

# The roots of music: Emotional expression, dialogue and affect attunement in the psychogenesis of music

Musicae Scientiae

16(2) 200–216

© The Author(s) 2012

Reprints and permission: [sagepub.co.uk/journalsPermissions.nav](http://sagepub.co.uk/journalsPermissions.nav)

DOI: 10.1177/1029864912440778

[msx.sagepub.com](http://msx.sagepub.com)

 SAGE

**Ulrik Volgsten**

Örebro University, Sweden

## Abstract

In this paper I sketch the outlines for a comprehensive theory of the psychogenesis of music. That is, a theory of how human beings may come to hear certain sounds and combinations of sounds as music. It is a theory that takes its empirical starting point in previous and well-known research findings on fundamental human interaction and communication. As such it incorporates at its core a developmental-psychological theory about the human being's development of a sense of self in relation to others, from infancy on, and is further supported by findings from research on infants' behavior and reactions to music. It is argued that human interaction and communication is at the outset musical – or *protomusical* – and that which makes interaction and communication work is the emotive, or affective power of sound (“communicative musicality” is another term that has been used for mainly the same phenomenon). Although the empirical foundations are familiar, the comprehensive picture offered by the theory is new. The theory is structured according to a main thesis that states that music is a way of “making special” human self-development, our “sense of self”. However, for this affective-communicative theory to explain not only our reactions to protomusical sound, but also music at large (music understood in a broad universal sense), it must be extended to answer certain questions about human cognition. Therefore I start by referring to research in cognitive-psychology that explains cognition in part as a capacity to categorize phenomena according to their level of detail or generality.

## Keywords

affect attunement, communicative musicality, developmental psychology of music, dialogism, making special, self development, non-verbal communication, sound categorization

## In the beginning was the voice – the mother's voice

Already at four and a half to six months of age, babies seem to prefer certain phrasing in music. In one study, Mozart minuets were played with pauses inserted that either did or did not correspond to the phrase indications of the score. The infants in the test spent significantly longer

---

### Corresponding author:

Ulrik Volgsten, School of Music, Theatre and Art, Örebro University, SE-701 82, Sweden.

Email: [ulrik.volgsten@oru.se](mailto:ulrik.volgsten@oru.se)

time facing the loudspeaker out of which came the versions with pauses between phrases, than they did facing the loudspeaker out of which came versions that had pauses inserted in the middle of the phrases (Krumhansl & Jusczyk, 1990). When did the children acquire this preference for stylistically correct phrasing of music? Are we innately hard wired to prefer correct phrasing to stylistically-deviant versions? Some would not hesitate to say yes. Mozartean phrasing resembles human prosody and research shows that babies react to sound already in utero (cf. Grimwade et al., 1971; Olds, 1985; Woodward & Guidozzi, 1992; Zimmer et al., 1982). Others would equally emphatically reject the claim. In either case, an explanation is needed to tell us why the infants react as they do.

Studying the human being after birth reveals that already within the first three days of life, a newborn baby can distinguish its mother's voice from other female voices. Not only does it recognize its mother's voice; it shows a clear preference for it (DeCasper & Fifer, 1980; Karmiloff & Karmiloff-Smith, 2001). At two months after birth babies react differently to different kinds of prosodic speech patterns: falling speech melodies soothe, rising melodies attract attention, bell-shaped and falling melodies maintain attention, while bell-shaped and unilevel voice melodies discourage ongoing behavior. The effects are similar both in American English and Mandarin Chinese (Papousek et al., 1991). It has been suggested that the capacity to recognize the mother's voice among other female voices serves "to prime the newborn to respond preferentially to its mother and . . . provide important feedback to the mother who is looking for signs of recognition" (Hepper et al., 1993) and the recognition of one's mother's speech melodies has survival value (for a wider social significance, see Cordes, 2003, who found similarities between prosody in motherese and different types of socially functional songs). Without a mother (or other caretaker) to feed it, the child would starve.

This answers the *why* part of the question. An answer as to *how* the infant comes to prefer its mother's voice and certain phrasing of music has been offered by Sandra Trehub and her colleagues.

Trehub has studied both the attention-invoking dispositions of different types of speech contours (the exaggerated way of speaking that is sometimes referred to as baby talk, or motherese) and the soothing capacity of lullabies. She suggests as an explanation that these patterns correspond to prototypical categories that our cognitive apparatus is particularly apt at recognizing. The prosodic and the melodic contours are basic categories, with similarly perceived overall shapes that make them easily encoded and remembered. Trehub further suggests that the prototypicality of categories is accounted for by Gestalt principles (such as similarity, proximity and common direction). In particular she refers to the law of good continuation. The rising and falling contours, as well as the bell-shaped contour, are said to display "good forms", which make them particularly easy to perceive by the infant. In addition, the child will also notice deviations from such good patterns more easily than it will notice deviations from less good patterns.

According to Trehub, this might explain how the mother's voice comes to be recognized by the newborn (see Karmiloff & Karmiloff-Smith, 2001, for detailed explanation). The idiosyncratic deviations from the prototypicality of good patterning that the mother's speech and melodic phrasing displays captures the child's attention, and makes it possible for the child not only to recognize the general contour properties, but also to categorize the more fine-grained details of the mother's voice: Trehub says that

It is possible that infants go beyond a contour processing strategy, encoding the precise extent of the mother's pitch excursions or intervals. This would provide them with a basis for recognizing the mother by her unique yet familiar tunes, which may also be presented in a personalized set of rhythms. (Trehub & Trainor, 1990; Trehub & Unyk, 1991)

Trehub's explanatory framework shows how infants' reactions to prosodic and melodic contours to some extent are both systematic and predictable. But we may go further and ask by what biological or physiological mechanisms these systematic reaction patterns are carried out. The research I will refer to in providing an answer is particularly interesting for our discussion, since it will show that our cognitive capacity for the categorization of sound is fundamentally affective, a matter of feeling.<sup>1</sup>

## Neurophysiology of contours

According to René Spitz (1965, p. 44), the newborn's perception of the world is limited to the experience of "pure differences". A newborn cannot perceive any "thing" in the world, it has not yet formed any concepts for any particular objects. Instead it experiences abstract quality-changes such as rhythm, tempo, duration (Spitz, 1965), intensity, and shape (Lewkowicz & Turkewitz, 1980; Meltzoff, 1981). These experiences of various abstract quality changes are "amodal": instead of distinguishing between different sensory modalities they function as a "common sense", enabling the comparison between different modalities of sensory stimulation (Stern, 1985). What do these abstract quality changes consist of?

