A PERSPECTIVE ON THE LIMITED POTENTIAL FOR SIMULTANEITY IN AUDITORY DISPLAY

Joachim Gossmann

UC San Diego
Center for Research and Computing in the Arts
9500 Gilman Drive La Jolla, California 92093-0037
jgossmann@ucsd.edu

ABSTRACT

The auditory environment has been described as a *biased competition*: The juxtaposition of an array of pre-formed auditory streams and a process of attentional selection [1, 2]. The orientation of attentional selection toward environmental streams is differentiated towards different *modes* of streaming: Speech, music and sound effects are only three examples in a potentially open polymorphism of *perceptual strategies* through which we access the sounding world.

This differentiable-simultaneous manifold of environmental streams allows perceptual participation only within a certain number of processes at the same time—only one speaking voice, one sense of ”harmony”, a single ”rhythm”, and so forth.

We propose a re-basing of sonification strategies not on the definition of external mechanisms, but on the definition and application of new *modal strategies* that are circumscribed and accessible through *what is not possible to perceive at the same time*.

1. INTRODUCTION: THE DIFFERENTIABLE-SIMULTANEOUS MANIFOLD

The phenomenon of *multiple parallel channels of information* encounters on many structural levels in time-based media artifacts: The distinct intertwined voices of contrapuntal music, the parallel polymorphism of dialog, music and sound effects projected onto the audience from a multichannel loudspeaker array in movie soundtracks, radio drama and news reports that combine location-sound with added voice-over, and the two ears that we both hear with at the same time. We find ourselves addressed by representations and expressions of a multiplicity of simultaneously present streams, objects and events. Auditory media, which unfold exclusively in temporal developments, seem to imply the potential to display a manifold of simultaneous signals and processes to the participant. But before we can approach a structural description of the phenomena of perceptual simultaneity, we should first generate transparency in an area of potential misunderstanding.

1.1. The distinction between audio channels, sensory channels and environmental streams

We can distinguish three structural levels on which we find arrays of parallel streams:

1. the *array of audio channels* that are stored and transmitted by the medium and projected by the loudspeakers or headphone
2. the *sensory array* of the participant
3. the *manifold of environmental streams* that make up the auditory scene the observers and participants find themselves immersed in

Evidently, we find the polyphonies we experience in the audio content itself (layer 3 in this model) encoded and transmitted through layers 1 and 2. However, each of these connected layers is characterized by the a potential for structural independence. Especially the relationship between a loudspeaker signal and an environmental stream is a source of potential confusion. We usually do not encounter the voices that make up a musical polyphony projected from distinct physical sources, channels or spatial locations—a string quartet represented by four discrete loudspeakers for example. Instead, the count of transmitted and projected media channels tends to conform to the properties of the sensory array of the participant—stereo loudspeakers, headphones, (video screens in the audio-visual case, sometimes with two simultaneous images, one for each eye). But we are increasingly confronted with cases in which the count of discrete audio channels that are projected from loudspeakers is greater than the number of ears in a listener's head. We can shed light on this by describing the *environmental role* of a loudspeaker as an interpolation within a structural triangle with the following corners:

- The audio channel projected from a single loudspeaker is a stand-in for an *environmental stream*.
- The audio channel is directly connected to one of the ears of the participant as a *sensory channel*, e.g. by headphone.
- The channel is part of a multi-channel array to be projected from loudspeakers that are each heard by both ears. Spatial impressions are encoded in inter-channel signals.

We find the first case realized for example by the projection of film dialog from the center channel in order to constrain the localization of the actor's voices to the center of the screen. The second case conforms to the binaural application of sound to the listener's ears via headphones, and the third case is found in all loudspeaker arrays that surpass the two stereo channels in number, such as the cinema and home-theater audio formats promoted by the movie industry (5.1, 7.1, 9.1, et cetera). and finds its most extreme realization in wavefield synthesis systems in which a single loudspeaker is never heard as a discrete *sound-source* on its own and instead always appears as a contributing element in the synthetic creation of an environmental sound field. A more detailed investigation into the relationship environmental streams, audio channels and the sensory array of the participant needs to be topic of a future publication.

