Philosophical Communications, Web Series, No. 34
Dept. of Philosophy, Göteborg University, Sweden
ISSN 1652-0459

Why does music move us?

Björn Vickhoff (Dept. of Musicology and Film Studies)
Helge Malmgren (Dept. of Philosophy)
September 1, 2004

Abstract

The communication of emotion in music has with few exceptions, as L. B. Meyer´s *Emotion and Meaning in Music* (1956) and the contour theory (Kivy 1989, 2002), focused on music structure as representations of emotions. This implies a semiotic approach - the assumption that music is a kind of language that could be read and decoded. Such an approach is largely restricted to the conscious level of knowing, understanding and communication. We suggest an understanding of music and emotion based on action-perception theory - present moment perception, implicit knowledge and imitation. This theory does not demand consciousness or the use of signs. Neuroscientific findings (adaptive oscillators, mirror neurons) are in concordance with our suggestion. Recently these findings have generated articles on empathy – relevant to the understanding of music and emotion. Key words: empathy, implicit knowledge, imitation, entrainment, mirror neurons, perceptual learning.

Intro

An obvious and intriguing aspect of music is its ability to communicate emotion. How this is done is a very old question. It has been an interest of many scientific fields such as philosophy, neuroscience, psychology, and musicology (Juslin & Sloboda 2001). Since the 1970s we have witnessed a semiotic approach to address this subject. The flow of music has been stopped and fragmentized into structural factors - elements carrying meaning (Tagg 1979). Obviously there are events and sounds in music that could be analysed as signs. But, the semiotic side is certainly weaker in music than in ordinary language. It is just as true to say that a walk in the forest is semiotic, since almost anything could be taken as a sign of something else, if we choose that perspective: traces of animals, signs of season, signs of

geology, signs of meteorology, signs of God's creation, nature as poetic metaphor of the soul etc. Even if music is semiotic, it is not just semiotic. It is, as is the forest, also a reality of its own, and is perceived so.

The Swedish film director Ingemar Bergman recently played suite nr 6 in D-major for violoncello by Johan Sebastian Bach on the radio. Then he asked: *"Can anybody tell me – Why do we have music? Man is the only species who has music. Can anybody answer this question? I am not religious, but I think we have received music as a gift to give us an idea of other worlds than the one we inhabit".* Implicit in the question is the recognition that Bach is not signifying this world. We have to admit the suite is not telling us much about the world we know. It is just imagined beauty.

Action-perception theory

We need to ask: Are we, in everyday listening, really perceiving music as signs to be decoded? Could it be the other way around: that everything in the present moment (motivation, intention, state of mind, and memories of all kinds) is automatically affecting our perception of music? This makes listening to music spontaneous and emotional – a lived experience rather than a decoding of signs. The difference is that we enjoy music as such – as melody development, harmonies and rhythms, just as we can enjoy nature without "reading it". The philosophical consequence of this approach is that if music is not a sign of emotion, the emotion some way or other has to be inherent in the music heard. We intend to discuss this. In the following we will outline a theoretical framework for the analysis of the perception of music and its emotional implications. This framework is built of elements from many works in several different research traditions. What unifies them is the emphasis on the dynamical nature of perception and the close connections between perception and action. Hence we have drawn from Husserl's phenomenology of time consciousness, from the early psychological perception-action theories of Lipps and Rorschach, from Merleau-Ponty's phenomenology of perception and action, from the ecological psychology of J. J. Gibson, and from recent theorists and experimentalists in the fields of child psychology, ethology, neuropsychology and neurophysiology.

The temporality aspect is crucial. Daniel Stern's concept the present moment is based on Edmund Husserl's intentional arc which implies:

> *... being aware of what I have just seen, heard or thought – retention – as well as anticipating somehow a continuation of what I am now seeing, hearing or thinking – protention. (Stern 2004)*

The present moment is thus not just momentary *presence* but a time span framing perception. This includes an experience of what we just lived and an anticipation and intention based on what just happened. The usual span of the present moment has been estimated to 2-4 seconds – the length of a breath and thus the length of a phrase. It is the dividing line between experiencing the world directly ("live" and "on-line") and indirectly as the world symbolized (talked about or thought about). It is our window to life. Since the perceived moment has some expansion in time, we can perceive movements and actions – life as a process. And in fact we can perceive time itself as durations.[i]

It is just not possible or useful to be aware of everything in the present moment. As we perceive, we focus our attention to something that becomes *figure* of our perception. The rest is background. We are aware of the figure but not of the background. If something happens in the background, such as a false note in the violins, or a child crossing the street in front of our car while we are speaking on the cell phone, we direct our attention to it and it becomes figure (cf also Malmgren 2004).

We wish to stress that under the semiotic level of knowing and understanding, there is an immediate "perceive-and-act" level of knowing and understanding, designed to enable species to live in their environment. Perception is intertwined with action. We act on what we perceive, we learn to perceive by our interactions with the world (i.e. the infant's cross-modal exploration of the world), and we have to act to perceive (move our eyes, walk around, turn our heads etc.). The knowledge how to move is called *procedural* knowledge by today's psychologists (see, e.g., Snyder 2000:33) or (following the philosopher Ryle) *knowledge-how,* as opposed to *declarative* knowledge or *knowledge-that* (Ryle 1949). Procedural knowledge is mainly a kind of *implicit* knowledge in that it need not involve consciousness. The acquisition of procedural knowledge by experience is called *procedural memory*. *Perceptual memory*, or

the way perception is modified by experience, is similar to procedural memory in many respects. Not only because perception usually involves perceiving action possibilities, but also in its purely sensory aspects. When one has finally succeeded in perceiving a certain ambiguous figure as a definite Gestalt, the same perception will automatically reappear when one sees the same figure in the future, without any conscious act of interpreting.

In several pathological conditions, declarative and procedural knowledge are dissociated from each other. Merleau-Ponty describes a case with a man who suffered from a brain injury (Merleau-Ponty 1945). This man could not, when asked to, put his hand on his head (although he clearly understood the instruction), but when a fly landed on his head, he smashed it. The conclusion is that although he could not do it deliberately, his body knew how to do it, when the lived situation demanded so. Procedural knowledge is bodily skill (stored in the nervous system). It is the knowledge how to walk, move the mouth as we speak, ride a bike etc. - actions we are normally not conscious of. It includes knowledge of the space and terrain we move in; how to reach, catch and find our way. Some actions, like the newborn calf's ability to walk, are inborn, some are learned by imitation, and some are consciously learned but become automated after practice. Most of these abilities are too complicated to be taught by communication of signs, or even consciously understood. They contain too much information for our rather limited capacity to consciously hold and handle information. We could try to teach a child about the physical mechanisms at play in bicycling, or to verbally describe all the movements she has to perform, but this would not make the child able to ride the bike.

