From maximal to minimal voice – concepts for evaluating vocal sounds in electroacoustic music

Andreas Bergsland, NTNU (Norwegian University of Science and Technology), Music Dept.
Presented at EMS – 05, 22.oct. 2005

Introduction

In this paper I would like to give the outlines of a theoretical framework that can be used in analysing, evaluating and classifying electronically transformed voices within acousmatic electroacoustic works. The presentation is based on an ongoing doctoral project in its second year, and is by no means fully developed theory, but will nevertheless give an idea of the direction my research has taken so far. The aim of this presentation will be to show how different conditions can contribute in more or less integrated, coherent and stable images of an active sound-producing being responsible for the perceived sound, a being which we, adopting one of Michel Chion's terms, can call an *acousmêtre*.[1]

Thus, my topic can be seen as a special case of what commentators as Denis Smalley and would call a source/cause relationship, dealing not only with sound in itself but the imagined origin of sounds.[2] However, in the case of the voice, this relationship is highly complex and multidimensional due to the voice's special status in human relations and communication. Also, our relationship with sounds perceived as vocal is of a particular kind, due to our general familiarity with the voice, both from a perception and a production point of view.[3] Michel Chion's comments on the *vocosentric* character of cinema, would probably also apply to "vocal" electroacoustic music: "*There are voices, and then everything else*" (Chion: 5), meaning that when a human voice is present it sets up a hierarchy of perception, centring the voice. As, Chion states, paraphrasing Christiane Sacco, the "*presence of the human voice structures the sonic space containing it* […] If a human voice is a part of [a sonic space], the ear is inevitably carried toward it, picking it out, and structuring the perception of the whole around it. The ear attempts to analyze the sound in order to extract meaning from it – as one peels and sqeezes a fruit – and always tries to *localize* and if possible *identify* the voice" (*loc.cit.*). However, in electroacoustic music, this structuring, localizing, identifying and extracting meaning is wholly dependent on the "condition" of the voices heard, their felt presence and meaningfulness, or the degree that they are perceived as voices all together. It is

---

[1] Chion, M. (1999).
[2] See e.g. Smalley 1997.
[3] Belin et al. (2000) have even found evidence of voice selective areas in the human brain.

exactly this gradation that I will try to come to terms with in this presentation through the concepts of maximal and minimal voice.

Theoretical background

The concept of maximal and minimal voice is borrowed from the literary theorists Donald Wesling and Tadeusz Slawek.[4] Concerned with subjectivity, intentionality, expressivity and the dismantling of these conceptions within postmodern literary practices, the authors introduces the notions of maximal and minimal voice as tools in analysis of literary texts. The maximal and minimal voice are presented as ends, or extremes of literary voice, which the authors theorize mainly by analyzing texts that are seen as belonging to each of the extreme positions. They see the reader as an integrating and constructive force that faced with the even the most disjunctive and indeterminate postmodern texts, "hunt for continuity, individuality, centeredness, bounding outline, [and] social tones of voice" (Wesling & Slawek 1995: 11). My proposition in this article is that Wesling and Slaweks concepts can be useful tools also in the description and analysis of electroacoustic music centring on vocal sounds, albeit in need of considerable qualification. Literature and music are very different as forms of art, the former being primarily read whereas the latter is primarily listened to. Still, both a reader of a literary text and a listener to "vocal" electroacoustic music will have a tendency to try to create and structure meaning around one or several enunciating subjective forces from such stimuli, as well as identify, localize and set the boundaries around them. The ways that these processes work is of course very different in the two artforms, but they both relate whatever that is perceived to some form of expression, that is, a being involved in some form of production that conveys some form of meaning.

Furthermore, I find their approach through poles or extremes fruitful. This is by all means not a new way of structuring a representation of a quality in electroacoustic music, but it is to my knowledge a new way of dealing with questions of subjectivity.[5] One problematic issue in applying maximal-minimal as poles on a continuous scale, however, is that almost by definition, the maximal voice is seldom or never found in its fullest sense in electroacoustic music. Still, it is my claim that the maximal pole constitutes a frame of reference that even sounds perceived as only faintly vocal will relate to – a sort of centre of gravity that electronically manipulated and transformed voices will orbit around.

---

[4] "Wesling D. and Slawek T. (1995) Literary Voice - The Calling of Jonah. State University of New York Press, Albany.

[5] For an example of such a structuring see e.g. the reality-abstraction continuum proposed by Young (1996).

Maximal voice

I would like to start out by trying to define what I mean by maximal voice in relation to what Smalley and Emmerson defines as the "acousmatic" branch of electroacoustic music.[6] A short and tentative definition of the *maximal voice* would be as follows: The sense, created in the mind of the listener through hearing the a vocal sound, of *enduring* presence and centrality of a temporally stable, clearly identifiable, localizable and definable human *acousmêtre*, conveying highly over-determined meanings. As in Wesling and Slawek's account of the literary maximal voice, one is still here concerned about questions of continuity, individuality, centeredness, bounding outline and of rendering from it a relatively unambiguous meaning, not only hinging on the qualities of the work, but also on the mental structuring and assigning of meaning of its receiver. The maximal voice does therefore not only refer to the sounds that are listened to, but is as much a way of listening to the sounds, where this way of listening is a product of a conventionalized practices both of production and reception that has evolved since the last half of the nineteenth century together with the development of sound reproduction technologies.[7]