1.2. The Auditory Scene: Stream formation and selection or *perception-as-action*?

The process by which acoustic energy that arrives at the ear is transformed into auditory experience is the concern of psycho-acoustics research. The description of principles and processes involved in the formation of objects and streams in the perception of time-based content can be approached from a variety of perspectives. A very influential school of thought in the area of perceptual object formation are the *Gestalt Principles of Perception*, a set of rules and tendencies that seem to underlie our structural interpretation of the environment—the emergence of forms, boundaries, shapes, foregrounds and backgrounds and so forth [3]. While Gestalt Psychology has its origin and focus in the analysis and description of *visual perception*, we can interpret A.Bregman's well known work on *Auditory Scene Analysis* as a correlate for auditory domain [1]. Similar to the grouping principles of gestalt psychology, Bregman sees auditory perception as a process of fusion and segregation that results from properties and features of the acoustic signal: On the one hand the fusion of perceptual elements depending on their spectro-temporal structure (harmonicity, common onset/offset, common fate in the frequency or amplitude domain, et cetera), and on the other hand the linking of distinct events into perceptual streams depending on their similarity in auditory *feature-spaces*: For example, the distinct timbre- and pitch-spaces of a flutes, violins, birds and cars cause them to segregate into distinct perceptual objects and continuous perceptual streams. Here, spatial location is one factor among others.

It has been argued that the role of the *perceptual object* is not sufficiently described as a bottom-up coagulation juxtaposed to the process of attentional selection, but that there exists an important infusion of low-level stream segregation by cognitive processing, and that the *objects of perception* can in fact simultaneously be regarded as a basic unit of both cognition and attention [4]. In the psycho-acoustic domain these relationships are being investigated in the work of B.Shinn-Cunningham [2].

Another approach to the structural interpretation of perception occurs in the wake of the theory of environmental perception established by J.J.Gibson [5]. Gibson avoids the bottom-up and top-down structures of gestalt theory and instead sees perception as a *direct* process that dispenses with the differentiation between the stimulus, the environment and its perception. Alva Noë in turn interprets this direct perception *as action*—the involvement of the participant's body in a direct performance of perceptual enactment [6].

From these diverse backgrounds, we can consider the segregation of perceptual objects, streams and behaviors that are available to selection by focus and attention not only as the outcome of a feature-based coagulation, but also as inference of patterns and expectations by the observer and finally, following Noë, the activation and involvement of specific *perceptual strategies*: In the context of this presentation, we would like to address this conceptual fusion between the formation of *perceptual streams and objects* and the involved strategies of it active perception as the an outward perceptual activity of *modal streaming* that is performed by participants. Perceptual involvement with media displays can be regarded as an application of modal strategies by which participants discover, approach and become involved with the environment. Modal streams are distinct from *sensory streams* as they can alternatively span multiple sensory modalities or become segregated within a single sensory stream—but also in distinction from *perceptual streams* that emerge from a bottom-up fusion of sensory stimuli. What we mean by *modal streams* is the performance of a perceptual strategy by the perceiving participant in a continuous process of active perception in the senses of Noë —a perceptual involvement the participant might be unaware of [6]: Both the conscious effort of looking up a youtube video and involuntary eye movements in the observation of an image can be regarded as aspects of a *modal strategy of active perception*.

1.3. The simultaneous manifold

In audio-visual media, perceptual objects and streams can span multiple sensory modalities: A car driving by, people talking in the background, a record player playing diegetic (in-scene) music, et cetera. We experience independent simultaneous multi-modal objects that form relationships and groupings, a whole that consists of simultaneous parts: Our experience of a time-based media artifact could be described as a *differentiable simultaneous manifold*.

As we attend the multiple seemingly independent entities that occur in juxtaposition, superposition and sequence within the mediated content, we tend to become oblivious to the technological transmission channels or the way the media system addresses our sensory channels we have described in 1.1. And instead become immersed in a mobile panorama of perceptual objects and streams that is at the same time *coherent* and *navigable*.

While the strict definition of attention allows the perceptual selection of only a single object or stream [2], the perceptual simultaneity of distinct but coherent perceptual streams we encounter in auditory media suggest that the *shape of what we can attend to simultaneously* is wider than a single *perceptual object* or *auditory stream* in the definitions of Bregman and Koehler.

Evidently, our potential for simultaneous perception is characterized by limitations. Barbara Shinn-Cunningham describes the middle-ground between perceptual object formation and attentional selection as a *biased competition* that is decided either by the volition and attentional direction of the perceiver or the salience of the perceptual object. Following the idea of perception as combination of simultaneously activated *modal strategies*, we may describe these potentials for simultaneous perception as a repository of perceptual resources that is available to the observer.