Neonates (30 minutes!) have been proven to imitate mouth movements (Meltzoff & Moore 1987). Findings like these have been rejected by some cognitive psychologists as paradoxes since these children could not possibly represent another person's face. This is because imitation, as they understand it, involves a separation of the self from the other and thus has to be conscious. But the child does not seem to care, it still imitates. Therefore we must accept that this imitation is inborn primary behaviour and that conscious understanding of the relation of self and other is a secondary capacity. The conclusion is that there are two kinds of imitation: one hardwired instant imitation and one intended conscious imitation.

Symbolization makes it possible to see our self from the outside. This leads to a cleft between the subject as the one living in the world and the subject as the one being aware of this subject. The symbol is the tool separating the "I" from the "you", the "I" from the "world" and finally the "I" from the one denoting herself "I" (the ontological cleft, the loss of innocence - expulsion from paradise). Symbolization liberates us from the present moment and makes it possible to actualize things and events not present. By means of symbolization we are able to talk and think (we can also think in images). When we are thinking we are not really there, not really present, since we are out on a little journey in our thoughts. And thus:

> *In the conscious movement, the velocity is slower and accuracy is poorer. This means that you have lost the on-line control of the movement, which works automatically: instead, you have used a system that is slower and is not adapted to doing automatic movement*. (Jeannerod & Gallagher 2002:)

The knowledge on this conscious level is called *explicit.* Explicit knowledge is not limited to conscious knowledge how to do things, but includes most of our knowledge about fact and events. Knowledge about facts is called *semantic* or *declarative*. It is our knowledge about the world, learnt by verbal communication through teachers and media. Knowledge about personally experienced events is called *episodic* - memories of lived experiences of present moments. A well known example of episodic memories is Marcel Proust's memory from childhood, triggered by the taste of a Madeleine biscuit soaked with lime blossom tea. As in the case of Proust, these memories are often brought to consciousness by our senses – a scent, a melody. Episodic memories may be partly constructed from knowledge categories and then depend on semantic memories. And, when we construct a narrative from episodic memories, as Proust did, they become totally semantic.

Conscious levels and unconscious levels work together in everyday action.

- ✓ We may be conscious about intentions, but not conscious about means.
- ✓ There is a connection between implicit and explicit knowledge: "With continuous, regular use, concepts in semantic memory are evoked automatically, and hence become implicit" says Snyder (2000), and exemplifies with the sight of an apple. This sight would immediately and automatically actualize the word "apple".

- ✓ And contrary: The sound of the word “apple” automatically triggers sensory images and bodily reactions. The conscious act of decoding is automated into the act of perception.
- ✓ Spontaneous expressions can be consciously decoded and deliberate representations can be perceived without conscious decoding.
- ✓ Implicit as well as explicit knowledge affects expression.
- ✓ Thus every level of awareness generates bodily expressions that can be perceived by others.

How is emotion communicated?

In order to say that an emotion is communicated, the emotion of the other must somehow be transmitted to the subject. On the conscious, representational level we can communicate the truth (or the lie) *that* we have a certain emotion. But this is not equivalent with communicating *the* emotion. There is a difference between listening to somebody telling us she is happy and being affected by her smile. Sadly, we can be quite indifferent when somebody tells us about her happiness.

We can see in herd living animals that they communicate by means of movements and sounds, and that each species have their own inherited system. It is hard to see music as such a system. But it seems as though music has much in common with prosody and body language. We know that prosody, facial expression and body “language” (another language which is not a language) are affecting us in interpersonal communication. These expressions could be analysed as signs of emotion, but on the daily level we do not decode them. We are merely contaminated by the other person’s expression. Perception does not separate expression from emotion. We live the expression of the other *as* an emotion.

Prosody, facial expression and body language are all motor activities. Emotion is intertwined with action in several ways, for example in the sense that emotion reflects motivation which is thought to steer and moderate action. As animals move through the landscape, the *affordances* (vital information in the environment, Gibson 1986:127) affect the bodily state of the animal (hunger, thirst, fatigue, sexual drives, satisfaction, fear, rivalry etc.). And the bodily state

affects perception. This interplay decides intention and steers action. Automatic two-way connections exist between emotions and movements. When we scratch a dog behind the ear, she starts kicking. And just by performing a facial expression mechanically, we get affected. This feedback from the manifestation of an emotion to the emotion itself is one of the keys to the understanding of the communication of emotions. Another key is *immediate motor identification:* If a person is thirsty and drinks – we need to drink, if she vomits – it makes us nauseous, if she urinates - we want to urinate, if she yawns – we have this sudden urge to yawn, if she cries – it brings tears to our eyes, if she laughs – she makes us laugh, if she gets angry – we might have a slight tendency to get angry too, if she sings a song – we tend to fall in, if she tries to make 2m in a high jump – we "help" her by lifting our own leg (as noticed already by Darwin 1872) if she loves us and gives us a hug – it is easy to love her back, and if she is sexually aroused – it is hard to ignore her. Together these keys can explain how a perceived manifestation of an emotion in another person can evoke that same emotion in me - without any intermediaries in the form of conscious representations. We are contaminated by the other's basic actions. The action of the other triggers a spontaneous tendency to imitate, and this tendency makes us feel the emotion.