Neurophysiological research findings (that are too technical to summarize here; see Frijda et al., 1991; Wallin, 1991; Panksepp & Trevarthen, 2009, for emotional phenomena and their neural correlates; Wittmann & Pöppel, 1999/2000, for the neural basis of time perception; Nagy et al., 2010; Vickhoff, 2008, on mirror neurons and imitative behavior) indicate that experiences of abstract quality changes may consist of what Daniel Stern has described as "temporal pattern[s] of changes in density of neural firing": "[n]o matter if the object [is] encountered with the eye or the touch, and perhaps even with the ear, it produce[s] the same overall pattern of activation contour" (Stern, 1995, p. 84). As a result of these patterns of firing being hedonically appraised by the organism (Berlyne, 1971), the phenomenological experience is *affective*, it is a certain kind of feeling-experience. As Stern explains,

whenever a motive is enacted (whether initiated internally or externally, as in drinking water when thirsty or receiving and adjusting to bad news), there is necessarily a shift in pleasure, arousal, level of motivation or goal achievement, and so on, that accompanies the enactment. These shifts unfold in time and each describe a temporal contour. The temporal contours, although neurophysiologically separate, act in concert and seem to be subjectively experienced as one complex feeling. (Stern, 1985, p. 84).

Transferred to music, I suggest that the prosodic and melodic contours emitted in lullabies and motherese cause corresponding neural activation contours in the recipient listener, experienced as shifts between tension and relaxation. In other words, to hear a melodic contour is fundamentally to feel it (cf. Sammler et al., 2007). The activation contours also trigger attention to their external causes due to the shift in hedonic tone that they entail, which enable subordinate categorization and identification of the individual contours.

I have discussed elsewhere how the affective component of cognition may explain the motivation of further categorization and conceptualization (Volgsten, 1999). The question of motivation is important, since in contrast to the protomusicality of prosodic contours and the short melodic phrases of lullabies, music seems to have no natural function; music is very much a socio-cultural phenomenon (Brown et al., 2001; Volgsten & Brown, 2006). What I want to do now, therefore, is to show by which principles these short melodic and prosodic contours may be expanded and extended to account for the complex cultural phenomenon that we call music (I intentionally speak about music in the singular here, using the term in a

culturally wide and inclusive sense, cf. Nettl, 1983). Thereafter I will discuss in more detail the socio-cultural functioning of music.

## Levels of categorization

Although Trehub does not use the terminology, it is possible to describe the child's recognition of the mother's voice in terms of "basic" and "subordinate" levels of categorization. What is the difference between these levels of categorization? And what is "categorization" in the first place? Categorization, to categorize, means simply to group distinct phenomena together as being the same (i.e., belonging to the same category). In Western music, an "authentic cadence" and a "deceptive cadence" both belong to the category of "cadence", whereas "Search and Destroy" and "Innsbruck, ich muss dich lassen" both belong to the category of "song". Cadence and song are different categories. Level of categorization, on the other hand, refers to the amount of detail that distinguishes one category from another. Songs are categorized at a higher and more general level (in some cases encompassing the category of cadences as a distinguishing feature). At a lower, less general and more particular level of categorization we find subordinate categories such as "psalm", "lullaby", "jazz standard", and "schlager". Further down the categorical continuum we find particular psalms, lullabies and jazz standards, as well as cadences and pitch intervals. At lower levels we have to perceive more detailed attributes to distinguish the categories, while at higher levels the categories become increasingly abstract and general (cf. Mervis & Crisafi, 1982; Rosch & Mervis, 1981; Starkey, 1981).

A somewhat similar hierarchy of categorical levels can be found in the Liberian Kpelle terminology for musical performance. For instance, the term *pêle* refers to certain styles or types of singing, dancing and playing, whereas *wule* means song (or singing), and *Te-te-píúu* is the name of a certain song, each song being made up of man-made sounds, or voices *wóo*, and so on (Stone, 1981).

The basic level is thus not at the bottom of this continuum, but rather somewhere in the middle, between the extremes of particularity and generality. The distinguishing feature of basic-level categories is that they are easily encoded and remembered, they have similar affective shapes and show high intracategorical similarity and intercategory dissimilarity. Tellingly, they are the most easily acquired by children, some of them perhaps even innate. Subordinate categories, on the other hand, encode more detail for their differentiation, such as the mother's "unique . . . pitch excursions" and "personalized set of rhythms".

Considering Trehub's examples, we may assume there is a more or less innate preference for the good forms of the melodic contours at the basic level. Hedonic variation triggers attention to the specific details of the mother's voice and leads to subordinate levels of categorization, levels that can be biologically explained with reference to survival value. The capacity for subordinate discrimination and categorization of a brief melodic contour can be described as a *holistic pattern-recognition*. Holistic pattern-recognition means that the listener recognizes the global quality of a melodic or prosodic contour (pattern) without thereby necessarily being able to decompose and identify its constituent notes. A melodic line with a certain arrangement of notes has a particular perceptual quality, or affective tone, that does not equal the sum of its parts, whereas a different arrangement of the same constituent notes will display a different affective characteristic (cf. Warren, 1993). And while the lay listener may not be capable of repeating or analyzing the melodic compound into its constituent notes, the trained listener can (DeBellis, 1995; however, neither DeBellis nor Warren explains perceptual quality in terms of affective tone). Another example highlighting the difference between basic and subordinate

levels of categorization is the case of an untrained music listener spontaneously tapping his or her foot to an intermediate metric level of music (the strong basic-level beats), whereas the skilled musician has no problem attending and coordinating to both higher and lower levels of pulse and rhythm (Drake et al., 1997).

If we regard the brief melodic and prosodic contours of lullabies and motherese as basic level, many songs, melodies and compositions supersede the former's "basicity". Simply put, most music in most cultures is more extensive and complex than mere basic level contours. Various functional explanations may be given for this universally occurring phenomenon (i.e., music being more complex than mere basic-level contours). Such explanations may claim that music originally is a courtship device, or that music is primarily an enhancer of group cohesion (Brown et al., 2001), that it serves to display socially valued resources, or the control and channelling of aggression, or that music facilitates the establishment and maintenance of social identity through rites of passage (Dissanayake, 2006). The logic seems to be that, from an evolutionary perspective, "more music" makes a stronger impression and thus fulfils its function more efficiently (and that "more music" entails more elaboration and the quality of being "special", as I will argue).

### **Expanding the basic level repertoire of contours**

How then, is "more music" made from the basic level building blocks of prosodic contour? An important and efficient means of extending the affectively efficient principles of the basic level is metaphorical mapping. For instance, as Michael Spitzer points out, "When we hear the figured textures of Bach's passion as counterpoint, we are mapping from basic level categories of musical experience – the familiar realm of conventional contrapuntal technique – onto a level that is more distant from firsthand experience" (Spitzer, 2004, p. 21). A particular way of making sense of performance, composition or listening is "scaled up" (or "down") to make sense a new dimension of music. Of course, regarding counterpoint as a basic level category, as Spitzer does, already assumes a substantive metaphorical mapping from what I have defined as biologically basic categories, that is, the prosodic contours perceived already at infancy. The point, however, is that we seem to have a more or less innate capacity to metaphorically project and extend the qualities of basic level contours onto wider time spans (or spatial regions, as in the three-dimensional conceptualization of Kpelle music).