2. PERCEPTUAL STREAMS AS PERSISTENT PERCEPTUAL INTERFACE

Auditory streams in the sense of Bregman are characterized by a dichotomy of *mobility* and *persistence*: On the one hand, the stream itself persists over time and is attributed to or accountable for the emergence of persistent objects within our environment. On the other hand, its appearance can change and modulate, and its variability has the potential to encode information within itself: A speaking voice, figuring prominently in the famous auditory scene example of the *cocktail-party* [7], is characterized by a persistence that allows the party guest to navigate the auditory scene with their attentional focus. But the interior, the *content* of the stream is characterized by variability: What is being talked about, how it is being said, the specific sounds of vowels, consonants, phonemes, how the physiological performance of the speaker contextualize the individual voice, et cetera: The modal stream can be interpreted as an *interface* that allows the discovery of previously unknown aspects and properties of the environment. Upon closer inspection, streams can in turn disintegrate into a manifold of independently observable features: Streams within streams, accessible within one another through progressive attentional disclosure as it was described for example in Merleau-Ponty's phenomenological analysis of perception [8].

As a *perceptual interface* toward our environment, modal streams provide us with an access of relative persistence through which we provide attention to environmental processes. In this way, we can see them as a bidirectional relationship: On the one hand, they form a channel through which environmental information reaches us, on the other hand, a pre-set strategy to interpret the environment is already implied in the establishment of the stream itself.

3. APPROACH FROM INSIDE: PERCEPTUAL RESOURCES

Multiple streams can be present in our environment simultaneously, but often we can not attend all of them at the same time: We see ourselves surrounded by opportunities to involve our perception and action, but we can only realize a very limited subset of them at any given time. In cognitive science, we find this formalized as a juxtaposition between an array of disclosed perceptual objects and streams on the one hand and the process of our shifting attentional selection on the other hand [9, 2].

However, we need to acknowledge that in the pre-attentional formation of perceptual objects the a *type* of object is already defined, and moreover, these different *phenomenological types* of streams are characterized by a different potential to be attended simultaneously. More than a general *sensitivity for sound waves*, hearing involves an a priori *listening-for*, a perceptual top-down pre-organization, and it appears to be that different types of listening engagement are characterized by a varying potential for simultaneity, to be occurring in parallel or at the same time as other engagements.

For example, it seems evident that we only have the potential to fully engage and understand a single stream of type *speech*. Multiple simultaneous language streams will lead to a discrimination of the streams into *attended* and *peripherally attended* speech—or, if that is not possible, confusion and unintelligibility are the consequence. We find an even more extreme case in music, in which the addition of a second music stream into the environment leads to an effective destruction of the music with only very limited potential to selectively attend one of the coinciding streams. Then again, we seem to be able to let multiple different non-speech environmental sounds occur simultaneously without a similar destructive effect. In a structural analysis of these relationships, we can distinguish the following cases:

3.1. The navigable multiple and polyphony

3.1.1. Navigable multiple

As we can see in the example of the cocktail party, perceptual streams can form a *navigable multiple*: While not all streams can be attended simultaneously, the streams are still accessible to participant's select and engagement. We can only attend to one conversation at a time, but which one is up to our attentional navigation of the auditory scene.

3.1.2. Parallel simultaneity and polyphony

In certain cases, modal streams can become accessible in parallel simultaneity: We can experience a collection of streams in simultaneous superposition while they still retain their own identity and potential for an increase of depth of attention. We can see an example in the potential of speech and music to be present simultaneously—as opposed to the superposition of two *musics* or the presence of two speaking voices simultaneously which is immediately characterized by conflict. We can compare this to *musical polyphony* which represents another example: In a 4-part fugue, the voices retain independence to an approach of analytic listening, but cohere to form an aggregate: Attentive selection may shift between focusing on a single stream or the global perception of the harmonic relationships resulting from their combination. The layers of a movie soundtrack can be seen as another example: Each of the layers of the soundtrack—dialog, music and the various sound effects—is characterized independence that allows them to be created by different production teams, can reside in a different phenomenological area as Michel Chion describes in his book *Audio-Vision*[10]. Nevertheless, a coherent experience is created that has the potential to subsume the individual constituents within it. In contrast to the *navigable multiple* from which the participant can freely pick streams to attend, we can call this case in which distinct streams form a new coherent whole the *polyphonic multiple*.

But next to the formation of navigable and polyphonic manifolds, perceptual objects and streams can also merge or obstruct each other.

3.2. Correlative merge

If modal streams contain correlated behaviors this may result in their perceptual fusion into a single more complex stream or group of connected developments. This is the case for example for complex sound objects or audio-visual coherence in the context of cinema sound (for example, a car drive-by).