As earlier said, what I have called maximal voice is rarely found in electroacoustic music in its fullest sense, but rather in other sonic expressions as radio and audio-book genres like the interview, causeries, speeches and lectures.[8] In all these genres, the voice appears either undisturbed as the only sound source, or as the foreground phenomenon, with very few other sound sources interfering that potentially could threaten the intelligibility of what is being said. Neither environmental sounds nor sounds implicit in the recording process are allowed to come to the foreground, preferably by minimizing noise through using recording devices and practices such as close miking or soundproof studios to shut out interfering sounds. Alternatively, noise can be kept stable and predictable, to allow the listener to "listen through" the noise and eventually ignore it. Characteristic of all the mentioned genres is that the voice is presented as conventional and meaningful speech in them, unquestionably giving centrality to the verbal meaning that it conveys and at the same time eschewing everything that could disrupt or undermine the clarity of the verbal message. The "radio voice of authority" as defined by Katharine Norman, quoting Dyson, fits clearly within my definition

---

[6] "Acousmatic music is intended for loudspeaker listening and exists only in recorded form (tape, compact disc, computer storage)", Emmerson & Smalley (2006)

[7] See Sterne (2003) for a thorough account of the genealogy of this practice.

[8] It can also be found other expressions as well, for instance there is a section in John Coulter's multi-media work, *Shifting Ground*, that displays both maximal and minimal voices.

of the maximal voice and demonstrates the centrality of the verbal aptly: "[the radio voice of authority] does not mumble or stutter, it pronounces full and meaningful sentences, it says something" (Dyson, quoted in Norman 2004:104). Further on Norman adds that "the voice of authority […] assumes that you will listen because it has information that you need to know" (Norman 2004: 104). Thus, Norman's and Dyson's conception of the voice of authority underlines the privilege of the verbal at the same time as it shows how this could be related to issues of power.

With a centrality assigned to the verbal content of speech in the maximal voice, what is it that becomes periphery? Firstly, there are of course all the aspects of the vocal sound that are related to the act of producing the utterance – all the physical and physiological causes that we perceive to be responsible for the sound wee interpret as voice as well as all the abstracted sonic qualities that can be perceived from them. Naturally, these aspects are still very much present and part and parcel of the message, and indeed a part of what we think of as voice, but nevertheless such aspects are relegated to playing a supporting rather than central role; they are background rather than foreground, creating emphasis and presence, uniting segments into larger units, connecting the utterance with the utterer and so forth. Secondly, we do not only have to consider voice "in itself" when we listen to a reproduced voice, as the case is with the maximal voice, we also have to think about the context or environment that the maximal voice is speaking from and situated in. Because, whenever we hear a voice emitted from a speaker, it is speaking from *somewhere*, and when we are aware of the reproduced nature of the voice, that *somewhere* is primarily not the location of the loudspeaker, but the *virtual* space that surrounds the speaking source placed "within", so to speak, the framing technological mediation. But still, for the maximal voice we see that these aspects are given a supportive and even more peripheral role than that of voice production. From the production perspective one has been using the techniques already mentioned to close off or downplay potentially interruptive sounds. The studio, as the ideal environment to consign transparency or inaudibility to this dimension, thus becomes a place that can be heard only negatively as the absence of acoustical markers that can point to the environment that surrounds the speaker; like reverberation, echo or other sound sources. Even if it is hard to draw exact borders between this context/environment and the technology that mediates the voice within this environment, exactly because the practices of e.g. microphone placement and soundproof studios that constitute the context are one part of this technology, one can designate a final peripheral zone; the technology that plays a part in the mediation separated from the "there and then" of the event of the recording. Again, from the production

perspective, one has always been aspiring to the conditions of "high-fidelity" where technology should seek to be what Jonathan Sterne calls a "vanishing mediator – a means that would obliterate itself in achieving its end" (Sterne 2003: 283). Thus, for both the context/environment and the technology aspects we can also see that there is a logic of transparency at work – they are there, but are not supposed to be heard. Taken together, the maximal voice establishes a kind of perceptual hierarchy that can be viewed as having a focus or centre with degrees of periphery surrounding it, where the verbal is the focus point, and where the other mentioned aspects, for simplicity's sake labelled *vocal production*, *context/environment*, and *technology* are peripheral to the centre in different degrees, here visualized as subsequent layers removed from a central area (see Fig.1). The lines between the *verbal* and the *vocal production* and between *context/environment* and *technology* are made dotted to indicate less defined boundaries between the areas.
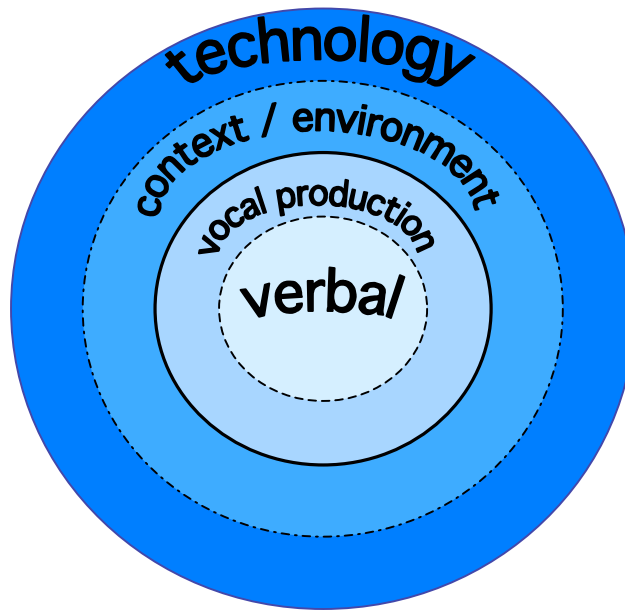