As Spitzer's example shows, basic level categories are not absolute and static, but flexible and modifiable (within limits and with respect to interest). For instance it has been observed that basic level categories may differ not only between cultures (Dougherty, 1981), but also between adults and their children (Mervis, 1984). Although this latter research has not considered music, I think musical examples would provide us with very similar conclusions.

So far I may have unduly emphasized research dealing with the perception and cognition of music heard. It is therefore important to point out the interactive aspects of music (i.e., playing, singing and dancing). Music is something we do, and passive listening (apart from lullabies) is largely a derivative form of musicking, which has developed in socially stratified cultures. As mentioned already, the newborn responds to its mother, it provides feedback. Infants respond by imitating and they also have "the capacity to provoke an imitative response, thus sustaining an interaction" (Nagy & Molnar, 2004). In addition, most (if not all) cultures engage their children from an early age in ritual or ritualistic dance, song and play, expecting participation and response (Turner, 1982). Children are equipped with a "communicative musicality", which not only accounts for their ability to engage in musical play, but also for their ongoing interpersonal

relations and social development (Malloch, 1999/2000; Malloch & Trevarthen, 2009). Thus metaphorical mapping should be seen not only as a way of cognitively extending one's ability to hear, but also as a way of making sense of what one is doing and of developing one's social interactions. Likewise, basic and superordinate levels of categorization are assumed to underlie performance as much as perception (according to David Hargreaves, 2011, listening to music is a creative practice; see also Harris Berger, 1997, on the heavy metal drummer's "practice of perception").

However, the basic and subordinate levels do not tell the whole story about categorical levels. Equally important is the *superordinate* level of categorization. In a way this superordinate level of categorization is even more important than the subordinate and basic levels, since it seems to be a uniquely human capacity. The main difference between lower and higher levels of categorization, as we have already seen, concerns the degree of particularity and concreteness versus the degree of generality and abstractness. High levels are more general and abstract, whereas low levels are more particular and concrete. In consequence this means that as we ascend from the basic level to higher and more general levels of categorization, it will become increasingly difficult to group phenomena on the basis of similarly perceived affective tone (the differences become too numerous), and it will become increasingly difficult to invoke single mental images that do justice to all instantiations of the category, whereas the action patterns required for performance of the category members become increasingly varied. At the topmost end of the continuum we reach the level where affective similarity is a minimal requirement and where *verbal stipulation* becomes the guarantee for categorical membership. In other words, what enables the acquisition of a superordinate category is verbal language (Benelli, 1988; Horton & Markman, 1980; Markman, 1989).

A good example is the general term *music* and its many cognates in the cultures of the world (whether they function as nouns or verbs). Music is a language-dependent superordinate category, and to experience sounding phenomena as examples of such general and abstract categories as (for instance) power ballads, ragas, taqsim or ngoma requires language. It is not the case that language gives us a term for a complex perceptual category that we already have. Instead, language is the very capacity that enables us to categorize such different-sounding instances as are covered by any of the above terms (power ballad, raga, taqsim, ngoma) under a single superordinate category. The same goes for the grouping of all categories together as an instance of, say, World Music. This might seem odd, even unlikely, though less so when we get the explanation.

## Superordinate categorization in action

Consider a situation when what counts as music for some is just strange noise for others.

What the superordinate category can do, or rather what the verbal label referring to the superordinate category can do – "music", say (or ngoma, or raga, or taqsim) – is that it can be used to tell the listener how to hear the encountered sounds, namely to hear them "as" music (or ngoma, or raga, or taqsim). The superordinate category can be used to tell the listener that this seemingly unmusical sound is "the same" in some aesthetically relevant aspect as that other sound which we already categorize as music (or ngoma, or raga, or taqsim). The superordinate category can thus function as a metaphor inviting us to assimilate the unfamiliar.

This may explain the phenomenon that John Blacking called "radical change". Radical change in a musical culture is a phenomenon that cannot be explained in the same way as

mere variation, Blacking says, since it “is the result of conscious human decisions”, involving “essentially *verbal* ideas” about music. Exactly how verbal ideas about music come to be necessary for change in “musical ideas” (the latter are said to “presuppose” the former), which are “nonverbal and . . . performative”, is not clear on Blacking’s account, but the notion of superordinate categorization provides an answer.

Any novelty must make sense to its audience to be accepted as music. As Blacking puts it, “individual Einsteins are of no significance unless their thoughts and actions can be shared by other members of society” (Blacking, 1986). If a new musical idea won’t make sense it will (at best) be dismissed as “original nonsense” (Kant, 1974, §46). In contrast, superordinate categorization, as I have explained above, makes it possible to subsume the radically new into one’s own preestablished cultural framework on the basis of metaphorical imagination (turning Blacking’s individual Einstein into a musical genius, as it were).

In addition to innovation, radical musical change may also consist in the assimilation of a foreign music (or function) into a previously established repertoire. Once superordinate categorization is applied to one’s own music it enables a decontextualization and recontextualization of the experienced sounds and their requisite activities. This amounts to a cultural self-reflexivity that enables, on the one hand, the recognition that a foreign music which sounds totally unfamiliar compared to one’s own can fulfil the same socio-cultural function, or on the other, that a foreign music regarded as similar to one’s own may nevertheless be understood as fulfilling a completely different function in its original (unfamiliar) context.

Superordinate category-labels do their metaphorical work by temporarily bracketing off musical sounds and the activities necessary for their production from worldly circumstances, such as social function and context. They enable the separation of the two parameters. Superordinate categorization enables us to say that two otherwise different instances of music (or musical performance) are the same. Superordinate categorization thus enables the public acceptance of radical change, either conceived of as the creative novelties and innovations that extend mere variation, or as stylistic pluralism within a culture, where style is understood as referring to functionally similar but perceptually different musics (Volgsten, 2006). In short, it enables the culturally self-aware development of art.

## Music as “making special”

According to Ellen Dissanayake, art is a way of “making special” the important aspects of our everyday lives (Dissanayake, 1980). I have mentioned some evolutionary important functions that music – or rather, musical sounds, protomusic – may have fulfilled, such as courtship, group cohesion or the establishment and maintenance of social identity. Music could, in line with Dissanayake’s thesis, be understood as a way of “making special” functionally efficient sound behaviour, where the “making special” means culture-specific subarticulations and superordinate extensions of a universally available basic-level repertoire of sounds. But music as a human cultural artefact involves more than this. What I suggest is that music “makes special” nothing less than the development of human identity, the dialectical articulation of an “I”, “you”, “we” and “them”. This very particular way of making the everyday special can be said to supervene on more “obvious” functions such as those already mentioned. What makes the enhancement of self-development a less “obvious” function than, say, the establishment and maintenance of social identity (through rites of passage), is that the former goes unnoticed and that it is hardly ever intentional. This does not mean that it would be less important in any sense; rather the contrary.