It is important to note that while correlative effects occur within our perceptual environment, for example the micro-correlation between a sound source and its reflection that leads to the encapsulation of the reflection into the *spatial timbre* of the sound source, correlation can also be discovered as an effect of self-motion: We may hypothesize that the impression of spatial persistence, for example of architecture, could be interpreted as an effect of correlation between the self-motion of a participant and the perceptual change in the appearance of the architectural environment. The merging of perceptual elements that show correlated behavior is in accordance with the rules governing the perceptual fusion and segregation of streams [1].

3.3. Destructive merge

The destructive merge is an everyday experience: Streams mingle together and overlap making each other mutually indistinguishable, comparable to two layers of handwriting written in top of each other. For example the projection of two speaking voices from the same loudspeaker, or the simultaneous presence of two violin sonatas usually lead to a destructive merge of the simultaneous streams.

In the hierarchical perspective of bottom-up and top-down formation of perceptual objects, the mutual obstruction of perceptual objects and streams can occur on any level of formation or attentional selection—from *energetic masking* in the sensory channel to various effects of *informational masking* or failure in attentional selection [2]. Coming from the perspective of direct perception, we can describe the mutual merging and masking of modal streams as *perceptual resource conflict*. Like the navigable and polyphonic manifold, we can interpret merging and masking as a structural dependence and relationship between the perceptual resources that we apply to different aspects of the environment over time.

4. INTERLUDE1: PITCH, SPECTRAL MORPHOLOGY AND THE MODAL STRATEGY OF MELODIC LISTENING

A popular example of perceptual fusion is the phenomenon of instrumental timbre. As we know, the perception of timbre is related to the amplitude and phase relationships of partial frequencies that are connected by a *common fate* in frequency and amplitude. Preferably, the partial frequencies have *harmonic* ratios [11].

But beyond the emergence of *pitch* and *timbre* as independent categories, we might say that to hear a sound as a musical note, as an element within the context of a melody, is more than just an effect that emerges from a partial relationship within the signal itself. Music implies a self-application of the participant to the melody through a strategy of *melodic listening*. What we mean by that is exemplified in the *speech-to-song* illusion described by Diana Deutsch[12]: A repeated fragment of spoken word is initially approached with a strategy of *speech listening*. Upon multiple repetition, the strategy shifts, and what is heard becomes more and more a sung melody. The signal has stayed the same, what has moved is the listener. We can say that the strategy of *melodic listening* we apply to music in fact determines our attitude and thereby our interpretation of the music.

In the opposite direction, we can also find musical examples in which our—intuitive or trained—strategies of *melodic listening* have been intentionally subverted: If the harmonicity of the spectrum or the common fate of the partials is disturbed, the fusion into a sound characterized by a single pitch and timbre can break up and begin to sound bell-like: We may hear *multiple simultaneous pitches* within a single sound, especially if we have trained ourselves to navigate such frequency mixtures. If furthermore the common fate of the partials is disturbed, the experience of the sound can split up into even more independent entities all together.

A music piece in which these effects can be experienced in an exemplary way is Karlheinz Stockhausen's piece *Cosmic Pulses*in which sound layers, clearly delineated by a common fate in the area of frequency, amplitude, spatialization, develop interior worlds due to the inharmonic *split* spectra and the micromodulations within the spectral composition of each layer: An unsettling experience as we find our modal approach to the hearing of sound constantly challenged and on the edge of disintegration, all the while new layers are piled atop one another [13]. In his own words, Stockhausen admits that one might not be able to attend all contained streams during one individual listening run:

> If it is possible to hear everything, I do not yet know–it depends on how often one can experience an 8-channel performance. In any case, the experiment is extremely fascinating! [14]

5. PERCEPTUAL RESOURCES: LISTENING AS SELF-APPLICATION

We often find music tracks organized into a *playlist*, the reason being that we are generally unable to appreciate two musics playing simultaneously—we prefer to attend them in sequence. When we superimpose two *musics*, they usually do not combine *navigable multiple*. While details of each music track remain accessible to attentive selection, others merge into a combined perception that appears not so much a summation of its parts but a different experience in itself. We may pick up on familiar instrumental timbres, vocalists, melodic fragments and recognizable moments of each music even when it is superimposed with another music, but certain aspects become very hard or even impossible to perceive when presented in temporal coincidence. To pull it down to a common sense statement: Music is a time-based art and lives from the fact that elements are presented in succession, with specific duration, intensity—and the attentive presence of the listener.

While simultaneous melodic lines for example can add up to a navigable polyphony—whether this occurs in the confines of musical meter and harmonic counterpoint as in Bach's music or as a stochastic and *chaotic* process one such as in Xenakis or Ligeti shall be another question—but it appears that only one sense of *harmony* or *tonality* seems to be possible at any moment: If multiple harmonies coincide, we do not hear both at the same time. In the case of harmony, we also have difficulties to listen to them as navigable parallel presence in the same way that we might attend to two talkers at a cocktail party. What emerges is a new *bi-tonal* harmony—a new tonality in itself.