Figure 1- The hierarchy of the maximal voice

We can also see that at the outer levels there is a logic of transparency at work. In supporting the ideas of *expression* and *reproduction* the sounds from the loudspeakers are this way firmly embedded as a faithful re-presentation of an event projecting meaning or knowledge as relating to a particular somebody's production. Through establishing itself as a socially constructed norm involving both production and reception the problems of uncertainty and absence inherent in the temporal and/or spatial displacements of the acousmatic situation are in this way counteracted.

Minimal voice

So, to the opposite pole of the concept pair; the *minimal voice*. If a short and tentative description were to be given at this moment, it would be that *minimal voices are distant, abstracted, and hardly associable with human vocal production, on the border of not being voice any more, and of becoming "pure" sound, where no clearly definable meanings are conveyed*.[9] However, it might be just as pertinent to define it as the negative of the maximal voice lacking one or several of its constituent properties; *enduring* presence and centrality of a temporally stable, clearly identifiable, localizable and definable human *acousmêtre*, conveying highly over-determined meanings. The first thing to be noticed about this definition when related to the use of voice in electroacoustic music is that whereas the maximal is relatively narrow and well-defined as a category, the minimal voice would potentially include broad categories of sound, comprising a wide range of different possibilities, all departing from features that we could identify in the maximal voice.

The second thing to be noticed about the minimal voice is that the primacy of the verbal is subverted, resulting in "no clearly definable meanings". If we still maintain that the aspects of *vocal production*, *context/environment* and *technology* play a part in the perception of minimal voices, it is clear that the mentioned hierarchy of the maximal voice must be re-arranged in some sort of way, privileging one of the other aspects. But, the relative openness of the definition of the minimal does not prescribe *how* the verbal is subverted, it rather suggests that one here is at the border where the mentioned aspects can cease to be relevant for the perception of the sound in becoming "pure", that is, disengaged from notions of origin in favour of more abstract qualities.[10] It seems then, that it is much more unclear in the case of the minimal voice whether the aspects that were relevant to the maximal voice are at all relevant to the minimal, and if they are, it is equally unclear how the aspects are organized, whether there is a particular perceptual focus and what aspect that would constitute such a centre. To visualize this would therefore be futile.

However, it would make sense to try to visualize how the minimal voices of electroacoustic music relate to the maximal. As a preliminary simplistic model capturing the mentioned narrow-broad relationship, one can visualize the relationship between the maximal and minimal voice as a centre – periphery model. (See fig.2) This model also facilitates the possibility of classifying something as lying closer to the centre vice versa the periphery, it

---

[9] Thus, the continuum from maximal to minimal voice can thus be seen as a developed source-specific gradation Smalley's degrees of surrogacy (Smalley 1997)or of the realism – abstraction continuum of John Young (Young 1996).

[10] Such a perception would be close to Schaeffer's ideas about *reduced listening* (Schaeffer 2002).

visualizes that there are certain critical limits that delineates a bordering onto what is not voice, and the possibilities of having different trajectories between maximal and minimal.
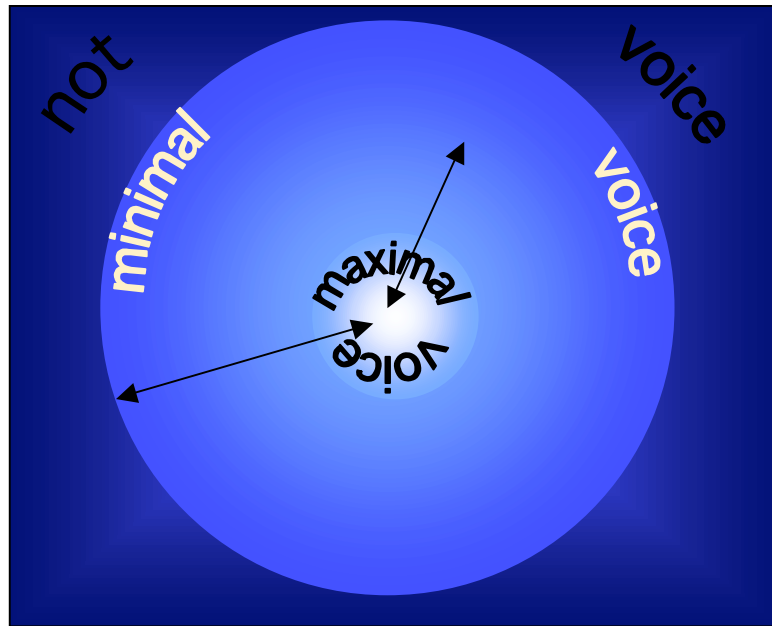