However, the idea of an unintentional way of “making special”, the result of which goes unnoticed, may strike anyone as odd unless it be clear that “making special” is regarded here primarily as a function, not as a purposive action or “use” (cf. Merriam, 1964). Music, as a human cultural artefact, is a result of “making special” the development of human identity, our sense of being a self in an interpersonal world. The means to this end is music’s emotional power, or rather (as we have already seen) the emotional power of sound, its affective component.

To support my claim I will refer to the developmental-psychological theory of Daniel Stern (to whom I have already referred), which articulates the human development of a “sense of self” in a number of interrelated and overlapping stages, or phases. I will first give a brief outline of Stern’s theory. Thereafter I will show how we, in analogy with the model proposed by Stern, develop our human capacity to perceive sounds as music. It should be kept in mind that in these trajectories (Stern’s and mine) each succeeding stage presupposes the preceding one, but it does not replace it (see also Stern, 2000, xiff.). A fully developed sense of self, as well as of music, requires that all the developmental stages, or phases, act in concert.

## The developmental psychology of self

According to Stern, the qualities of sensation to which the newborn child attends, such as its mother’s voice, constitute the earliest temporally organized islands of coherence and coordination in an otherwise non-differentiated chaos (what Spitz called an experience of “pure differences”). These experiences serve as the earliest points of reference between which relationships of significance can be inferred. Most notably, the child will increasingly experience the difference between events that it may enact itself and events that are beyond the limits of its own immediate volition. The experience of being the agent of certain coherent events but not of others gives the child a first sense of what Stern calls *self versus other*, while the regulation of the infant’s affective state by another’s ministrations lead the child on to a second level, to a sense of *self with other* (Stern, 1985, 69ff; 100f).

These togetherness feelings – which are radically different from the earliest non-differentiation and symbiosis with the mother – constitute the earliest developmental phases of our personal selves and point further to the level of a *subjective self* (Stern, 1985, p. 138). At the level of a subjective self, perceived behavior begins to acquire social significance. At this level, affective contours become related to the context of purposeful interaction, whereas earlier amodal perception was merely affective. It is at this level that a sense of intentionality begins to emerge – the self having already been articulated against and with an *other*: now this relationship takes on a rudimentary dimension of subjective purposiveness.

The crucial means to this development Stern calls “affect attunement” (Stern, 1985, p. 138). Affect attunement is a more or less unconscious communicative behavior building on the amodal similarities between the infant’s and the parent’s respective behavior. Without necessarily being aware of the fact the parent attunes to the child’s activities by performing an analogous action that retains the amodal properties of the original action. The attuning activity performed by the parent shares with the child’s activity the underlying affective contour. Both activities are similar to the respective agents with regard to the amodal qualities of shape, rhythm and intensity (as experiments show, whenever the parent misattunes, for instance by exaggerating the intensity or showing no sense of rhythmic timing, the child reacts by becoming upset).

Basically this is a communication of feelings, through feelings, which becomes successively more “dialogical”. The prosodic contours of the parent’s talk to the baby has an affective impact,



the contours are *felt* as they are heard. Likewise the baby responds with non-verbal vocalizations that are felt by the parent (and only subsequently interpreted as conveying information). This non-verbal, affective communication follows an intersubjective dialogical pattern rhythmically regulating the “calls” and “responses” (as it were) from parent and child. The dialogical pattern requires an affective attunement on the part of both. One senses the unfolding of the other’s prosodic contour and anticipates its phrasing off, at which point one’s own response may catch on (Jaffe et al., 2001; Stern, 1985).

Stern emphatically stresses that attunement is not a case of strict imitation (or “mirroring”), since that would direct attention to the external qualities of behavior. Rather, affect attunement “is the performance of behaviors that express the quality of feeling of a shared affect state without imitating the exact behavioral expression of the inner state” (Stern, 1985, p. 142). Affect attunement is more like a reciprocal variation of a mutually shared theme. It allows the growing awareness that the other is feeling something quite similar to what I feel, and that the activity is continued or interrupted for the same affective reason on both parts.

It is through affect attunement that we start to get a feeling for subjective selves and others. As this sense of a subjective self and a subjective other develops we become motivated to expand our grasp of affective contours in a temporally extensive and inclusive way. We sense that the ongoing activity (with its particular affective tone) is the outcome of what has happened earlier (with its particular tone), and we anticipate a continuation and perhaps even an end, a “closure” (all with their particular affective characteristics).

Stern speaks about “protonarrative envelopes” that add a temporally extended, narrative-like feeling shape to the perceived world, and where basic level affective contours come to function as elements in a subjectively articulated plot:

The elements of plot get temporally distributed on a line of dramatic tension. And the dramatic line of tension is invariably synchronous with the temporal feeling shape. This is natural, since the motive-goal-tension is played out in terms of temporal shifts in arousal, pleasure, motivational strength, and goal attainment. In a sense, the perceived plot is superimposed or rather dispersed along the temporal feeling shape, which then acts as the line of tension to carry the narrative. (Stern, 1995, p. 91)

When affective contours are experienced as linked together, they function as protonarrative envelopes, as a supportive scaffold for the distribution of perceived and successively denominated and narrated subjects, objects, actions and events. Subsequently language and narration themselves provide unifying themes that further extend the affective contour. Along with the acquisition of language, the sense of a subjective self will help to shape the child’s beliefs about its personal history and character, eventually enabling the final stage of development, at which we acquire a sense of a *verbal self* (Stern, 1985, p. 162). As Stern says, the role of language for the verbal self “is not primarily another means for individuation, nor is it primarily another means for creating togetherness”. The uniqueness of language is that it

ultimately brings about the ability to narrate one’s own life story with all the potential that holds for changing how one views oneself. The making of a narrative . . . involves thinking in terms of persons who act as agents with intentions and goals that unfold in some causal sequence with a beginning, middle and an end. (Stern, 1985, p. 173f)

As the verbal self develops, so does our linguistic ability to verbally specify personal characteristics (an ability sometimes referred to as “folk psychology”), the function of which is to predict and explain other people’s behavior (whereas emotion- and mood-terms are used for

less predictable behavior, see Griffiths, 1989). All in all, subjectivity and personality, that of others as well as our own, is not given in advance of social relations, and it is very much a question of verbal ascription of characteristics that emerge in, and as a result of, particular interactions with others. It is as much a result of one's developing sense of self as it is a requisite for it. We do not become fully developed personal subjects until we are able to ascribe subjectivity to others. Part of this ability is language-bound. Language, to quote Daniel Dennett, establishes a self as its source:

The strings or streams of narrative issue forth *as if* from a single source – not just in the obvious physical sense of flowing from just one mouth, or one pencil or pen, but in a more subtle sense: their effect on any audience is to encourage them to (try to) posit a unified agent whose words they are, about whom they are: in short, to posit a *center of narrative gravity*. (Dennett, 1991, p. 418)

## A developmental psychology of music?

As I have claimed, music is the result of “making special” the development of our sense of self in an interpersonal world. This does not mean that music depicts or denotes the process of self-development and identity-making in some imitative or symbolic way. Rather, it is the means of sound already employed in this process that are made special, elaborated into music. Music is a way of making self-development special by means of exploiting affective qualities at the core of human self-development. We have seen the various levels of categorization at work in this process, as well as the developmental stages, or phases, of self. Let me now add the finishing notes to the score by giving some examples that show how musical phenomena relate to this self-development model. At the end, we shall also be able to see to what extent the self-developmental trajectory is *necessary* for our perception of music.

That we experience short musical phrases as cohesive gestalts depends on the same principles of basic level categorization as does the child's preference for its mother's voice. We can hypothesize that it is at the earliest stages of self-development – the development of our senses of *self versus other*, and *self with other* – that perception of musical phrasing begins. We saw at the outset that infants are fully capable of perceiving the melodic phrases of a Mozart minuet. But not all music in the world is as simple in phrase structure as a Mozart minuet (an experiment with Japanese music played to American infants gave negative results, see Jusczyk, 2003). The point is of course that development starts with the simple and proceeds to the more complex. At the levels of *self versus other* and *self with other* the simplest musical phrases are distinguished from their affective impact, from the way they feel “in me”, as *other than myself*. Musical phrases are, for the first time, heard as being “out there”, as sound.

In contrast to simple phrasing, to perceive even the simplest melodic structure (an ABA form, for instance) as a continuous whole, rather than as phrases strung contingently one after the other, would require the capacity for metaphorical projection mentioned earlier. The counterpart in human self-development is the protonarrative envelope, as envisioned by Stern. The “elements of plot” in the case of music would be the short melodic phrases and motives of the melody, with their own affective contours, while the protonarrative envelope is the affective contour of the melody in its entirety, whether this melody is a simple children's song, a symphonic theme, or even a complete symphonic movement. The subordinate categorization of such envelopes enable us to differentiate and generalize different “schemas” at the basic level, sound schemas prescribing beginnings, middles and ends, which we come to expect and anticipate in our future encounters with music (Gjerdingen, 1988; Rosner & Meyer, 1986; Spitzer, 2004). Something similar to protonarrative envelopes is also, I believe, what Michel Imberty

has in mind when speaking about the “macrostructure” of a musical work (Imberty, 1997), and it seems to share affinities with Fred Lerdahl and Ray Jackendoff’s (1983) “prolongational reduction”, according to which tension and relaxation in a well-formed tonal piece is said to occur in a hierarchy of structural levels, the topmost of which embraces the entire piece – though Lerdahl and Jackendoff’s generative theory lacks an account of the affective nature of short phrases at the “surface level” (and neither do they draw any connection to the development of our sense of self).

Our ability to project protonarrative envelopes develops reciprocally with our sense of a subjective self in relation to others. As affect attunement gives me a sense that my dialogical counterpart feels similarly about our common interactivity, I can project my own growing sense of intentionality and expectancy, whereas any unexpected behavior by the other (which inevitably will occur) enables me to infer other intentions and expectations than my own. This has implications also for our growing sense of music as something other than our own affective experiences (as something objectively “out there”). The developmental level enabling a sense of a subjective self is also, I further hypothesize, the level at which we come to experience music emotively, as in some sense *expressing* feelings (it’s still too early to speak about specific emotions). This common phenomenon – to hear music as expressions of feelings – has been variously explained with reference to our propensity to animate the world around us – as R. T. Allen says, “we first take the world to be animate and expressive and then learn to de-personify and de-animate parts of it” (Allen, 1990) – or with reference to the rhetorical figure of *prosopopoeia*, the metaphorical figure attributing a voice or face to an inanimate phenomenon. Hearing music, Fred Everett Maus claims, is “a matter of continuous temptation” to engage in “the interpretative move from musical gesture to the unity of a persona” (Maus, 1989), a move less mysterious when we consider that for most of the time, music has been experienced in a direct fashion, either performed by ourselves or in performance by another person face-to-face with us, rather than in the mass-mediated way so overly common today.

The final level of Stern’s developmental trajectory involves the sense of a verbal self. As we have seen, language is a means of specifying one’s self and its history and character, and so on. Likewise, verbal accounts specifying the identity, history and character of music abound in most music cultures of the world, although focus may lie at different levels (as Alan Merriam said, there is no culture that has nothing to say about its music, Merriam, 1964, p. 117). For instance, as Steven Feld reports, the Kaluli speak of their music as a characteristic “lifting up over sound”, which is only possible for a song that has become “hard”. This view is directly related to their everyday environment in the forest, where sounds constantly shift figure and ground – an environment where only people that have overcome the softness of childhood and become hardened can survive (Feld, 1981). Another example is the Tiv, who conceive of composition as “a quick subtraction after slow addition”. The phrase (which is a translation of a single Tiv term) reflects a zero-sum view of life and may also characterize the catching of fish, figurative speech, withdrawals from the bank, the digging of a well, and miscarriage (Keil, 1979). Similarly, in the Western classical tradition, Anthony Newcomb has identified “plot archetypes” in some of the symphonies by Beethoven, Schumann and Mahler, expressing “renewed harmony to heal the wounds inflicted by mankind’s alienation from nature”, a progress from “Arcadia forward to Elysium” (Newcomb, 1992).

Whereas these examples concern performances or works, Susan McClary (1993) and Philip Tagg (1979) have both analysed how extensive “plots” are built up by smaller musical phrases and units (what Tagg calls “musemes”) in both classical and popular music of Western culture. The examples also exemplify the transition in music from protonarrative to narrative “proper”,

that is, when the affectively felt protonarrative envelope becomes consciously attended to as such, and analytically (i.e., verbally) specified at a component level.