We can find a similar behavior in the perception of rhythm. If two different repetitive rhythmic structures coincide, we seem to be unable to hear them as two separate rhythms at the same time. In some cases they might form a navigable multiple if they can be attributed to different modal streams, but more often they will combine into a new rhythmic structure. Even while we might be able to discern what meter each music piece is by selectively attending to individual instrument timbres if one of the coinciding *musics* is characterized by repetitive patterns, the overall impression of the rhythm will be lost.

The phenomena of *harmony* and *rhythm* contain phenomenological aspects that resist the formation of a *navigable multiple* or even a *polyphonic multiple*. We can describe them as *perceptual resources*: A limited potential to simultaneously attend to environmental phenomena. The musical features of *harmony* and *rhythm* are akin to our ability to only attend to one language stream at any given time, albeit with different structural demands on simultaneity and another navigation strategy for the participant. While cocktail parties encourage a manifold of simultaneous conversations, there usually is only a single music track playing in the room. Our listening can handle a coincidence of rhythm, harmony and an environment of navigable conversations, but not two incoherent harmonies and rhythms. [1]

[1] The first modern composer to exploit the collision of different harmonies and rhythm was arguable Charles Ives who is known for experimenting with marching bands performing pieces of different harmony and rhythm while marching through his home town—an experience he would later emulate in the polymetric sections of his symphonies.

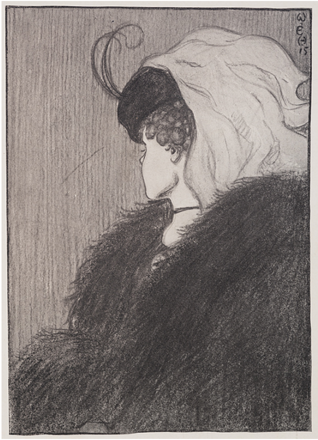

Figure 1: You can shift between seeing an old or young woman in this famous image [15]. However, it appears problematic to see both at the same time.

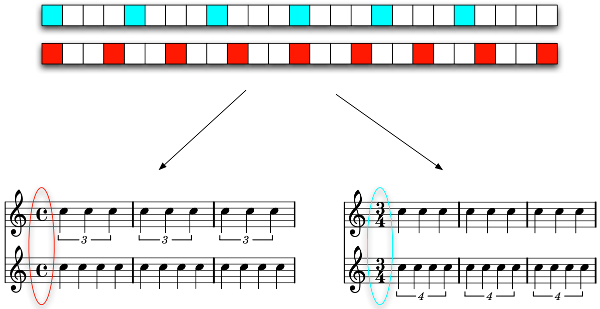

Figure 2: You can shift between hearing this time sequence as 4/3 or 3/4 polyrhythm—as two distinct rhythms occurring in 4/4 or 3/4 time respectively.

6. POTENTIAL ORIGINS OF MODAL STRATEGIES

The different morphology of the modal strategies involved in the perception of speech and music begs the question what origin they can be traced back to.

Of course we have to assume that the establishment of *see-as* and *listen-for* patterns that underlie these phenomena is subject to continuous improvisational adaptation, optimization and intuitive experimentation. Our taste in music changes, as does our perspective on all other aspects of life. One way to describe this open epistemological field is the area of the *cognitive body* I have described in [16]. However, we could for example list three potential channels through which modal strategies could emerge: *Learning and experience, evolutionary development and emergence*.

Evolutionary we can assume that basic modal strategies are made available to us through an expression of our genes. For example, our basic sense of hearing—the potential to perceive sound in general can be attributed to the fact that we have ears which evidently evolved through natural selection. In this area there are also the physiological and neuro-physiological properties of our body that can become an active element in the task of perceiving sound—for example the experience of groove. In his book *Sweet Anticipation*, David Huron traces musical experience back to the evolutionary history of auditory processing the central nervous system [17].

Emergent modalities address us from a stream of perceptual events that enters our perception from our environment: Something catches our attention without a clear pre-formed interpretation or expectation: There is an a-priori sense and experience of *potential meaning* in the experience of the signal, motivating a process of attentional observation which leads to the accumulation of hypotheses, inferred persistencies like patterns, objects and agencies: The self-organizing emergent collection of assumed and expected underlying behaviors. This can immediately be observed in the process of *listening to music*.