Figure 2. Centre-periphery model of maximal and minimal voice

However, this model does not signal any qualitative differences between occurrences other than their distance to the centre, so it does not tell us anything of the features that contributes in placing a vocal sound at one or the other location, or how the levels in the mentioned hierarchy plays a role in this. It would be tempting then, to try to refine the model by including more dimensions or aspects to it, so that it could be a better tool in describing the interaction between listener and vocal sounds in electroacoustic music. In the following part of this article I have tried to describe a set of such dimensions, knowing that such a formalization both can gain and inhibit the description of a segment of music. Gain, because one then could have a tool that could be used to compare different segments within one piece or between different pieces. Inhibit, because such a formalization always will implicate a simplification of matters and a shortcoming to the singularities of each piece and each listening. Still, most form of representations struggle with similar inhibitions, and the solution to deal with these shortcomings can be to strive towards a multiplicity of representations. At least, this is what Godøy proposes: "It is in abundance, even superabundance of different representations, both of different overviews images and different sonorizations of the musical object, that awareness and knowledge are created" (Godøy 1997: 302). Also, I have to stress that the model I am proposing in this article is by no means fully developed, but must be seen

as a first step in the development of one of several possible tools for analysis and description of voice in electroacoustic music.

Immediacy

The first aspect I have chosen to describe her can be labelled *immediacy*. The label *presence* could possibly be quite as pertinent, since I am starting from the intuition that hearing an acousmatic voice will implicate a sense the listener of the acousmêtre "being there" within or behind the sound. However, this sense of immediacy or presence is understood in a particular sense, namely in a negative relation to the medium of experience, which can be seen as equivalent to the *technology* aspect in the earlier mentioned hierarchy. Actually, perceiving the technology involved or the mediated nature of the experience tend to be an impediment to experiencing the acousmêtre as "being there", as mediation through technology implies a distance. The question of immediacy is at the centre of any mediated experience, are we to believe the new community of empirical researchers focusing on mediated experiences, human-computer interaction and virtual reality, however giving it the label *presence* and defining it as a "failing to", "overlooking" or "behaving as if" the technology did not play a part in the experience (Lee 2004: 203). According to this explication, then, presence or immediacy is an effect of a lowered attention to the role of the mediating technology in the experience, and accordingly that awareness of technology would cause the experience of presence to fade. The media theoreticians Bolter and Gruisin have also set up an opposite term for this kind of immediacy, namely *hypermediacy.* In their view immediacy implies a belief in "some necessary contact point between the medium and what it represents" (Bolter & Gruisin 1999: 30) whereas hypermediacy "multiplies the signs of mediation" (*ibid.*: 34). If we transfer this to the maximal and minimal poles of voice, immediacy is naturally situated at the maximal pole, and an increasing attention to technology would then entail a movement towards the minimal, where minimal immediacy would converge with the situation of hypermediacy.

Among the huge variety of sounds that can be heard in electroacoustic music, there is little doubt that electronic technology has an essential role, at least on the production side. In Grove Music Online electro-acoustic music is defined like this: "Music in which electronic technology, now primarily computer-based, is used to access, generate, explore and configure sound materials, and in which loudspeakers are the prime medium of transmission" (Emmerson & Smalley 2006). This does not mean that how the sounds are produced necessarily have to be reflected in how they are perceived. But, even if there is no necessary

and causal link between means of production and perception, this does not mean that technology must be totally irrelevant to how the music is perceived, like Smalley seem to advocate in his spectromorphological approach (Smalley 1997). Such a view is supported by Emmerson, who in a review and critique of Smalley's notions of indicative fields opens up for the possibility that "technology itself, may become a reference field, drawn attention to as crucial signifier; the acousmatic veil torn down and the transparent means of production and dissemination become the subject of the discourse" (Emmerson 2000, 205).

If some sounds are profoundly marked by a certain production technology other sounds seem less marked, and even approach what one often would describe as "unprocessed" or "mere reproduction", at least for segments of a piece. And, as I have already stated, to perceive it in this way would be one of the conditions for us to hear the voice as *maximal*, in that it would be experienced as if "overlooking" technology. However, as the early discourse on sound fidelity shows, what is experienced as "marked" by technology is highly relative to historical and socio-cultural conditions.[11] The listeners' ability to "listen through" whatever hiss and crackle is present to the reproduced sound, demonstrates the importance of the listener's frame of reference when it comes to what is commonly called the "quality" of the reproduction. Exactly this relativity is in my view highly relevant also within a single piece when it comes to the experience of a vocal sound as more or less processed. Any changes or differences in the way technology contributes as a component in the perception of the sound, let's say between "unmarked", "less marked", "clearly marked" and "heavily marked" sounds, can place a focus on or give an awareness of the mediating or production technology. Such a focus will at the same time make the experience of an acousmêtre into a mediated rather than an immediate experience, making the distancing of time and space between the "here and now" of the listener and the "there and then" of the acousmêtre apparent. Thus, the awareness of technology will affect the sense of immediacy in a negative way; interrupt it, introduce distance or even destroy any sense of presence. If there are little differences when it comes to this aspect, i.e. when the sounds perceived differ little in the degree that technology is experienced to play a part, it is less likely that it will play a prominent role in perception. There are two ways in which this may happen. Either, as when something is perceived as reproduction ("overlooking"), or as when