Emotion is a complicated concept. It is hard to distinguish the nuances of feelings, moods, drives, needs, motivation, relational bonds or attitude. There is an emotional aspect of all these phenomena. What interests us is the communication aspect. As noticed by Gallese (2003) this is implicit in the concept *Empathy,* originally a translation of the German word *'Einfühlung'*, used in the psychology of aesthetic experience

> *...to denote the relationship between an artwork and the observer, who imaginatively projects himself/herself into the contemplated object. Lipps (1903), who wrote extensively on empathy, extended the concept of Einfühlung to the domain of intersubjectivity, which he characterized in terms of inner imitation of the perceived movements of others. When I am watching an acrobat walking on a suspended wire, Lipps notes, I feel myself so inside of him (ich fühle mich so in ihm).* (Gallese 2003:)

Here empathy is connected to the aesthetic art object and to intersubjectivity in general. It is not restricted to a specified emotion such as sympathy or urge to assist. What are the

mechanisms behind this transference of emotions from the observed to the observer (Scherer 1998)? Lipps theorizes that the essence of empathy is inner imitation of the perceived movements of others. Similar thoughts were voiced by the psychiatrist Hermann Rorschach, who argues that motor identification is involved even when we perceive human movement in a static picture (Malmgren 2000), and by the philosopher Merleau-Ponty. As we shall see later, the basic idea which unites Lipps, Rorschach and Merleau-Ponty has been extensively confirmed by recent research in psychology and neurophysiology.

Stern (1993) identifies three independent registers of feeling:

1. *Categorical affects*. These are emotional states like the Darwinian basic six (happiness, sorrow, fear, disgust, anger and surprise – detectable as facial expressions). Primary, early affects are happiness, anger and sadness. Happiness is developed very early in interpersonal relations (returned smile). Anger and sadness are developed from distress – happiness hindered. Considered secondary are affects of shame, guilt, and embarrassment requiring self consciousness and consciousness of the attention of others. [ii]

2. *Vitality affects*[iii], captured by kinetic terms: "crescendo, decrescendo, fading, exploding, bursting, elongated, pulsing, wavering, effortful and easy".

3. *Relational affects* ("being loved, esteemed, thought wonderful, special or hated, and the feelings of being secure, safe, attached, alone, isolated, or separated").

The *vitality affect* is Stern's most interesting contribution since it adds a dynamic aspect to the understanding of emotion. It requires shared "almost identical capacities for temporal discrimination" between the participants. Stern shows how emotions are cross modally evoked, shared and regulated. The inner subjective qualities of feeling are displayed as form, intensity and timing, observable and directly perceived as a physical expression of emotions. For instance, the child smiles to her mother, the mother answers y-e-a-h in a smoothly arched pitch form showing the rise and fall of the smile of the baby, meaning - "Oh! This is pretty

much how you must have felt, and I am showing you I know by matching the temporal contour of my activity to yours" (Stern 1993). Intuitive *motherese* (the way mothers address their babies) is universal. There is a focus in this conversation on the time course of the feeling intensity, the *temporal feeling shape* (Stern 1995:87). The temporal feeling shape of the observer will correspond to that of the observed, resulting in *affective attunement* - shared affective states. This connects with experiments where light-dot displays were placed on people in dark environments in order to capture pure movements (Cutting/Kozlowski 1997). The results indicate that we are not only capable of recognizing the identity of a walking person, but also the emotion of this person from pure movement observation (cheerful or depressed and sad.) Other experiments show that temporal information is transferred intermodally (Kamachi et al 2003).

Entrainment

An intriguing aspect of the communication of vitality affect is *entrainment,* referring to our capacity to share the present moment dynamics and to synchronize movements with each other. In fact we have to share the present moment to be able to apprehend the change in intensity. We perceive each other on line. *Emotional contagion* has been defined (Hatfield et al. 1994) as "*the tendency to automatically mimic and synchronize facial expressions, vocalizations, postures and movements with those of another person and, consequently, to converge emotionally".* If we read this strictly, it seems that just by synchronizing, we share the emotion. Although radical, it fits with the feedback connections between emotions and their manifestations, described above. It is also in line with Stern's observations. The importance of the time factor was been verified in his video experiments of mothers presented on screen to their babies. When a short time displacement between the baby's utterance and the mother's reaction was made, the baby lost interest. Even if there is much more to explain in this process, we cannot ignore that in a broad field of research entrainment is a condition for communication of emotions. It is *the* common factor. As Stern notes neurological research seem to substantiate the possibility of entrainment by the finding of so called *adaptive oscillators*

> *…that act like clocks within our body…These inner clocks use the real-time properties of incoming signals to "set" your adaptive oscillators, so that they*

> *immediately bring their own rate of neural firing into synch with the periodicity of the incoming signal.* (Stern 2004:)

For a mathematical analysis of adaptive oscillators see Zacksenhouse (2001). The oscillator networks responsible for entrainment of musical rhythms seem to be widely distributed in the brain (Thaut 2003).

Daniel Stern again:

> *…when people move synchronously or in temporal coordination, they are participating in an aspect of each others experience. They are partly living from the others centre.* (ibid.:)

If two persons were to look at the same object and move until the object looks exactly the same, they would end up standing inside one another. From this point everything looks the same and feels the same what ever direction they look at. This seems to be what happens when we share the present moment. The astonishing effect is that this makes us share perspective not only of time but of the whole spectrum of perception. However, identification in the real world is only a partial overlap. We are still able to distinguish ourselves from the other.

The perspective of the other is crucial for empathy. Stern gives a striking example, taken from Bråten (1999). Autistic children, by definition lacking in empathy, imitate, but not by putting themselves in the other's position. When a mother holds up her palms, children usually put their palms so that the palms touch. But children diagnosed as autistic tend to put their hands with their palms facing themselves! They imitate from their own point of view. This indicates that the description of empathy as the ability to take the other's perspective should be taken literally - not metaphorically! In one experiment a preverbal infant was watching an experimenter unsuccessfully trying to knob off a dumbell-like object. When the child was given the opportunity, it immediately tried to and succeeded to knob it off and seemed pleased at the result (Stern 2004). Thus the whole intentional complex of motivation and emotion is transferred. This is *action understanding* - an interesting form of understanding since it relies on implicit knowledge and takes place without conscious effort. We automatically perceive

movements as intentions and things as invitations to act. This is done by co-producing these movements/actions in our own body, living the intention.

Mirror Neurons and Empathy

Neurophysiological studies on monkeys revealed 1992 mirror neurons in the frontal lobe firing as the monkey watches movements related to hand and mouth (Di Pellegrino et al. 1992). The very same neurons fire when the monkey performs the same movements. Hence it seems that when the monkey observes hand movements that are part of its repertoire, the seen action is automatically retrieved (without having to be executed) in the motor system. Mirror neurons thus prepare us for imitating motor sequences (Rizzolatti et al. 2001). These findings provide additional understanding of entrainment – sharing of the present moment. This sharing occurs when the mirror neurons fire as a consequence of the actions of the other.[iv] There is strong evidence from fMRI and other studies that mirror neurons exist in humans (for a review, see Rizzolatti & Craighero 2004). As in the dum-bell experiment referred to earlier, experiments have shown that monkeys, like humans, can infer the goal of the action, even if the visual information is incomplete. Supporting evidence in humans is found in a group of patients with prefrontal lesions (Lhermitte, 1983). These patients will compulsively imitate gestures or even complex actions performed by an experimenter. This behaviour has been explained as an impairment of the inhibitory control normally exerted on the elementary motor schemata.