A *learned* modality can be seen in the ability to attend speech: While we might be endowed with an innate, potentially *physiologically pre-disposed* [18] tendency to attribute meaning to reoccurring sound patterns, the specific language we speak comes toward us from the environment we grow up in—the interactions we have as children with our environment. We might say the speech channel emerges in a self-driving process of improvisatory rehearsal by a continued contribution of trial, error, conscious effort in production and attention.

7. INTERLUDE2: POLYRHYTHMS AND THE SHIFT OF PERSPECTIVE AS PERCEPTUAL SELF-APPLICATION

The strategies by which we listen to our environment are characterized by a degree of conscious control. We can see this in the case of polyrhythm perception. The perception of polyrhythms is split into the perception of a *primary beat* that conforms to the perceived *meter* of the rhythmic structure, and a *secondary beat* which is heard as being offset or as “standing against” the primary beat. While the temporal structure of the events themselves stay identical, listeners have the potential to consciously navigate between different listening perspectives on the polyrhythm by applying the modal strategy of the *meter* to each of the two layers, shifting the way the polyrhythmic stream of beats. We can compare this process to way ambiguous images appear, for example the famous picture that can be seen as an old or a young woman, depending on the way we apply our strategy of *seeing a face*. In both cases, we can not take both perspectives at the same time.

8. MUSIC, SPEECH, THE *NATURAL ENVIRONMENT* AND SONIFICATION: DISTINCT MODAL STRATEGIES

Taking a closer look at the activity of listening to music, speech and sounds from the natural environment, we can distinguish different relationship of the activity and the participant: We find *modal strategies* in the interpretation, approach, following and tracking of the sound and what is encoded within it that imply a different kind of involvement.

8.1. Environmental sound

When we are immersed in natural sound scenes, we are experiencing sounds in their natural state, as an *identity of the sound with its source*. Unlike speech and music, which are strategies used by human beings to target the perception of other human beings in order to achieve a specific effect, the sound caused by the wind in our ears is a property of the air and the wind. Animal sounds are an aspect of the animal. The presence of water is announced by its specific look as well as its characteristic sounds, et cetera. Of course it has been argued that the perceptual approach toward our *natural* environment has been developed and optimized in the process of evolution, and a perceptual theory that underlines this identity of perception and the environment can be found in J.J.Gibson's work on environmental perception [5]. From this perspective, musical listening tends to appear as a secondary category—a *cheesecake of the mind*[19], and speech listening becomes yet another even more extraordinary involvement.

8.2. Music

Music is generally expected *to produce a desired effect by itself*, without any analytical effort of the participant. What we hear is not experienced as property of the external environment, but an emotion, meter, rhythm, melody, et cetera, that emerges within an inherently *human way of listening*. Arguably, listening to music is not an involvement with the outside world but in fact with our own potentials of having an aesthetic experience. In order for music to appear, the participant has to provide specific perceptual resources—for example what we have previously circumscribed as the potentials for *harmonic* and *rhythmic listening* or the potential to experience sublime emotions as laid out by David Huron [17]. We could describe the musical experience as a *massage* of these resources, and the participant has little more to contribute than to remove potential distractions from the environment to make sure nothing else will occupy the required perceptual potentials and thereby *mask* and *occlude* the musical experience. As we accumulate experience throughout our lives, new perceptual resources form, and our taste of music changes: We can continuously discover new and interesting aspects in music, however, when the music *doesn't work*, when it causes dissatisfaction or confusion, we usually do not blame ourselves: The composer, the performer, the sound engineer or the home stereo is at fault, while our ability to listen to and enjoy music is often considered an innate aspect of our humanity.

8.3. Speech

Speech on the other hand is very obviously an *acquired* perceptual strategy. We are not born with the language that our parents speak, and we have to learn both the production of speech as well as its understanding: Native language is acquired through attention, rehearsal, repetition, optimization, reflection, trial-and-error, adaptation, et cetera. Listening to language is evidently the involvement of a specific learned resource of the participant: We can only do it for one speaker at a time. In speech, the difference between the transmission channel and its content becomes evident: The fact that a person is talking is to a large degree independent of what they are going to say. The involvement of decoding language has a degree of independence from the circumstances the language is heard in—even though we take the situation of what is being said into account.

8.4. Sonification

When we interpret sonification not only as a strategy to organize, create and render sound, but inversely as a *modal listening strategy* or, to put it simpler, a *way of listening*, we can see how it is different from environmental sounds, speech and also music:

In comparison to *natural environmental listening*, sonification necessarily has to communicate its data by using properties of sounds that are *inherently detached from their source*. As such sonification is comparable to a learned listening strategy like language. It is designed to target our perceptual potentials in a specific way, but in order to *encode something other than itself* in a similar way speech or a technological media channel would.

