Motor-Mimetic Music Cognition

Rolf Inge Godøy

MUSIC IS IMPURE

Although music is usually considered a sonic form of art—as opposed to, for instance, various visual arts—it is quite clear that music appeals more to than just our sense of hearing. There are innumerable accounts of music appealing to, and making use of, various "extra-musical" elements such as visual images, colors, bodily movements, sensations of effort, touch, olfaction, etc., and this is reflected in the very many metaphorical terms used for characterizing musical sound, such as "bright," "dark," "slim," "fat," "sharp," "soft," "high," "low," etc.

If we were to postulate a theoretical distinction between on the one hand "pure" sound, defined as the acoustic signal (represented by various analog or digital means) and on the other hand "impure" sound, defined as the complex and multifaceted image we have in our minds in listening situations, I believe it could be argued that it is actually the "impure" that is primordial to the "pure," and not the other way around. In other words, the cooperation or interaction of the senses is "spontaneously" at work in any act of human perception and cognition. There is now a growing amount of evidence in the cognitive sciences, both in the domain of audition research [1] and in general [2], indicating that this sensory integration is not a secondary "by-product" of a "pure" sound stimulus, as has been the dominant Western view of perception and cognition in the past couple of centuries. In particular, I believe that we have good reason to suspect that images of sound-producing actions such as hitting, stroking, blowing, etc., play an important role in our images of musical sound, hence my idea here of motor-mimetic elements in music cognition.

ECOLOGICAL APPROACHES

Western composers have for centuries applied various "extra-musical" visual and/or numerical structural relationships in music, as in, for example, the famous contrapuntal principles of using melodies in original, reversed, mirrored, and reverse-mirrored versions, thus creating four versions of a melody arranged symmetrically around a central point. In the visual domain, such symmetries may be easily perceivable, but in music, it will in most cases be quite difficult to perceive the symmetry relations, as music is time-dependent and detecting such symmetry relations would require very clear cumulative visualizations of the tones. There are a number of other such "hidden" structural relationships used by Western composers and music theorists; common to them all is their violation of the principle of coherence in music by perceivable sonorous qualities.

Transferring structural or numerical features from one domain to another may thus in many cases result in some quite bad category mistakes, i.e. mistaking principles of organization in one domain as valid for other, totally different domains [3]. Rather than transferring any structural relationship directly from one sensory modality to another, what I therefore propose is to give a sketch of how music as a fundamentally cross-modal phenomenon can provide the basis for the relationship between the senses in an ecologically conditioned way.

The term "cross-modal" denotes the cooperation of the different sense modalities, and the term "ecological" denotes the idea that the human senses have evolved in interaction with the environment in service of our orientation and survival. Furthermore, "ecological" implies that there are a number of constraints on human perception and cognition, constraints that become all the more obvious when we compare human perception with what machine-based, artificial systems are capable, and not capable, of doing.

In audition research, it has become abundantly clear that artificial systems relying exclusively on information contained in the "pure" signal encounter great difficulties in tasks of recognition and orientation [4]. This is so because purely "bottom-up," data-driven systems without any prior knowledge, or previously established schemata, will not be able to untangle the complex fabric of sound in the real world, as for instance encountered in the frequently mentioned "cocktail-party" scene. Making sense of what is said in a noisy environment necessitates constantly making deductions from a large number of cues, as well as reconstructing or guessing what is going on in the face of degraded or missing information. In fact, it has been documented that information from one sensory modality can even override that of another, so that when people look at a video displaying the facial movements of a person pronouncing a different sound than what they are actually hearing, they may be convinced that they heard something other than what was fed to them in the acoustic signal [5].

Although research on cross-modality is a rapidly growing field [6], there are of course still many unexplored elements here. However, I have previously proposed a simple but, I hope, productive model of cross-modality at work in music perception and cognition. I called this a "triangular model" [7], because it depicts inextricable relationships between action, vision and sound in music perception and cognition. The basic assumption of this triangular model is this: Any sound
can be understood as included in an action-trajectory. Furthermore, this image of sound-production will have visual and motor components in addition to that of the “pure” sound, and I believe images of sound-producing actions can play the role of mediating between the visual and the sonorous, i.e. that actions can translate from the sonic to the visual and, conversely, from the visual to the sonic.

EXCITATION AND RESONANCE

In line with the “ecological” way of thinking presented here, it could be fruitful to separate two components of sound production that often are fused together. Sound production has an action component—what could be called excitation, as well as the reaction to the excitation—what could be called resonance. Separating excitation and resonance means not only separating the images of what we do from the images of the effects of what we do, but also separating ecological knowledge of action, i.e. motor and kinesthetic knowledge, from our knowledge of the material properties of objects and environments in our surrounding world. For example, in the case of a drum sound, this means separating our image of the action of the mallet hitting the drum from the image of the vibrating membrane, shape, size, material, etc., of the drum, or in the case of a trumpet fanfare, separating the image of energetic blowing from the image of the mouthpiece, cone, tubes, etc., of the trumpet.