Several theoretical models have been proposed for the mirror neuron system, but we need not go into the controversial issues. We will use the term *sensorimotor schema* to refer to the whole complex of perceptual memories and motor schemata which is involved in the perception of an action and in the execution of the same action. The correspondence between the sensory and motor components of such a schema cannot always be perfect, since it is obvious that in many cases an observed action can at most be partly mapped onto an existing motor schema. Else we would never have to learn to perform.

As Kosslyn (1993) and others have argued, the neural substrates of perception and imagination overlap to a large degree. Hence it is natural to ask whether sensorimotor

schemata are activated when we only imagine an action. The following quote from Marc Jeannerod suggests that this is the case:

> *...mental images, at least in the visual modality, rely on the same neural substrate as the perceptual images that are generated during normal perception. ...motor imagery pertains to motor physiology in the same way as visual imagery pertains to visual physiology... motor images must also be considered to be functionally related to the imagined movement*. Jeannerod (1994:)

Jeannerod cites a convincing example of the connection between imagery and motor skill fetched from the field of sport:

Yue and Cole (1992) compared the increase in muscular strength produced by actual training (repeated maximal isometric contractions) and by mental training (imagining the same effortful contractions). In both conditions, the maximal force produced by the trained muscles increased significantly (by 30% and 22%, respectively). *(ibid.:)*

In 2002 Stephanie Preston and Francis de Waal published an article on empathy introducing the PAM – model (Perception – Action Model), influenced by the findings of mirror neurons. The article is treating empathy as a communication of an emotion from one person to the other so that the emotion evoked in the subject is the same as the emotion of the other. The authors argue

> *[that] attended perception of the object's state automatically activates the subject's representations of the state, situation, and object, and that activation of these parameters automatically primes or generates the associated and somatic responses, unless inhibited.* (Preston & deWaal 2002:4)

The "object" in this quote is "the other". The importance of imitation to empathy is stressed by the notion that expression, and recognition of expression and gestures are impaired in individuals with autism. And, even more, by the observation that autistic infants lack in coordination of activity with their caregiver.

The authors present a list of catalysts facilitating empathy:

- ✓ Similarity

- ✓ Familiarity
- ✓ Cue salience

Among the invited commentaries published with this article, Iacobini and Lenzi) report that they have identified areas with mirror properties in the human brain (the fronto-parietal circuit). When imitation and observation tasks were compared, they found a complete spatial overlap of activation between imitation and observation. The only difference was the magnitude of the response. This part of the brain is connected to the *insula* which has a

> *...key role in affect generation and serving as a relay between frontal cortices and limbic structures, thus representing a possible pathway for empathic resonance*. Iacobini and Lenzi (2002)

It is generally thought that just because mirror neurons have been found in restricted areas, this does not mean that they are not present elsewhere. For mirror neuron theory to be applicable to music and emotion we would need proof that the theory is not restricted to visual perception. And, in October 2003 the discovery of audiovisual mirror neurons was reported. They represent actions independently of whether these actions are performed, heard or seen (Keysers et al. 2003).

All and all, the research could be summed up in a simple, symmetrical and reversible general model.

Emotion↔Motor activity↔Audiovisual information↔Motor activity ↔Emotion

For this general model to be applicable to music we need to show that it makes sense to contemplate music as implicit knowledge. This entails that

- ✓ music is movement
- ✓ music movement is an expression of emotion
- ✓ music movement can be imitated through entrainment

Who is imitated? And, what about the conditions – similarity, familiarity, cue salience, entrainment and intentionality? How do they operate in music? Since this is a new approach to music and emotion, there is little research to refer to. The following is anchored in the first

author's own practical experience of the profession of music, including composing, recording, concerting, rehearsals, suffering, and the bliss of sharing emotions with audiences. We are not testing of the truth of the theory, but merely discussing how reasonable it is to apply it to music, how it agrees with professional and everyday observations of music, and discussing what the implications would be.

Application to Music

Music as movement. An obvious feature in music is that it makes people move. We jog, dance, run, pump weights, play air guitar, and mime in skipping-ropes to music (some of us). Rap music is largely a way to move, even to the extent that there is a fashion of lose cloths to go with this lax rhythmic pattern. Rhythm in beat music is indissolubly connected to dance moves. There is a connection between popular music and the construction of the human body. If you cannot move to it, it is no good. How can we explain this connection between music and movement?

To play an instrument is to get involved in an *acoustic-haptic-somatic* system. For a guitarist, this includes:

- ✓ *Exterospecific* information: The feeling of the guitar in her lap. This feeling is an essential part of the musical experience of the performer, as is the feeling and resistance of the strings.
- ✓ *Propriospecific* information: The posture of the body when she plays the instrument, which involves necessary balance, coordination of fingers and relaxation to be able to use the best technique.
- ✓ *Expropriospecific* information: The relation of the active body to the guitar – the sensory feedback from the instrument to her body as she attacks the strings with her nails in combination with the acoustic feedback.

Musicians have to move to produce music. Move their hands over the instrument. Move their entire bodies to sing, or to blow wind or brass instruments. In fact, if we were to invent a device to register in sound how a person is moving her fingers over a table, we would end up

with something like the piano. The instrument registers the attack, the direction (left-right), the intervallic leaps, number of fingers used, the spread of the fingers and the speed as sounds; as well as very subtle nuances in tempo and touch. Every heard change in pitch and dynamics corresponds to a movement by the player. This goes for any instrument including the voice. Thus we can state that music is heard musician movement. Movement is immanent in the heard information.