This involvement of the listener *to see something in the sound which is not itself* is also a difference between sonification and music. To Paul Vicker's dichotomy of *sonification concrete* or *sonification abstraite*[20] I would like to add that it is not sufficient to place the accountability for the appearance of sound into the human strategy for sound/music-generation alone. This would be comparable to placing the accountability for the meaning of speech only into the act of speaking while disregarding the involvement of *understanding*.

When we listen to Xenakis, John Cage and Alvin Lucier, we may indeed hear something that is comparable to *sonification heard as music*. The use of data appears as an element subverting the continuum of intentionality that is seen to reach from the composer to the experience of the music listener in order to evoke *open potential* in the participating listeners can be seen in the context of a larger cultural context of this era, as outlined by Umberto Eco's idea of the *Open Work* [21]. A further superficial kinship is generated in the sense of *unfamiliarity* and potentially *initial discomfort* that results from the fact that this strategy of *New Music* and sonification require ways of listening that are unfamiliar to the listeners of speech, natural environments and pre-20ieth century music.

But it is evident that the relationship between the sound and the listener as well as within the listener's involvement is very distinct: In the first case, a composer is exploring a strategy of generating an *aesthetic experience within the sound and its performance itself* that appears as new and unfamiliar to the listener. The plan is to invoke the curiosity of the listener and tap into our innate tendency to react to new experiences in our environment with the development of a complementary listening strategy: We always want to make sense of the world of course, we want to know what's going on, so we reach out and gather around what we do not understand.

The end state of successful sonification however is that the sound, or any aspects of a musical experience in fact *vanish from the listeners perception*, and what shines up behind the auditory transmission of information are the data that underlie the sonification: The listener is not consciously involved in listening to sound, but becomes connected to the data and relationships that are encoded within it, in a similar way that the listener of speech become oblivious to the sound of phonemes, and the pitch of the voice, and instead focuses on *what is being said*—a process we saw reversed in Deutsch's Speech-To-Song Illusion [12].

The sound features become an intermediate encoding step in the communication of data, and the experience is mediated by music, but in the end primarily non-musical: The difference between *message* and *massage* in the sense of McLuhan [22]. Whether the sounds embodied in this process are derived from sound-making properties of our natural environment or electro-acoustic *acusmatic* sound that has no other source than a loudspeaker [10], or whether the sound properties share a kinship to *musique concrete* or tonal music—even whether the sound is comfortable, aesthetically pleasing, beautiful et cetera—become secondary criteria similar to whether the sound of the announcer's voice on the train platform is pleasant to listen to.

That being said, evidently *New Music* has shown is the way of opening up musical accountability to *non-intentional* elements such as data values and thereby created a bridge for listeners to open their ear to the qualities of *sounds detached from their cause*, and this achievement is of course interesting to acknowledge from the perspective of sonification. In a previous publication we have argued that referential sound, for example the famous use of *piano samples* as carriers of pitch information, can lead to a loss in perceptual detail—the technological transformations that lead to the formulation of *musique concrete* have shown us the way how to *listen to spectral qualities* of sound and thereby made a new perceptual approach possible. In this sense, we might indeed be able to *let music shows us the way*, but the focus has to be the activity of the listener and participant.

What makes the world behind the sound appear is the listening strategy of the participant, the artist and composer ideally becomes as invisible as the *designer of a language*.

9. SUMMARY: DISCOVERING THE MODAL STRATEGIES OF SONIFICATION THROUGH THEIR *POTENTIALS FOR SIMULTANEITY*

I derived the concept of *modal strategy* from a structural description of our potential to appreciate simultaneous multitudes of specific kinds of processes in our environment—speech, harmony, rhythm are three examples. From this position I argued that listening is characterized by specific potentials for simultaneity that are inherent in the perceptual approach toward our surroundings, for example the ones listed in 3.1.

From here we may ask: What needs to be *moved out of the way* if a sonification strategy should be perceived successfully? Do sonification strategies allow to be perceived simultaneously (like music and sound effects), or do they mask each other? What is the specific domain the competition, collision or masking occurs in—is the masking *energetic*, *informational*, or inherent in the the activity of *participation*, such as attentional selection, focus, following and other aspects of *perception-as-action*? Under what circumstances can a sonification strategy generate a *navigable multiple* or *polyphony*?