However, it is not the narrative character of these accounts that make them language-dependent (to the extent that protonarrative envelopes develop prior to language); rather it is the abstract and general character of the concepts employed in the respective narrative accounts that require language. That two differently sounding phenomena may be categorized together as either a “lifting up over sound”, a “quick subtraction after slow addition”, or a “renewed harmony to heal the wounds” – this is what calls for a superordinate category.<sup>2</sup>

These examples also exemplify the introduction of the distinction between subject and object in our musical experiences. The verbal level is that at which we definitely separate the animate from the inanimate, when we say that music does not express emotions, at most it is “expressive of” them (Kivy, 1989). The verbal level is that at which we distinguish subject from object. The sound that I sense as something other than my own subjective experience may be conceived of either as a type of action – music is something we do, a “performance” – or it may be conceived of as “that which” we perform (or hear performed). These distinctions require a sense of a verbal self, with the mastering of language that the verbal self implies, and also what we may call a *verbal sense of music*, enabling verbal specifications and theoretical distinctions of what we hear. Paraphrasing Dennett, whom I quoted earlier, we may say that the Western cultural custom to posit a “work” as the objective source for various similar performances of music is a result of a verbal urge to posit a center of narrative gravity – in this case the narrative can be historically traced to the renaissance composition of *res facta* and the “equiparation” of artists with the holy Pope, assumedly enabling artistic creation *ex nihilo* (see Blackburn, 1987; Kantorowicz, 1961; Volgsten, 2012).

However, as Newcomb says with reference to the musical narrativity he verbally specifies in his analyses, narrativity need not be an “extramusical” matter, need not be “something external to the musical happenings themselves”; musical narrativity may rather be something that appears from the “interaction of formal paradigm, thematic character and recurrence, and plot archetype” (Newcomb, 1992). As I stated earlier, the developmental levels act in concert. We presumably never hear a piece of music exclusively from the vantage point of a verbal self (with an exclusive verbal sense of music); it seems much more likely that we fluctuate between the levels, giving rise to more analytically determined experiences at times, and more affective experiences of music at other times – thus irritatingly confusing the listener types listed by a sociologist such as Theodor Adorno. Nevertheless, Adorno’s opposite point is equally true, that critical analysis “is not an external factor added to aesthetic experience”; it “is inherent in that experience” (Adorno, 1989). Put differently, our musical experiences are always categorized to some extent by the discursive logosphere in which they appear. This holds, as we have seen, not only for the individual listener but for musical cultures at large (as well as for types of analysis that would not count as “critical” to Adorno).

In analogue to a fully developed sense of self (a state which one can perhaps at best continue to aspire to), a *full* sense of music – what I have called music as a human cultural artefact – would thus involve all the developmental stages pointed out. We can even say that “music as a cultural artefact” and “a full sense of music” is more than a mere analogue to a full sense of a self (with all its component stages); it is a real homologue. Music builds on many of the same affective components as our sense of self (such as prosodic contours), and our capacity to perceive sounds as music also builds on – and “makes special” – many of the same developmental principles as our sense of self. In particular I want to emphasize the role of affect attunement, of which I have so far said relatively little.

## **Final remarks: Affect attunement and its consequences for our understanding of music**

First, attunement is at work in any successful attending to a temporal sequence, music or other. According to Mari Riess Jones, attention to temporal sequences is primarily governed by certain biological “attending rhythms”. In order to attend to the temporal unfolding levels of a sound event, the listener has to synchronize his or her attending rhythms with those of the event. As she puts it, the listener’s “attending rhythms become phase-locked to corresponding time spans marked within the event”. This phase-locking between the listener’s internal rhythms and the various rhythmical layers of a piece of music “involves a synchronous interplay between an attender and an event in which the former comes to partially share the event’s rhythmic pattern”. In other words,

the biological basis for responses to event time takes the form of attunement rhythms that selectively entrain, that shift over nested levels, and eventually are shaped by the event itself . . . In effect, the structure of a temporally coherent event can function as a natural time keeper for the attender. Instead of a clock-timed world, the attender responds to a dynamic event-timed world in which time judgments depend on an event’s characteristic timing and on how its structure confirms or disconfirms some expected course. (Jones & Boltz, 1989)

Affect attunement, as described by Daniel Stern, goes beyond this ability to attune to temporal events. It involves feeling that which one attunes to (the prosodic contour, for instance), as well as maintaining a dialogical relationship to it. The dialogicity, the give and take, which Stern exemplifies with reference to childrens’ games, is what allows the participants to sense the other’s sensations through the enactment of similarly felt behavior. Attunement hereby becomes a means for articulating one’s own subjectivity in relation to another equally subjective agent. Attunement turns into affect attunement. It is important to see that subjectivity is not “objectively” there and “discovered” through the developmental process; subjectivity is intersubjectively created by the dialogically-related agents as they develop a sense of self in relation to another self. What is new in Stern’s account of this development is his articulation and stress on the role of affect attunement as the means for this self-development.

My aim, on the other hand, has been to highlight the role of affect and affect attunement in our experiences of music. Music, as I have suggested, ultimately builds on the same proto-musical substrate as early communication between parent and infant (of course there are also other emotion-invoking aspects in music, of the biological, personal and cultural kind). The sounds are felt, and the participants engage in a dialogue of feeling. In the infant-parent context this enables the ascription of subjective intentions to the other – as well as to oneself. When hearing music the listener likewise attunes, engages emotionally, with the sounds. And likewise intentionality is ascribed. The music is heard as meaningful, purposeful. Even the simplest music “talks” to the listener (which is why even bird song, simple or not, can be heard as music). And in the same way as language enables us verbally to narrate our identities, language enables us to determine our musical experiences, both individually and culturally. Our vehicle on this trip is our capacity for affect attunement.

Sounds become music only in so far as sensitive human listeners relate to the sounds, attune to them, and ascribe (more or less culturally specified) musical qualities to them. As a definition of music, this is more than a little circular; it is not until we view the issue from a developmental and psychobiological perspective that this circularity gains supportive strength, and we can start to disentangle the “spiralling” forms of music as a human cultural artefact.

Thus my final point. In sound there is no music until a listener *adds* it, appreciates it, passes it on to others, through affect attunement (the first listener may be the composer or musician, playing or imagining). The music is *added* to the work as the musicians follow the composer's instructions (more or less). The music is *added* to the sounding work as the audiences hear what the composer and musicians intended (more or less). And, as Kant would say, we expect others to join in with our judgement.

Music, in other words, is social. Take away the affective attunement, the emotional investment, to music and there is no more left in sound than a Mozart effect to trigger some mice in a maze (Brown & Theorell, 2006). Music certainly is more amazing than that. Music is a human cultural artefact, and its communication – its potential for human communion – should not be restricted by any tendency to reify its identity conditions (Volgsten & Åkerberg, 2006; Volgsten, forthcoming). Like *I* and *You*, music is a result of *You and I*. Here lie the roots of music, and all that makes music so special.

## Notes

1. I will not make any strict theoretical distinction between my uses of the terms “affect”, “emotion” and “feeling”, but rely on the reader's ability to infer from context what meaning is intended.
2. Likewise, the point is not that the verbal self involves the verbal narration of our selves (that is Stern's point, with all its important implications), but rather that the verbal self admits narration with the abstract and general terms that enable superordinate categorization, such as my “self”.

