionary perspective on the cognition of verbal art. *Neohelicon*. doi:10.1007/s11059-012-0148-7

Hall, T. A. (2000). *Phonologie: Eine Einführung [Phonology: An introduction]*. Berlin, Germany: de Gruyter.

Herder, J. G. (1772). *Abhandlung über den Ursprung der Sprache [On the origin of language]*. Berlin, Germany: Christian Friedrich Boss.

Janik, V. M., & Slater, P. J. B. (1997). Vocal learning in mammals. *Advances in the Study of Behavior*, *26*, 59–99.

Konishi, M. (2004). The role of auditory feedback in birdsong. In H.P. Ziegler & P. Marler (Eds.) *The Behavioral Neurobiology of Birdsong. Annals of the New York Academy of Sciences*, *1016*, 463–475.

Levman, B. G. (1992). The genesis of music and language. *Ethnomusicology*, *36*(2), 147–170.

List, G. (1963). The boundaries of speech and song. *Ethnomusicology*, 7(1), 1–16.

Lomax, A. (1967). The good and the beautiful in folksong. *The Journal of American Folklore*, *80*(317), 213–235.

Lomax, A. (1977). Universals in song. *The World of Music*, *19*(1/2), 117–130.

Lourenço, O. M. (2016). Developmental stages, Piagetian stages in particular: A critical review. *New Ideas in Psychology*, *40*, 123–137.

Masataka, N. (2005). *The onset of language*. Cambridge studies in cognitive and perceptual development. Cambridge: Cambridge University Press.

Merker, B. (2002). Music: The missing Humboldt system. *Musicae Scientiae*, *6*, 3–21.

Merker, B. (2005). The conformal motive in birdsong, music, and language: An introduction. In G. Avanzini, L. Lopez, S. Koelsch, & M. Majno (Eds.), *The Neurosciences and Music. Annals of the New York Academy of Sciences* (pp. 17–28). New York: Annals of the New York Academy of Sciences. doi:10.1196/annals.1360.003

Merker, B. (2009). Ritual foundations of human uniqueness. In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality: Exploring the basis of human companionship* (pp. 45–59). Oxford: Oxford University 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.

Merker, B. (2015). Seven theses on the biology of music and language. In P. A. Brandt & J. R. do Carmo (Eds.), *Sémiotique de la musique – Music and meaning*. Signata 6, Annals of Semiotics (pp. 195–213). Liège, France: Presses Universitaires.

Merker, B., Morley, I., & Zuidema, W. (2015). Five fundamental constraints on theories of the origins of music. *Philosophical Transactions of the Royal Society B*, *370*(20140095). doi:10.1098/rstb.2014.0095

Nathani, S., Ertmer, D. J., & Stark, R. E. (2006). Assessing vocal development in infants and toddlers. *Clinical Linguistics & Phonetics*, *20*(5), 351–369.

Nettl, B. (2000). An ethnomusicologist contemplates universals in musical sound and musical culture. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 463–472). London: MIT-Press.

Nottebohm, F. (1972). The origins of vocal learning. *The American Naturalist*, *106*(947), 116–140.

Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Erlbaum.

Oller, D. K. (2014). Phonation takes precedence over articulation in development as well as evolution of language. *Behavioral and Brain Sciences*, *37*, 567–568. doi:10.1017/S0140525X13004159

Papoušek, M. (1996). Intuitive parenting: A hidden source of musical stimulation in infancy. In I. Deliège & J. Sloboda (Eds.), *Musical beginnings: Origins and development of musical competence* (pp. 88–112). Oxford: Oxford University Press.

Papoušek, M., & Papoušek, H. (1981). Musical elements in the infant's vocalizations: Their significance for communication, cognition and creativity. In L.P. Lippsitt (Ed.) *Advances in Infancy Research*, *1*, 163–224.

Rousseau, J. J. (1781). Essai sur l'origine des langues, où il est parlé de la mélodie, et de l'imitation musicale. Retrieved from http://classiques.uqac.ca/classiques/Rousseau_jj/essai_origine_des_langues/essai_origine_langues.html

Schreiner, M. (1954). *Die grammatische Terminologie bei Quintilian [The grammatical terminology of Quintilian]* (Inaugural-Dissertation). München: Ludwig-Maximilians-Universität.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.

Snow, D., & Balog, H. L. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua*, *112*, 1025–1058.

Spencer, H. (1857). The origin and function of music. *Fraser's Magazine*, *56*, 396–408.

Stadler Elmer, S. (2002). *Kinder singen Lieder: Über den Prozess der Kultivierung des vokalen Ausdrucks [Children's song singing: On the process of cultivating vocal expression]*. Münster, Germany: Waxmann.

Stadler Elmer, S. (2011). Human singing: Towards a developmental theory. *Psychomusicology: Music, Mind & Brain*, *21*(1 & 2), 13–30. doi:10.1037/h0094001

Stadler Elmer, S. (2012). Structural aspects of early song singing. In A. Baldassare (Ed.), *Music – Space – Chord – Image. Festschrift for Dorothea Baumann's 65th birthday* (pp. 765–782). Bern, Switzerland: Peter Lang.

Stadler Elmer, S. (2015). *Kind und Musik.- Das Entwicklungspotenzial erkennen und verstehen [Child and music: Recognizing and understanding the developmental potential]*. Heidelberg, Germany: Springer.

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

Stark, R. E. (1980). Stages of speech development in the first year of life. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology, Vol. 1: Production* (pp. 113–142). New York: Academic Press.

Valsiner, J. (2003). *Culture and human development*. London: Sage.

Vihman, M. M. (1996). *Phonological development. The origins of language in the child*. Oxford: Blackwell.

von Humboldt, W. (1836). *Über die Verschiedenheit des menschlichen Sprachbaus und ihren Einfluss auf geistige Entwicklung des Menschengeschlechts [On the diversity of human language and its influence on the spiritual development of the human race]*. Berlin, Germany: Druckerei der Königlichen Akademie der Wissenschaften.

Vygotsky, L. S. (1976). *Psychologie der Kunst [Psychology of the arts]*. Dresden, Germany: VEB Verlag der Kunst. (originally published in Russian, 1925).

Wermke, K., & Mende, W. (2009). Musical elements in human infants' cries: In the beginning is the melody. *Musicae Scientiae Special Issue*, *13*, 151–175.