This separation of excitation and resonance is useful for many purposes in music, such as in orchestration and in musical analysis, but most of all in providing us with a possible clue as to how music is remembered and efficiently “re-coded” [8], and may help us better understand the flexibility and dynamic nature of musical imagery [9]. In particular, there seems to be evidence for motor coding as an essential element in musical memory [10]. As for resonance, there are several manifestations of the vast repertoire of “ecological” knowledge of resonant properties that we seem to possess, accumulated through a massive process of learning from birth (or perhaps even before that), both in everyday situations [11] and in more music-specific situations [12].

One consequence of separating excitation and resonance in our context is that we more clearly see the mediating role of action: Action “translates” between the senses, something that remarkably has been observed earlier by various artists and scientists [13] and seems to be well documented in contemporary neurological research [14]. Alain Berthoz reminds us of the famous line from Goethe’s Faust, “In the beginning was movement,” and demonstrates furthermore how action is an ingredient in most cases of perception as well: “Perception is not representation: it is an action simulated and projected upon the world” [15].

ACTIONS

There are different kinds of actions associated with music, for instance in dancing and marching; in this context, however, I will focus on the actions associated with the production of musical sound. There are two main groups of sound-producing actions, which I like to call ballistic and sustained. The term ballistic is here taken from the study of human movement [16] and denotes a brief, concentrated effort immediately followed by a phase of relaxation, as is the case in hitting or kicking and in performance on percussion instruments and certain keyboard instruments such as the piano (which, strictly speaking, is a kind of percussion instrument). In ballistic sound production, the moment of impact is immediately followed by a shorter or longer period of energy dissipation in which the sound usually has a decaying shape. Sustained sound-producing actions denote the more continuous effort required in blowing, singing and bowing, and the shape of the sound may just as well be flat or even ascending (i.e. having a crescendo) because of the continuous transfer of energy in the sound-producing actions.

Furthermore, sound-producing actions may either be singular, such as in a singular beating of a drum or in a long, sustained tone on a flute, or may be concatenated into more composite actions, such as several hittings in a drum roll or several sustained tones in a long melody. Interestingly, such concatenations of singular actions may make the singular actions fuse into hierarchies, just as in daily-life actions several finger movements fuse into more superordinate actions, e.g. typing on a computer keyboard, turning the pages of a book, lifting a glass of water, etc.

Our images of actions, be they singular or complex, are called motor programs. A motor program is an overview image or a script of what is to be done when, where and how. An image of how to play a certain tone on a cello, a phrase on a flute or an entire Beethoven sonata on a piano are all instances of motor programs, just as the images of how to turn on the light or walk to the park are motor programs. One important feature of motor programs is that they are flexible in the sense that they may exist at variable degrees of resolution and velocity, e.g. I can visualize my walk from here to the park either in great detail, including all the steps to be taken, or in a very fast and low-resolution version, where I only visualize certain turns or landmarks along the walk.

MOTOR-MIMESIS

The influence of our images of sound production on our perception of sound is sometimes referred to as the motor theory of perception. It was first used in linguistics as it became clear to some researchers that a purely signal-based model of perception was not going to work and that it would be more fruitful to suggest that listeners also make an internal image of how the sounds were assumed to be produced [17]. The motor theory has been controversial, but owing to advances in neurological research, there is now increasing support for the idea that motor faculties of the brain are indeed involved in most kinds of perception and cognition, which has led Berthoz to suggest that we replace the term “representation” with the term “simulation” to denote what is going on in our minds [18]. This means that there is an incessant simulation and reenactment in our minds of what we perceive and a constant formation of hypotheses as to the causes of what we perceive and the appropriate actions we should take in the face of what we perceive.

I believe this points in the direction of what I would like to call a motor-mimetic element in music perception and cognition, meaning that we mentally imitate sound-producing actions when we listen attentively to music, or that we may imagine actively tracing or drawing the contours of the music as it unfolds. Although there are many unexplored aspects of such a motor-mimetic theory, I believe there is now enough material to support this as a hypothesis as well as a hopefully productive explanation for the cross-modal workings of music, which may be summarized as follows: Motor-mimesis translates from musical sound to visual images by a simulation of sound-producing actions, both of singular sounds and of more complex musical phrases and textures, forming motor programs that re-code and help store musical sound in our minds.

In some cases, or for some people, the reverse may also be the case, i.e. that...
motor-mimesis can translate from visual images to sound by re-tracing the visual contours as sound-producing actions, “sonorizing” visual images. There is of course also the possibility of mentally tracing or drawing the resonance phase of the sound (i.e. after the onset of the sound), making more or less “subjective” renderings of the evolution of the sound or the tone color based on vocal imitations and/or other bodily based notions of shape, as found in common metaphors such as “open,” “closed,” “narrow,” “wide,” etc. [19]. All such mental tracing of musical sound would be in accordance with the motor-mimetic theory. Such reenactments of musical sound from the first person, egocentric and “subjective” point of view are in my opinion not only the privileged strategy for becoming intimately aware of richness of features in musical sound, but also the privileged strategy for relating musical sound with other forms of art.

References and Notes
18. Berthoz [2].