## References

- Adorno, T. W. (1989). *Introduction to the Sociology of Music*. Trans. E. B. Ashton. New York: Continuum.
- Allen, R. T. (1990). The arousal and expression of emotion by music. *British Journal of Aesthetics*, 30(1), 57–61.
- Benelli, B. (1988). On the linguistic origin of superordinate categorization. *Human Development*, 31, 20–27.
- Berger, H. M. (1997). The practice of perception: Multi-functionality and time in the musical experiences of a heavy metal drummer. *Ethnomusicology*, 41(3), 464–488.
- Berlyne, D. E. (1971). *Aesthetics and psychobiology*. New York: Appleton-Century-Crofts.
- Blackburn, B. (1987). On compositional process in the fifteenth century. *Journal of the American Musicological Society*, 40(2), 210–284.
- Blacking, J. (1986). Identifying processes of musical change. *The World of Music*, 28(1), 3–15.
- Brown, S. and Theorell, T. (2006). The social uses of background music for personal enhancement. In Steven Brown & Ulrik Volgsten (Eds.), *Music and manipulation: On the social uses and social control of music*. New York & Oxford: Berghahn Books.
- Brown, S. Merker, B., & Wallin, N. L. (2001). An introduction to evolutionary musicology. In Wallin, N. L. Merker, B. and Brown, S. (Eds.), *The origins of music* (pp. 3–24). Cambridge, MA: The MIT Press.
- Cordes, Inge (2003). Melodic contours as a connecting link between primate communication and human singing. Proceedings of the 5th Triennial ESCOM Conference. [http://www.epos.uni-osnabrueck.de/music/books/k/klww003/pdfs/143\\_Cordes\\_Proc.pdf](http://www.epos.uni-osnabrueck.de/music/books/k/klww003/pdfs/143_Cordes_Proc.pdf)
- DeBellis, M. (1995). *Music and conceptualization*. Cambridge: Cambridge University Press.
- DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mother's voices. *Science*, 208, 1174–1176.
- Dennett, D. C. (1991). *Consciousness explained*. London: Penguin Books.
- Dissanayake, E. (1980). Art as a human behavior: Toward an ethological view of art. *Journal of Aesthetics and Art Criticism*, 38(4), 397–406.
- Dissanayake, E. (2006). Ritual and ritualization: Musical means of conveying and shaping emotion in humans and other animals. In Brown, S., & Volgsten, U. (Eds.), *Music and manipulation: On the social uses and social control of music*. New York & Oxford: Berghahn Books.

- Dougherty, J. W. D. (1981). Salience and relativity in classification. *Language, culture and cognition: Anthropological perspectives*. New York: Macmillan.
- Drake, C., Penel, A., Bigand, E., and Stefan, L. (1997). *Tapping in time with musical and mechanical sequences*. Proceedings from the Third Triennial ESCOM Conference (pp. 286–291). Uppsala: Department of Psychology.
- Feld, S. (1981). ‘Flow Like a Waterfall’: The metaphors of Kaluli musical theory. *Yearbook for Traditional Music*, 13, 22–47.
- Frijda, N. H., Mesquita, B., Sonnemans, J., & van Goozen, S. (1991). The duration of affective phenomena or emotions, sentiments and passions. *International Review of Studies on Emotions*, 1, 247–259.
- Gjerdingen, R. O. (1988). *A classic turn of phrase: music and the psychology of convention*. Philadelphia: University of Pennsylvania Press.
- Griffiths, P. E. (1989). Folk, functional and neurochemical aspects of mood. *Philosophical Psychology*, 2(1), 17–30.
- Grimwade, J. C., Walker, D. W., Bartlett, M., Gordon, S., & Wood, C. (1971). Human fetal heart rate change and movement in response to sound and vibration. *American Journal of Obstetrics and Gynecology*, 109(1), 86–90.
- Hargreaves, D. (2011). *Musical beauty and imagination*. Paper presented at the conference “Music(s) and Human Beings II”. Örebro University, September 2010.
- Hepper, P. G., & Shahidullah, S. (1993). Newborn and fetal response to maternal voice. *Journal of Reproductive and Infant Psychology*, 11(3), 147–53.
- Horton, M. S., & Markman, E. M. (1980). Developmental differences in the acquisition of basic and superordinate categories. *Child Development*, 51, 708–19.
- Imberty, M. (1997). Can one seriously speak about narrativity in music? In Gabriellson, A. (Ed.), *Proceedings from the Third Triennial ESCOM Conference* (pp. 13–22). Uppsala: Department of Psychology.
- Jaffe, J., Beebe, B., Feldstein, S., Crown, C., & Jasnow, M. (2001). *Rhythms of dialogue in infancy*. Boston, MA, & Oxford: Blackwell.
- Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, 96(3), 459–491.
- Juszyk, P. W. (2003). *American infants’ perception of cues to grammatical units in non-native languages and music: Evidence from Polish and Japanese*. <http://hincapie.psych.purdue.edu/Juszyk/pdf/Polish.pdf>
- Kant, I. (1974). *The critique of judgement*. Trans. J. H. Bernard. New York: Hafner Press.
- Kantorowicz, E. (1961). The sovereignty of the artist: A note on legal maxims and Renaissance theories of art. In Meiss, M. (Ed.), *De artibus opuscula XL: Essays in honour of Erwin Panofsky* (pp. 267–279). New York: New York University Press.
- Karmiloff, Kyra and Karmiloff-Smith, Annette (2001). *Pathways to language: From fetus to adolescent*. Cambridge, MA: Harvard University Press.
- Keil, C. (1979). *Tiv song*. Chicago: The University of Chicago Press.
- Kivy, P. (1989). *Sound sentiment: An essay on the musical emotions*. Philadelphia: Temple University Press.
- Krumhansl, C. K., & Juszyk, P. W. (1990). Infants’ perception of phrase structure in music. *Psychological Science*, 1(1), 70–3.
- Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: The MIT Press.
- Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Audio-visual intensity matching. *Developmental Psychology*, 16, 597–607.
- Malloch, S. N. (1999/2000). Mothers and infants and communicative musicality. *Musicae Scientiae*, 3–4 [special issue], 29–57.
- Malloch, S. N., & Trevarthen, C. (2009). Musicality: Communicating the vitality and interests of life. In Malloch, S. N., & Trevarthen, C. (Eds.), *Communicative musicality: Exploring the basis of human companionship* (pp. 1–10). Oxford: Oxford University Press.
- Markman, E. D. (1989). How children constrain the possible meanings of words. In Neisser, U. (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 255–287). Cambridge: Cambridge University Press.
- Maus, F. E. (1989). Agency in instrumental music and song. *College Music Symposium*, 29, 31–43.