---

[11] In 1908 the quality of reproduction of voices from the Victor Talking Machine was so much better than earlier predecessors that its producers cast in doubt that the public would be able to tell the difference between the sound of a real singer and a recording played on a victor. See Sterne (2003) chapter 5 for a broad account of the early discourse of sound fidelity.

the listener's experience of involved technology can become totally redundant or implicit – "media numb", as John Richards has called it.[12]

Even if there are numerous ways that the attention to technology and mediation can be called in electroacoustic music, I would like to focus on the effects of experiencing departures from the linear unidirectional and uninterrupted temporal flow that constitutes the norm for maximal voice, and which is an important factor in establishing a sense of presence, where the acousmêtre "is there" for the listener. With that as a point of reference, one can experience that whenever disruptions, repetitions, skips or superposition of temporal layers are heard, the temporal flow in which the acousmêtre exists will momentarily loose its track, thus making it more difficult to maintain a stable mental image or a feeling of presence regarding the "being" behind the sound. At the same time, this break of realism will generally be accompanied by a consciousness of the technology that mediates and to smaller or larger degree transforms the components of the experience. The technology that formerly was transparent and implicit would then become explicit, getting what Katharine Norman would call the "unclean voice of the edit", something which in this context would clearly mark one kind of departure from the maximal voice (Norman 2004: 114-120). That a recording is actually edited will in this context be arbitrary, since an edit can be both audible and inaudible. Thus, there has to be perceived discontinuities in some form, for these kinds of effects to occur.

This is appropriately exemplified in a passage from Trevor Wishart's short text-sound piece *Blue Tulips*. Here, we hear an elderly woman retelling her dream about some blue tulips, and even if there are almost disturbingly long silent sections between the spoken phrases, and if the quality of the voice seems slightly marked by some filter or compression algorithm, the absence of discontinuities still manages to uphold a sense that the woman is sitting there telling us about it. This situation takes a turn, however, at the moment when the syllable "blu" is repeated and superimposed on the speech, and with the subsequent repetition of the whole phrase, this tendency is confirmed. The effect is not only that one suddenly gets an awareness of the mediating technology, but clearly also one of distancing and a "framing" of the syllable and phrases as manipulable objects. That this didn't happen until the felt disruption of the time flow despite the light colouring of the voice, can maybe tell us that marks of technology that are relatively constant and predictable, easily can be overlooked or ignored, to create the sense of immediate presence.

---

[12] John Richards, "Getting beyond the medium". Paper presented at EMS05, Montréal, October 2005.

We can also locate the same effect in a piece by the Norwegian composer Maja Ratkje, namely in *Insomnia* from *Voices*. Here, in the beginning and a few minutes out in the piece we hear a voice that judging by the narrow frequency range and high noise level, sounds as if it is recorded and played back on a Dictaphone or small radio. The second time we hear this voice it is abruptly stopped, and we hear the click of the button as if the playback of this voice is turned off. Thus, in recognizing that the voice is played back on some device we experience a form of secondary source, other than the voice itself, a source which in itself is a mediating technology, in some way "contained" and operated within the technology that mediates the music. In this way, it is not only the disruption of the temporal flow that makes us aware of a mediating and transforming technology, it is the sonic imprint of a recognisable recording device, highlighted through the click and the contrasting "silence" that follows, or to be more precise, the very low carpet of noise that implicitly marks "silence" within the frames of our current standards of technology. At the same time this contrast creates a "framing" effect, marking the segment with defined borders, as something, as an object, which is separated through the audible discontinuities, but also temporally, between a past event, the voice on the Dictaphone, and a quasi-present event; the click that marks the stopping of the playback of the recording. Thus, we get a kind of double distance here – it is distanced because of the explicit mediation and because of it being contextualized as "past". Even if the distancing in this example marks a departure from the "presence effect" of the maximal voice, it still uses it as reference point from which it has not moved too far.

To sum up, we have seen that through disruptions in the temporal flow, the technology layer is no longer a transparent mediator serving a verbal expression, but can be experienced as contributing to aspects of the musical work. Where disruptions reach a certain rate, e.g. through repetition, the attention to technology tends to fade and becomes redundant. Thus, the issue of immediacy can be categorized as having three main "modes" – from "overlooking" technology / experiencing immediacy, via "awareness" of technology / experiencing distancing to "numbness" of technology and of the existence of an acousmêtre. This can be visualized in this manner:

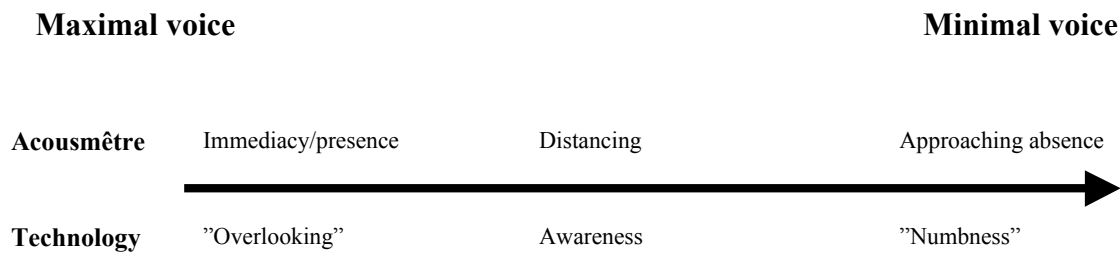|	**Maximal voice**		**Minimal voice**
Acousmêtre	Immediacy/presence	Distancing	Approaching absence
Technology	”Overlooking”	Awareness	”Numbness”

Figure 3-Maximal to minimal voice: Immediacy axis

In this way we have a structuring of one aspect relevant to the polarization between minimal and maximal voice that matches my intention of having a dimensional representation. Even if such a representation is by no means a fine grained instrument, having only three course categories, it still allows for some differentiation, which together with other dimensions will give some idea of how voice in electroacoustic music is perceived.

Salience

So, to the dimension that I have called *salience*. This dimension refers to the degree in which the different qualities of the voice "stand out" perceptually for the listener. Thus, such a dimension deals with the clarity, detectability and unequivocality in which the features of the voice are perceived. When these conditions are positively fulfilled one would be at the maximal end the dimension, and when they are negatively fulfilled, i.e. when the voice is perceptually close to absent, one would be located at the minimal end. In this relatively short discussion I would like to say something about the factors that can contribute to a move towards the minimal end of the salience dimension, and also something about the auditory processes involved.

To start with the former issue, one can make a distinction between the contribution of the sonic properties of the sounds that are perceived as belonging to the voice, and the contribution of any sonic context accompanying the voice in interfering or interacting with it. To take some very basic examples of the first, one can refer to *loudness*: To detect features from a voice, let's say a melody or a chain of words, which is presented with a loudness close to the threshold of hearing will naturally be a lot more difficult than with one to that is presented with a loudness equivalent to a face-to-face conversation. The same will be the case

if one hears only parts of the frequency contents of a voice, e.g. as a result of band-pass filtering. As for the second, sounds that accompany the voice can naturally also interfere with the listeners' ability to detect the features of the voice, and this effect is usually referred to as *masking* in literature dealing with human audition (see e.g. McAdams & Bigand 1993, Handel 1989 or Bregman 1990).[13] If a vocal sound was suddenly accompanied by a much louder sound, especially one that covers the same frequency range as the vocal sound, it would be difficult to perceive its features. Thus, for both kinds of contributions, one can speak of a *loss* of available features to perception, when compared to a speaking voice produced and reproduced with maximal clarity.[14]

This dimension can also be seen in relation to many of the grouping factors that are described by cognitive music research, most notably the auditory scene analysis theories formulated by Bregman (1990, 1993). In short, when the listener hears something, s/he groups sound according to innate criteria such as consistency of location, harmonicity of partials, and synchronicity of change in the spectrum, and/or learned knowledge about the behaviour of different sound sources like the human voice. These clues can be seen as "giving votes" to decisions whether something can be grouped into one or the other sound stream. However, the point is that when all these criteria point in the same direction, the decision is unequivocal in pointing to a source. A single voice, not interfered by any other sound, as in the maximal voice, would consequently also be unequivocally perceived as one vocal source. If other sounds are present that affects these decisions so that they become more ambiguous, however, it will also become more difficult to group the sounds heard, and the voice would stand out less for the listener. In this way, it would be more difficult to discern the bounding outlines of the voice, because it only partially stands out from the other sounds. Then again, this can be caused by both sounds that have similar qualities and those that are different. On this background, sounds can both stand in the way of the voice or *blend* or *merge* with the sound of the voice, thus making it difficult to discern the voice as a distinct source.[15] Here, an acousmatic section [01:24-03:40] from John Coulter's multimedia piece *Shifting Ground* (2005), provides a good example of these effects. In this section, we start hearing clearly separated segments of single voice in a sequence, making both words and properties of the

---

[13] Bregman defines masking as an effect that occurs "when a loud sound covers up or drowns out a softer one" (Bregman 1990: 27).

[14] However, it is not always easy to draw the line between effects related to properties of the voice in itself and related to contextual interference. In the case of reverberation and echo, the sound that interferes with the voice is made up of multiple delayed and attenuated versions of itself. Thus, in one sense it would be the contribution of the room surrounding the voice that interferes, in another sense it would be doublings of the voice.

[15] The term that is preferred in the study of auditory cognition is *fusion*. See e.g. Bregman (1990: 311-20).

speaker clearly perceptible to the listener. After a while the voices overlap more and more, masking each other, until they finally merged into the sound of a huge crowd. This also takes us to where the other end of this dimension points, namely a perceptual absence, either, as here, through masking and merging, or through fading out in silence. Other sorts of ambiguities can arise from mismatches between the mentioned knowledge about the voice as a sound source and the perceived sound. For instance, in the first part of Andres Lewin-Richter's text-sound piece *Caminando* (2002) [05:15-06:00], we hear a voice made by some kind of vocoding / cross-synthesis technique, where one can easily recognize the words that are articulated, but where the phonation part is replaced with a synthetically sounding major chord. In this case some auditory grouping mechanisms would "vote" for the segregation of the tones in the chord into separate streams not recognized as voice, whereas others would "vote" for integrating them into one single stream corresponding to one single vocal source.