The sensorimotor schemata – memories of perceptions partly mapped as patterns of movement – are, when applied to music, to be considered as the sedimentation of all music the subject has heard so far - the soundtrack of her life. The sensorimotor schema is relevant to intriguing phenomena in music:

1. It helps us to **recognize sound sequences** produced by human beings, not only structure but also continuous nuances. Our capacity for voice recognition is particularly well developed. We are able to identify a voice as a voice and separate it from other sounds. Even very ambitious strivings to produce artificial voices are revealed. Prosody is extremely hard to fake artificially. We also detect dialect and identity of speaker, and the emotion and intention of this speaker. This sensitivity to voices should according to action-perception theory be due to our highly developed motor ability to produce vocal sounds (the similarity condition for empathy).
2. It explains the automated **connection between a heard tone and the motor production of this tone**. We can step up to the piano, play a tone and then immediately sing it. A choir leader can give the voices tone and the whole choir starts singing at commando. Almost anyone can imagine a tone and then sing it. The explanation is that hearing it or imagining it is to rehearse the motor action latently without being aware of it.
3. It works as a **comparison tool in music technique training and embodies the results of this training**. The sensory component of the schema can work as the goal of the interaction player-instrument - the imagined "vision" of the sound – the ideal sound. For the professional player this goal or another ideal prototype is never achieved but she closes in on it during years of training. The learning process is automatic and deliberate imitation, but also other conscious work which gradually becomes automated. Training makes the player as sensitive to the sound quality of the instrument as others are to voices. Technique is to reach the imagined goal with a

maximum of precision with a minimum of effort. The player feels control and this interaction with the instrument is confirming her **musical self** as a co-perception of the sound. This grace is not achieved until the playing is absolutely deliberated from conscious effort. When we see this in a player, we will get the impression of elegance, or even grace. The sense of perfection – accuracy in hitting the pitch, tight rhythm, perfect sound - is an aesthetic impression in its own right and as such part of our enjoyment of music. As listeners we transmit this accuracy to our own bodies. This may be an explanation of the bodily felt satisfaction in listening to well performed music. Basically it is the same bodily satisfaction as when we manage to hit a target perfectly – the sensation of a perfect backhand volley. It is the feedback of success telling our bodies: This is how to do it! We feel the elegance of the move. Thus a living good performance is a confirmation of the musical self, not only of the performer but of the listener as well.

4. It is **a tool for anticipating the action of the other**. Years of training and listening render the artist a repertoire of sensorimotor schemata. She has an implicit knowledge of where the music is heading. The primed motor pathways of the listener resonate with the motion of the music as she listens. Music, even when complex, is recycling of limited material according to implicit rules connected to style and genre. This creates expected chord progressions (such as vamps), expected instrumental melodies (such as guitar riffs) and expected melodic moves. Thus there is both recognition from earlier heard music and recognition from the music at hand (the familiarity condition).
5. It **makes imitation possible.** A trained pianist can imitate another pianist by ear. She imitates the sound, but in doing so she has to re-produce the movement. This imitation has two steps: First, she is imitating the movement latently *as* she listens to it. The experienced instrumentalist feels the *acoustic-haptic-somatic* system of the other in her own body. Second, she does a deliberate imitation and when doing so the memory of what she just heard is still in her fingers. Musicians' amazing capacity to imitate by ear proves that the movement is communicated. In this sense music is choreography. The intention to imitate is conscious. The *how* to do this is implicit knowledge. A problem to be addressed is the fact that different instruments demand different movements, which can be referred to the similarity condition. The melody creates separate choreographies on different instruments. Thus there is no simple general relation between melody and body movement. This objection does not overthrow the fact that the information of movement is implicit in the music. Can the average listener

perceive this movement? Obviously we move to the music heard. Even people that are not musically trained can sing and thus imitate a singer, or even imitate an instrument by voice. We know implicitly how to move our bodies to produce not only rhythm and pitch but also the timbre of voices and instruments. Strangely, music seems to be spatially related to the body. Trevarthen (1999) reported on a 5-month baby, born blind, who "*without prompting or training, and, indeed without her mother being aware of her graceful rhythmic gestures, conducts portions of the melodies"* of her mother's nursery songs, with her left arm and hand. The baby raised her arm and spread her fingers as the verses were swelling in excitement and dropped her hand when the stanza ended. This is a perfect example of Sterns intermodal vitality affect. Most of us perceive "high" notes as spatially high and "low" notes as spatially low. There is an observable tendency in listeners to imitate the base drum with their feet and the melody with the upper body. The spatial relation to sound is reinforced by seeing the movements of the artist, the musicians, the dancers, and the conductor on stage, as well as the movements of the rest of the audience. Music is thus in the concert situation a multimodal communication of movement, even including vibration sense when you can feel the base drum "rearrange your bone structure". However and most important, even when the direction of the movement of the listener seems to be chosen spontaneously, the vitality quality and contour of the movement is still imitated.

6. It is **used in improvisation**. The expected continuation is so strong that it is heard and can be instantly imitated. Musicians just play the protention of the perception as a natural continuation. Improvisation can be a play with expectances, where the musician deceives and deludes the audience with artistic deviances from expected outcomes. This is possible because she has so many sensorimotor schemata as solutions to every now in the music. Even this is implicit. This renders richness to the music, surprise – wonder. It is noted by smiles and eye contact between musicians. It can be derived from the *socio-dramatic communication* between mother and child (Trevarthen 1992:128), described as a playful intersubjectivity. It is the play. It is like the unforeseen move of Ronaldinho. In jazz we see it in the in-out technique when the soloist plays for a while outside the harmony. This can be overtly observed as a bodily movement sideways as if the musician for a while stepped out of her body – choosing an alternative way. In flamenco the unexpected move is commented by the

connoisseur's introvert murmuring – *olé* - as recognition of *duende* (Spanish, literally meaning goblin or elf - the presence of magic) (Martin 1973:43).

7. And this is also to a large extent what **composers** of popular music do. "It just comes to me". Composing thus, is a conscious selection between possible continuations that make themselves heard.

Except from these heard and seen movements of the performers, music is movement in a second sense: A special case of *Gestalt* relevant to music is the principle of closure. This creates immanent dramaturgy. A spoken phrase can be open as in a question (upwards in pitch) or closed as in a statement (downwards). This call-response pattern is an archetype in music. The well-known four-bar II-V-I chord progression in jazz is harmonic closure. There are so many clichés in music for closing the present phrase (cadences, half cadences, riffs etc.) The entire song in western popular music is consists of phrases sequenced in longer rows of progressions to a final closure. This is perceived as motion because closure is expected. We perceive the direction to this closure. Listening to popular music is like walking in a familiar environment. It is the play between variant and invariant information as the landscape unfolds (Gibson 1986). We expect the things we see to have a closure (and not to continue endlessly in the invisible field) and as we move this hunch (perceived expectancy) is verified. Finally we are back where we started – home - the closure of the whole piece. Since the musical journey is not actually spatial, it is no real movement. But it is more than a metaphor since it is perceived as a longings, tensions, forces and movements.