I expect that an inquiry from this participant-centric perspective will in fact lead to more successful sonification designs that, insted of placing the accountability into the mappings and modals of data are motivated by a participant-oriented interest in *auditory scene synthesis*—a line of work that is already in process in the developments of stream-based sonification [23].

Through the development implementation and application of new modal listening strategies sonification can become an auditory interface that allow the active involvement of the participant, enabling them to experience accountable structures and perceptual properties far beyond an experience of *sound modulated by data*.

10. REFERENCES

[1] A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*. The MIT Press, Sept. 1994.

[2] B. G. Shinn-Cunningham, "Object-based auditory and visual attention," *Trends in Cognitive Sciences*, vol. 12, no. 5, pp. 182 – 186, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364661308000600

[3] W. Köhler, *Gestalt Psychology*. Liveright, New York, 1947.

[4] J. Feldman, "What is a visual object?" *Trends in Cognitive Sciences*, vol. 7, no. 6, pp. 252 – 256, 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364661303001116

[5] J. J. Gibson, *The Ecological Approach To Visual Perception*, new edition ed. Psychology Press, Sept. 1986.

[6] A. Noe, *Action in Perception*. The MIT Press, Mar. 2006.

[7] E. C. Cherry, "Some experiments on the recognition of speech, with one and with two ears," *The Journal of the Acoustical Society of America*, vol. 25, no. 5, pp. 975–979, 1953. [Online]. Available: http://link.aip.org/link/?JAS/25/975/1

[8] M. Merleau-Ponty, *Phenomenology of Perception*, 2nd ed. Routledge, May 2002.

[9] H. E. Pashler, *The Psychology of Attention*. Cambridge, Mass: MIT Press, 1998.

[10] M. Chion, C. Gorbman, and W. Murch, *Audio-Vision*. Columbia University Press, Apr. 1994.

[11] C. Plack, A. Oxenham, and R. Fay, *Pitch: neural coding and perception*, ser. Springer handbook of auditory research. Springer, 2005.

[12] D. Deutsch, T. Henthorn, and R. Lapidis, "Illusory transformation from speech to song," *The Journal of the Acoustical Society of America*, vol. 129, no. 4, p. 2245, 2011. [Online]. Available: http://asadl.org/jasa/resource/1/jasman/v129/i4/p2245_s1

[13] N. Collins, "Karlheinz stockhausen: Cosmic pulses," *Computer Music Journal*, vol. 32, no. 1, pp. 88–91, 2008. [Online]. Available: http://dx.doi.org/10.1162/comj.2008.32.1.88

[14] K. Stockhausen, "Cosmic pulses," CD Liner Notes, Kürten: Stockhausen-Verlag, 2007.

[15] W. E. Hill, "My wife and my mother-in-law. they are both in this picture - find them," in *Puck*. Washington, D.C. 20540: Library of Congress Prints and Photographs Division, 1915, vol. 78, no. 2018, p. 11. [Online]. Available: http://www.loc.gov/pictures/resource/ds.00175/

[16] J. Gossmann, "From metaphor to medium: Sonification as extension of our body," E. Brazil, Ed., International Community for Auditory Display. Washington, D.C., USA: International Community for Auditory Display, June 9-15 2010. [Online]. Available: http://icad.org/Proceedings/2010/Gossmann2010.pdf

[17] D. Huron, *Sweet Anticipation: Music and the Psychology of Expectation*. The MIT Press, Mar. 2008.

[18] C. P., "Temporal codes, timing nets, and music perception," *Journal of New Music Research*, vol. 30, pp. 107–135, June 2001. [Online]. Available: http://www.ingentaconnect.com/content/routledg/jnmr/2001/00000030/00000002/art00002

[19] S. Pinker, *How the Mind Works*. W. W. Norton & Company, Jan. 1999.

[20] P. Vickers and B. Hogg, "Sonification abstraite/sonification concrete: An 'aesthetic persepctive space' for classifying auditory displays in the ars musica domain," C. F. A. D. N. E. Tony Stockman, Louise Valgerur Nickerson and D. Brock, Eds., Department of Computer Science, Queen Mary, University of London, UK. London, UK:

Department of Computer Science, Queen Mary, University of London, UK, 2006, pp. 210–216. [Online]. Available: Proceedings/2006/VickersHogg2006.pdf

[21] U. Eco, *The Open Work*, 2nd ed. Harvard University Press, Apr. 1989.

[22] M. McLuhan and Q. Fiore, *The Medium is the Massage*. Gingko Press, Oct. 2005.

[23] S. Barrass and V. Best, "Stream-based sonification diagrams," Paris, France, 2008, inproceedings. [Online]. Available: Proceedings/2008/BarrassBest2008.pdf