To sum of this section of the article, I will present a simple schematic figure that includes some of the mentioned points for this dimension:
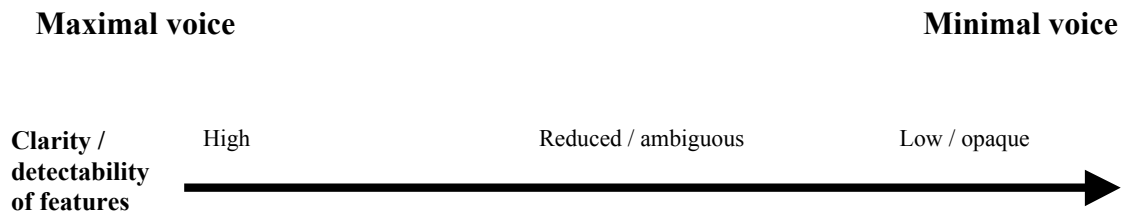
Maximal voice   **Minimal voice**

Clarity /   High   Reduced / ambiguous   Low / opaque
detectability
of features

Figure 4-Maximal to minimal voice: Salience axis

Source definition

The third dimension I will include in this theoretical framework I have called *source definition*, i.e. the degree of definition of the properties of the acousmêtre. By properties here, I mean the different aspects can be seen as "identity tags" or "traits" that together make up a particular identity configuration that is a part of the constitution of the acousmêtre. Such tags can be gender, age, personality and socio-cultural background, ethnicity, social status – and even a name and the memory of a face, if this person is known, and they can be inferred both

by verbal and non-verbal, or extra-linguistic clues. [16] A voice that can be defined and labelled easily and with high specificity with regard to physiological, psychological and socio-cultural aspects, would also be more easily be integrated into a coherent image of an acousmatic being, thus, for the maximal voice, these features are easily identifiable and have a high degree of definition. One example her would be Guiseppe Rappisarda's piece *The day before* where the use of Tony Blair's voice gives immediate recognition to anyone familiar with his voice from radio or television. In this case, I will guess that for most of us, we even can imagine his face and gestures when hearing the sound of his voice in this way. If we go to a relatively well known piece like Åke Parmerud's *Les Objets Obscures* (1991), however, the voice appearing would most likely be unfamiliar, but still quite close to the maximal because of the relatively unambiguous assigning of the mentioned identity "traits".[17] The voice we hear is intuitively recognized as an adult female, and linguistically and culturally she could by non-natives be identified as French - by French people possibly socially and culturally identified with a much higher degree of specificity, such as maybe upper middle class, Parisian, etc.

When we hear voices that are more ambiguous or indefinable according to these identity traits, however, it would indicate a motion further towards the minimal. When we experience a voice that evidently has the marks of increased playback speed (or decreased recording speed), like the introduction of Maja Ratkje's *Voices* (2002), the mentioned effects of attention to technology will tend to be evoked, but all the same this does not prevent us from trying to categorize through assigning identity tags to the voice. In fact, Clifford Nass and Scott Brave reports numerous experiments with human-computer interaction through voice that confirms a general tendency for people categorize voices according to gender, age, personality and place of origin, regardless of knowledge that the voice is synthesized. (Nass & Brave 2005). The relatively high-pitched voice we hear in Ratkje's piece, however, is not clearly lying within the characteristics of a certain category. The pitch is somewhat higher than an average woman's, perhaps suggesting that of an adolescent, but what makes the gender / age decision ambiguous here, is that the speech sounds do not have their characteristic peaks, or formants, where they would have been expected for neither a woman

---

[16] Laver (1991) defines dialect and characteristic sociolect as belonging to paralinguistic layer. However, since these features also give a lot of information about the speaker, I think it is logical that they are included in the extra-linguistic.

[17] The voice we hear in *Les Objets Obscures* by Parmerud belongs to Dominique Andriaux.

nor a child of any age.[18] So, due to this deviation it is difficult to unequivocally categorize the speaker according to the mentioned traits.[19] In this example, the verbal level amplifies this uncertainty of identification, not according to gender, but to place of origin, being a strange mix of Norwegian and English, both as phonemes and words.[20]

If we pursue the logic of this dimension towards the minimal end, we would end up with a voice that we would have few or no identity tags that we could attach to it, something we could label undefined, thus giving us this kind of figure:
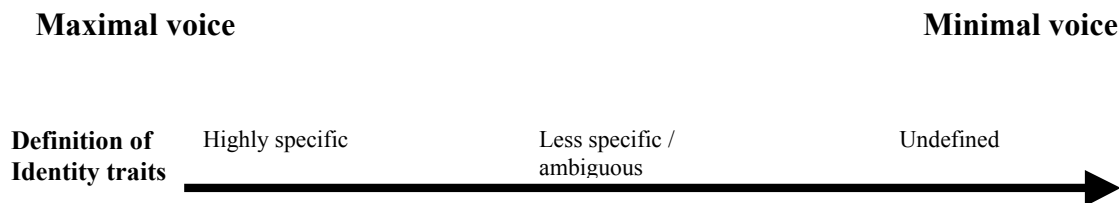

Maximal voice   **Minimal voice**

Definition of   Highly specific   Less specific /   Undefined
Identity traits   ambiguous

Figure 5-Minimal and maximal voice: Source definition axis

Information

Compared to the dimensions described so far, the dimension that I have called *information* is probably the most complex and challenging, and I am again underlining the work-in-progress character of my model. I have tried here to subsume the degree in which the voice conveys signification that is meaningful and coherent for the listener in any other way. However, what makes this extremely complicated is the many ways in which this information can be conveyed, the types of codes involved etc. Due to limitations of this article, I will only touch onto some of these problems.