Salient cues: So far we have commented on two of the conditions for communication of emotions: similarity (as knowledge of the instrument) and familiarity (as knowledge of the music). What about salient cues? Of course there are many cues to entrain in music. The most important one is rhythm, since rhythm generally facilitates entrainment (cf. note iv). Experiments have shown that a person tends to move her limbs in ratios of 1/1, 1/2, 1/3 or 2/3. This has been referred to as *self-entrainment* (Port et al. 1996). It has been proposed that self-entrainment is due to *"a natural and nearly unavoidable temporal constraint for systems of mutually coupled oscillators"* (op. cit.). If so, we have a biological ground for the ratios in musical rhythms making up to the discrete time intervals in music. This is how we naturally move our limbs when we play and dance. Rhythmically complex music like flamenco reveals that the skilled musician is able to coordinate highly complex ratios. Intriguingly, ratios are also what the natural tone system is built from, making up the discrete intervals in scales.

Together these two variables constitute a coordinate system with discrete pitch over discrete time – the musical score. Although the combinations are innumerable there are in western popular music just 12 possibilities in pitch within the octave and about 16 possibilities in each bar for every occurrence of an event. The discrete quality of music cuts down the number of possibilities for the next event in a melody (pitch) from a continuum (endless possibilities) to a limited number of what and when.[v] This renders cues to entrain. The most salient cue of all in popular music is of course the base drum insisting now, now, now - until we know exactly when the next now is going to happen – creating this now-trance. But all the other events are ratios of durations between these beats. And the larger forms are multiples of it. We know where the now is. And thus we know where we are in the music as a co-perception. We feel the sixteenth notes clicking, even in the breaks. Not approximately but exactly. The time cue is implicit in the music. The rhythmic cue in beat music is so salient that this alone triggers entrainment and opens up the listener to emotional responsiveness; and at times to trance, to possession of deities, to therapeutic music communication where no other communication works, and to mass psychological suggestion. Sublime experiences in music are often reported as a loss of the sense of time (Gabrielsson 2001). We may recall here that shared temporality is a condition for the communication of Sterns vitality affects.

Communication in the concert is not only communication from the stage to the individual listeners. It synchronizes the listeners to one body - to one moving mass. The single listener is just as affected by the mass as by the music. Known mass psychological effects are a lowering of self consciousness and critical thinking and an increase in spontaneity and emotions. Michail Bakhtin wrote about the medieval carnival:

> *The mass of the carnival is a popular unity…sensual and concrete…the crowd in itself, the very physical contact between the bodies. The individual… a limb in a crowd. In this unity the individual body to a certain extent ceases to be itself: you could so to say exchange bodies, renew yourself (by means of masks and costumes). At the same moment the people experience their sensual, concrete and material bodily unity and community.* (Bakhtin 1986)

Bakhtin emphasized the loss of individuality – the body is exchangeable in masses. What happens when individuals become a mass is interpersonal entrainment in full scale, uniting the individual movements, emotions and intentions to one movement, emotion and intention.

This can take any direction as in mobs or in religious/political meetings. Hate or love and anything in between. Everything said about entrainment so far is strengthened and pronounced when the entrainment involves an audience of individuals and transforms them into a mass. When an artist expresses love, her love thus can embrace the whole audience and even be enriched by the mass. There is an audience feed back to the artist creating superconductivity.

A drumbeat in a rock festival is not a monotone copy of something already heard and known, but every time a unique event enriched by all the other drumbeats. Where the analytical listener feels bored of the monotonous repetition, the listener entrained is captured in the present moment trance of 80.000 colleagues and throws his head more violently at each individual beat as if that particular beat was the ultimate experience. And this very beat explodes as he leaves the earth with a jump in a sustained n-o-o-o-o-w. And there he sails in a body no longer named Patrick, or Betty - but We. And the drum is no longer on stage but beats inside We.

Intention. Why do we have music? What is the purpose? There is no obvious Darwinian cause. This touches an old problem with our understanding of music. It is hard to see the intention. Probably intention is as manifold in music as in other human behaviour. Connected with this question are two other questions: Who is "the other" in music? and - What is the perspective from which this other is acting? We are helped by the lyrics, if there are lyrics and sometimes by the occasion (wedding, funeral, political manifestation etc.). The feelings towards the artist are vital to the experience of the music. Knowledge of the biography helps us identify the other. Often enough the artistic production is a biography in itself. Media provide an image. Here too the conditions familiarity (knowing the artist, experience of the situation) and similarity (the artist is one of us) are important for action understanding and sharing.

Let us recall Ingemar Bergman's imaginative proposal that music provides us with an idea of other worlds than the one we inhabit. It is obvious that a large body of music expresses a longing for other worlds – nostalgic reminiscences of childhood, innocence, home, mother, love, pure nature, etc., or a religious longing to a future Heaven. Maybe the reason why music so often lends it self to this, lies in its potential to express longing by means of rows of

disharmonic chord progressions, hindering the anticipated resolution. And, more important, as we entrain the present now, critical thinking is diminished and emotions and spontaneity is reinforced. This means that music has the power to bridge the ontological cleft, manifest in self consciousness, between our selves and the world. Paradise is a construction from episodic memories of perceiving the world directly. Significantly, it was the tree of knowledge that caused us to be expelled from the biblical Paradise. Music by means of entrainment has the power to reinsert this direct state, and thus to give us reminiscences of paradises lost. Music thus allows us to put ourselves in resonance with the world. The Laplander sings his jojk for no one - and for the whole world. We can pick up the guitar and play a song just for enjoying the sound - and to communicate with the evening. The urge for self expression is surely a fundamental drive in all human beings. If we should understand this in terms of intention we probably have to go down to a very deep and generally not conscious motive – the drive to connect. We think Stern's term *vitality affect* is well found because the force of life penetrates all expressions of feelings. This takes so many creative forms. The artist's cry for connection is one. The unspoken intention in a rock concert is to express the emotion on stage and involve thousands in this emotion, to touch them, to move them. The thin naked body of a Mick Jagger crying out: Look at me, I am singing my life here: my anger, my sufferings and my love. I give my heart to you. I am bleeding for you. This is a rite, a feast, a communion, a sacrifice - shared mythological experience. We embody the artist ritually. It is com-union and e-motion. He becomes us. His anger and suffering is our anger and suffering. We share this life, united in motion.