- McClary, S. (1993). Narrative agendas in 'Absolute music'. In Solie, R. A. (Ed.), *Musicology and difference: Gender and sexuality in musical scholarship* (pp. 326–344). Berkeley: University of California Press.
- Meltzoff, A. N. (1981). Imitation, intermodal co-ordination and representation in early infancy. In Butterworth, G. (Ed.), *Infancy and epistemology: An evaluation of Piaget's theory* (pp. 85–114). Brighton: Harvester Press.
- Merriam, A. P. (1964). *The anthropology of music*. Evanston: Northwestern University Press.
- Mervis, C. (1984). Early lexical development. In Sophian, C. (Ed.), *Origins of cognitive skills* (pp. 339–370). Hillsdale: Lawrence Earlbaum Associates.
- Mervis, C. B., & Crisafi, M. A. (1982). Order of acquisition of subordinate-, basic-, and superordinate-level categories. *Child Development*, 53, 258–66.
- Nagy, E., & Molnar, P. (2004). Homo imitans or homo provocans? Human imprinting model of neonatal imitation. *Infant Behavior & Development*, 27, 54–63.
- Nagy, E., Liotti, M., Brown, S., Waiter, G., Bromiley, A., Trevarthen, C., & Bardos, G. (2010). The neural mechanisms of reciprocal communication. *Brain Research*, 1353, 159–167.
- Nettl, B. (1983). *The study of ethnomusicology: Twenty-nine issues and concepts*. Urbana, Chicago and London: University of Illinois Press.
- Newcomb, A. (1992). Narrative archetypes and Mahler's Ninth Symphony. In Scher, S. P. (Ed.), *Music and text: Critical inquiries* (pp. 118–136). Cambridge: Cambridge University Press.
- Olds, C. (1985). Fetal response to music. *Midwives Chronicle and Nursing Notes*, 98, 202–203.
- Panksepp, J., and Trevarthen, C. (2009). The neuroscience of emotion in music. In Malloch, S. N., & Trevarthen, C. (Eds.), *Communicative musicality: Exploring the basis of human companionship* (pp. 105–146). Oxford: Oxford University Press.
- Papousek, M., Papousek, H., & Symmes, D. (1991). The meanings of melodies in motherese in tone and stress languages. *Infant Behavior and Development*, 14, 415–440.
- Rosch, E., & Mervis, C. B. (1981). Categorization of natural objects. *Annual Review of Psychology*, 32, 89–115.
- Rosner, B. S., & Meyer, L. B. (1986). The perceptual roles of melodic process, contour and form. *Music Perception*, 4(1), 1–39.
- Sammler, D., Grigutsch, M., Fritz, T., & Koelsch, S. (2007). Music and emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology*, 44, 293–304.
- Spitz, R. (1965). *The first year of life*. New York: International Universities Press.
- Spitzer, M. (2004). *Metaphor and musical thought*. Chicago & London: The University of Chicago Press.
- Starkey, D. (1981). The origins of concept formation: Object sorting and object preference in early infancy. *Child Development*, 52, 489–497.
- Stern, D. (1985). *The interpersonal world of the infant: A view from psychoanalysis and developmental psychology*. New York: Basic Books.
- Stern, D. (1995). *The motherhood constellation: A unified view of parent-infant psychotherapy*. New York: Basic Books.
- Stern, D. (2000). *The interpersonal world of the infant: A view from psychoanalysis and developmental psychology*. First paperback edition with a new introduction by the author. New York: Basic Books.
- Stone, R. M. (1981). Toward a Kpelle conceptualization of musical performance. *The Journal of American Folklore*, 94(372), 188–206.
- Tagg, P. (1979). *Kojak: 50 seconds of television music*. Göteborg: Skrifter från musikvetenskapliga institutionen.
- Trehub, S. E., & Trainor, L. J. (1990). Rules for listening in infancy. In Enns, J. T. (Ed.), *The development of attention: Research and theory* (pp. 87–119). North Holland: Elsevier Publishers.
- Trehub, S. E., & Unyk, A. M. (1991). Music prototypes in developmental perspective. *Psychomusicology*, 10, 73–87.
- Turner, Victor (1982). *From ritual to theatre: The human seriousness of play*. New York: Performing Arts Journal Publ. cop.
- Vickhoff, B. (2008). *A perspective theory of music perception and emotion*. Göteborg: Skrifter från musikvetenskapliga institutionen.



- Volgsten, U. (1999). *Music, mind and the serious zappa: The passions of a virtual listener*. Stockholm: Skrifter från Musikvetenskapliga institutionen.
- Volgsten, U. (2006). Between ideology and identity: Media, discourse and affect in the musical experience. In Brown, S., & Volgsten, U. (Eds.), *Music and manipulation: On the social uses and social control of music* (pp. 74–100). New York & Oxford: Berghahn Books.
- Volgsten, U. (2012). *Musiken, medierna och lagarna: musikverkets idéhistoria och dess betydelse för etablerandet av en idealistisk upphovsrätt* [Music, media and the law: The history of the musical work and its relevance for the establishment of an idealistic copyright]. Stockholm: Gidlunds.
- Volgsten, U. (forthcoming). Emotions, identity and copyright control – The constitutive role of affect attunement and its implications for the ontology of music. In T. Cochrane, B. Fantini & K. R. Scherer (Eds.), *The Emotional Power of Music*. Oxford: Oxford University Press.
- Volgsten, U and Brown, S. (2006). Preface. In Brown, S., & Volgsten, U. (Eds.), *Music and manipulation: On the social uses and social control of music* (pp. xxii–xvii). New York & Oxford: Berghahn Books.
- Volgsten, U., & Åkerberg, Y. (2006). Copyright, music and morals: Artistic expression and the public sphere. In Brown, S., & Volgsten, U. (Eds.), *Music and manipulation: On the social uses and social control of music* (pp. 336–364). New York & Oxford: Berghahn Books.
- Wallin, N. L. (1991). *Biomusicology*. Stuyvesant: Pendragon Press.
- Warren, R. M. (1993). Perception of acoustic sequences: Global integration versus temporal resolution. In McAdams, S., & Bigand, E. (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 37–68). Oxford: Clarendon Press.
- Wittmann, M., and Pöppel, E. (1999/2000). Temporal mechanisms of the brain as fundamentals of communication – with special reference to music perception and performance. *Musicae Scientiae* [special issue], 13–28.
- Woodward, S. C., & Guidozzi, F. (1992). Intrauterine rhythm and blues? *British Journal of Obstetrics and Gynaecology*, 99, 787–790.
- Zimmer, E. Z., Divon, M. Y., Vilensky, A., Sarna, Z., Peretz, B. A., & Paldi, E. (1982). Maternal exposure to music and fetal activity. *European Journal of Obstetrics and Gynaecology and Reproductive Biology*, 13(4), 209–213.