Still, we cannot ignore that music can move us even if we do not have any idea of who the other is and consequently do not discern any obvious intention. This leaves two alternatives: Either music experience shows, contrary to what the imitation theory is proposing, that we can be affected by pure expression without perceiving an intention, or intention is implicit in the musical expression. The two aspects of musical movement (the movement of the musician and the anticipated closure) are intertwined in the sense that listeners follow culturally learned musical paths (sensorimotor schemata) to expected universal goals (gestalt closure). Is this a situation where the listener can discern the intentional goals or end-states necessary for implicit action understanding?

Is music movement an expression of emotion? Music-psychological research shows that listeners reliably discern emotional categories (Gabrielsson & Lindström 2001). To take one

example: Melancholy/sadness is, in music as in prosody, the universally best communicated emotion. In music, melancholy is low register, legato, minor chords, small melodic leaps, low volume, lack of dynamics, slow, monotone and firm rhythm. Halftone movements (chromatics) and disharmonic (unsolved) chords are particularly melancholy. The richness of variables studied (multivariate analysis technique) gives us a hint of how many variables we "weigh together" in our perception of music without being conscious about it. The parameters studied are said to be structural features and the communication of emotions is understood as a decoding of these structures as signs. But, when we take a closer look at musical parameters, we find that with few exceptions they could very well be analysed as movements - as vitality affects. Intervallic leaps are the size of the movement. Legato is a description of the smoothness of the motion. Volume is, as pitch, related to intensity of the motion. Rhythm is the regularity of the motion. Dynamic change is the vitality contour of motion, and so on. As we can see, it is through vitality affects - in the case of melancholy, lack of vitality – that the emotional category is perceived. Emotional category is just another name for a certain combination of dynamic features, or – a certain combination of motions in the imitating listener: sensorimotor schemata. Also, we think the Stern relation affect is due to affect attunement and thus dependent on vitality affects. We love the artist because we have shared present moments of implicit action understanding with her. Thus it is wrong to think of category, relation and vitality affect as independent – at least when it comes to music. The vitality affects seems to be primary, underlying the others.

Traditionally, listening to music in emotional contexts in film, advertising, computer games etc. is thought to render signs of emotion. But we think these connections are usually automated into the perception resulting in instant feelings, instead of coding representations of emotion. As we listen, our total knowledge, implicit and explicit, affects perception. Let us take the general model: Emotion _ Motor activity _ audio-visual information _ Motor activity _ Emotion, and apply it to the communication of fear in music. The sensorimotor schemata of fear in the composer are intermodally transformed to musical dynamic patterns, these resonate with the sensorymotor schemata of fear in the listener and produce fear in her. The prototype of fear in music could be heard in Alfred Hitchcock's Psycho.

fear: sensorimotor schema→music: attack, staccato, high pitch, loudness→ sensorimotor schema: fear

Now the critical reader may remark: “Are we really to believe that the composer is scared when he is composing?” There are several possibilities here. Either he is working in a flow of inspiration and by imagination actualizes a sensorimotor schema for fear. Or, he is consciously using expressions of fear as representations to reach the emotional effect, but in doing so he experiences fear by means of feedback from these musical signs. In both cases the sensorimotor schema is activated and implicitly leads his musical expression. For the listener the result is the same.

Coda

Music is a perfect example of implicit knowledge. Composing, performing and listening to music are largely bodily acts. Initially we stated that the philosophical consequence of a non-semiotic approach would be that the emotion is implicit in the music as such. This is hard to accept, but the sensorimotor schema could be the key to this understanding. We propose that music is an auditory picture of the composer’s sensorimotor schema. This can be sedimented in a music score or a phonogram and brought to life later and act on the sensory motor schema of the listener. To be contaminated by music therefore does not necessarily demand the presence of *the other*, even if this helps as additional information and action understanding. When we listen to a record the artist is absent, so is the composer, still we entrain the flow. Alfred Schütz once wrote:

> *Although separated by hundreds of years, the latter (the listener) participates with quasi-simultaneity in the former´s (the composer’s) stream of consciousness, by performing with him step by step the ongoing articulation of his musical thought. The beholder, thus, is united with the composer by a time dimension common to both.* (Schütz quoted in Martin 1995:199)

We would say: Although the listener is separated from the composer by hundreds of years, she synchronizes with an auditory picture of his sensorimotor schema. Perceptual expectation makes her a co-composer - the motion and the emotion is shared. She, thus, is united with the composer by means of entrainment of an auditory picture of his present moment. And being united with the other’s present moment is, as we have argued, to (partially) share perspective, motion and emotion.

Music moves us. Music is dancing us until the dancers vanish. Everything is movement. This is our existence – gestures towards a dancing world.

How can we know the dancer from the dance?

Yeats

Acknowledgements

We are grateful to Jan Almäng and Olle Edström for helpful comments.

References

Bachtin, M. 1982, *L'oevre de François Rabelais et la culture populaire au Moyen Age et sou la Renaissance*. Gallimard, Paris.

Bråten, S. 1999, Intersubjective Communication and Emotion in Early Ontogeny (Studies in Interaction). Cambridge University Press.

Cutting, J. E., Kozlowski, L. T. 1977, Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society*; 9:353–356.

Darwin, C. 1872, *The Expression of the Emotions in Man and Animals*. J. Murray, London.

Di Pellegrino, G. Fadiga L., Fogassi, L., Gallese V. & Rizzolatti G., 1992. Understanding motor events: a neurophysiological study. *Experimental Brain Research*, 71: 491-507.

Gabrielson, A. 2001, Emotions in strong experiences with music. *Music and Emotion – theory and research*, edited by Juslin, P. & Sloboda, J., 431-449. Oxford University Press.

Gabrielsson, A., Lindström, E. 2001. The Influence of Musical Structure on Emotional Expression, *Music and Emotion, theory and research*, edited by Juslin, P. & Sloboda, J., 223-248. Oxford University Press.

Gallagher, S. & Jeannerod, M., From action to interaction. *Journal of Consciousness Studies* 9:3-26.

Gallese, V. 2003, The Roots of Empathy: The Shared Manifold Hypothesis and the Neural Basis of Intersubjectivity, *Psychopathology,* 36:171–180.

Gibson, J. J. 1986, *The Ecological Approach to Visual Perception*, Lawrence Erlbaum Associates., Hillsdale, New Jersey.

Hatfield, E., Cacioppo, J. T. & Rapson R. L. 1994. *Emotional contagion*. New York: Cambridge University Press.

Husserl, E. 1966. Zur Phänomenologie des inneren Zeitbewusstseins. *Husserliana* Bd.10. Nijhoff, Den Haag.

Iacobini, M. & Lenzi, L .L. 2002, Mirror neurons, the insula, and empathy. *Behavioural and Brain Sciences* 25:39-40.

Jeannerod, M. 1994. The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, 17 (2):187-245

Kamachi, M., Hill, H., Lander, K. & Vatikiotis-Bateson, E. 2003, 'Putting the face to the voice': matching across modality. Current Biology 13, 1709-14.

Keysers, C., Kohler, E., Umilt, M. A., Nanetti, L. Fogassi, L., Gallese V. 2003, Audiovisual mirror neurons and action recognition. *Experimental Brain Research* 153:628-636.

Kivy, P. 1989, *Sound Sentiment*, Philadelphia , PA: Temple University Press.

Kivy, P. 2002, Introduction to a philosophy of music. Clarendon Press, Oxford.Kosslyn, S. 1996, Image and brain: the resolution of the imagery debate. Bradford Books, Massachussets, MA.Lhermitte, F., Pillon, B. & Serdan, M. 1986, Human autonomy and the frontal lobes: Part 1. Imitation and utilization behaviour: A neurological study of 75 patients. *Annals of Neurology* 19: 326-34.

Malmgren, H. 2000, Rorschach's idea of a "movement" response in the light of recent philosophy and psychology of perception. *Rorschachiana* 24:1-27.

Malmgren, H. 2004, Why the past is sometimes perceived, and not only remembered. *Philosophical Communications, Web Series,* 31.

Martin J. 1978, *El Arte Flamenco de la Guitarra*, United Music Publishers Ltd, London.

Martin, P. 1995, *Sounds & Society*, Manchester University Press, Manchester & New York.

Melzoff, A.N. & Moore, M.K., Imitation of facial and manual gestures by human neonates. *Science* 1987:75-78.

Merleau-Ponty, M. 1945, *Phénomenologie de la Perception*. Gallimard, Paris.

Meyer B. M. 1956, *Emotion and Meaning in music*, The University of Chicago Press.

Neisser, U. 1994, *The Perceived self : ecological and interpersonal sources of self-knowledge,* Cambridge : Cambridge Univ. Press

Port, R., Tajima, K., & Cummins, F., 1996, Self-entrainment in animal behaviour and human speech. *Online Proceedings of the 1996 Midwest Artificial Intelligence and Cognitive Science Conference.* http:// www.cs.indiana.edu/event/maics96/Proceedings/Port/port/.html.

Preston S., deWaal F. 2002, Empathy: Its ultimate and proximate bases, *Behavioural and Brain Sciences*, 25:1-20.

Rizzolatti, G., Fogassi, L., Gallese, V. 2001, Perspectives – neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews/Neuroscience* vol. 2:667-668.

Rizzolatti, G. & Craighero, L., 2004, The mirror-neuron system. *Annual Review of Neuroscience* 27:169-192.

Ryle, G. 1949, *The Concept of Mind.*

Scherer, K. R. 1998. Emotionprozesse im Mediencontext: Forschungsillustrationen und Zukunftsperspectiven. *Medienpsychologie*, 10:276-293.

Snyder, B., 2000, *Music and Memory*, MIT Press, Cambridge Massachusetts.

Stern, D., 1985, *The interpersonal world of the infant.* New York, Basic Books.

Stern, D. 1993, The role of feelings for an interpersonal self, in *The Perceived Self.* Edited by Ulrich Neisser, Cambridge University Press. 205-215

Stern, D. (1995), *The Motherhood Constellation.* Basic Books, New York.

Stern, D. 2004. *The Present Moment.* W.W. Norton & Company, New York, London.

Tagg, P. 1979. *Kojak…50 seconds of television music: toward the ANALYSIS of AFFECT in POPULAR MUSIC*, Studies from the Department of Musicology, Gothenburg. No.2

Thaut, M. 2003, Neural basis of rhythmic timing networks in the human brain. In: Avanzini, G., Faienza, C, Minciacchi, D., Lopez, L. & Majno, M., The neurosciences and music (Annals of the New York Academy of Sciences, 999), 364-373. New York Academy of Sciences, New York.

Trevarthen, C. 1992, An Infant's Motives for Speaking and Thinking in the Culture. In A. H. Wold (Ed.). *The dialogical alternative* , pp. 99-137. Oslo and Oxford: Scandinavian University Press/Oxford University Press.

Trevarthen, C. 1993, The self born in intersubjectivity: The psychology of an infant communicating, in *The Perceived Self*, edited by Ulrich Neisser, Cambridge University Press.

Trevarthen, C. Musicality and the Intrinsic Motive Pulse: Evidence of human psychobiology and infant communication, Special Issue of *Musicae Scientiae: rhythm, Musical Narrative and Origins of Human Communication,* 1999:157-213.

Zacksenhouse, M. 2001, Sensitivity of basic oscillatory mechanisms for pattern generation and detection. Biolofical Cybernetics 85, 301-11.

[i] The span of our experience of the present moment must not be confused with the duration of the experience itself. The experience may change almost continually, even if each experience includes intentions and protentions with a span of several seconds.

[ii] If this is true, dogs, obviously showing shame and guilt when they have done something wrong, are self conscious.

[iii] The concept Vitality Affect was first introduced in Stern (1985)

[iv] How do mirror neurons relate to neural oscillators? Most probably, mirror neurons reflect the integrated action of many adaptive oscillators which together constitute the primary interface with the other's movement pattern. But this is the topic of another paper.

[v] We will not discuss the use (especially in classical music) of small temporal deviations from these discretized expectations in order to produce specific aesthetic and/or emotional effects.