Starting again from the maximal, as I have said, the priority is given to the verbal level, allowing what Wesling & Slawek call "over-determined meanings", to be inferred, that is, with a minimum of ambiguities and incoherence. So, if the verbal in its function as tool for

---

[18] This can be due to the fact that the difference between the two first, and perceptually most important formants, of man, woman or child, is not the same for different vowels (see Kent & Read 200?, s.204)
[19] As Bosma reports, gender patterns still can be inscribed in electronic voices, but then referring to the mode of production where this "stereotype relates woman to body, performance, tradition, non-verbal sound and singing, and man to electronic music technology, innovation, language and authority" (Bosma, 2003)
[20] According to Nass and Brave, effects of uncertainty of identity definition can lead to the classification as "strange, dislikable, dishonest and unintelligent", it can lead to difficulties in perception and processing, or disregard of the inconsistent information (Nass & Brave 2005: 16).

conveying coherent meaning is at the centre in the maximal voice, it would be evident that within music and sonic artforms, the verbal is rarely the sole centre of attention, thus departing from maximal voice in the fullest sense. In vocal music from the Western tradition, one has a repository of vocal styles that all display different relationship between the verbal and that of vocal production, from spoken recitation and *Sprechgesang*, to highly melismatic singing and pure vocalese.[21] In electroacoustic music there is an even broader palette of ways in which the significative functions of the verbal can be weakened or undermined. In such cases, the intelligibility of the verbal layer can be affected, and linguistic structures can be put under different degrees of pressure through a departure from conventional speech patterns, often combined with a focus on sonic properties. This pressure tends to affect referentiality, reducing the symbolic mode of signification, thereby decreasing the possibilities of meaningfulness and coherence using linguistic means.[22] Thus, the hierarchical arrangement of the levels earlier presented can obviously be affected, putting either the vocal production and the contextual levels in the centre of attention, allowing different types of non-verbal information, like para-linguistic information regarding emotions, moods and attitudes of the speaker, or gestures of the vocal production, to have more emphasis. This type of information, though, is less defined and precise, and has less potential for high density of information, compared the propositional and symbolical function of the verbal.

If the non-verbal aspects tend to be less precise and more ambiguous than the verbal, there are situations where even those referential aspects of sound tend to be exhausted, leaving the material qualities of the sound as the only relevant focus in perception. It therefore makes sense to let such a situation be represented by moving towards the minimal extreme of this axis, representing a way of perceiving a voice as almost ceasing to be one, approaching "pure" sound. A sound conveying little or redundant information, would to a much higher degree invite to attending to its sounding qualities than would a meaningful spoken sentence. One can experience that it is much easier to listen for sound qualities in a language one doesn't understand one's mother tongue, making it important to emphasize that this axis must be seen in relation to the listener's competence and training. Thus, we can see that it can be meaningful to place high degree of information at the maximal end of this dimension, and

---

[21] See Peter Stacey (1989) for a thorough account of different styles of vocalizations in Western art music in the 20th century.

[22] I have shown elsewhere how verbal meaning can be put under pressure from musical forces in the contemporary symphonic song *Vævet af Stængler* (1991) by Norwegian composer Olav Anton Thommesen. (Bergsland, 1999, 2000). Here, I took Aksnes' (1995) reworking of Julia Kristeva's theories of the symbolic and the semiotic as a starting point for my discussion. The symbolic mode is related to the act of predication, which is seen to constitute both the subject and the object. The semiotic however, is what is heterogeneous with regard to meaning, what undermines and "spills over" the verbal, contributing to ambiguity.

abstraction in the other, but that extra care must be taken to account for differences between listeners.

The movement along this dimension can be illustrated by a piece like Steve Reich's *Come Out* (1966). In this piece, we can detect a movement in the way we perceive what we hear from the first appearance of the segment and through each subsequent repetition. We start from something close to a maximal voice, where we pay attention to both verbal content, and what the vocal production is telling us about different features of the speaker. With the first cut, our attention is directed to the mediating technology, and we have the distancing and framing effects also affecting presence, and proceed eventually to a phase when all these effects become gradually more and more exhausted and redundant in favour of a focus to the slow changes of the sonic qualities of the slowly phase shifting superimposed versions of the loop. When we have reached this last phase, we still have a sequence of sound recognizable as voice, but by redundancy the information that it carries in the verbal level and sound producing levels, is approaching zero – it is no longer heard as relevant. Even if it still clearly human vocal production, the feeling of presence is lost through framing and distance, and through redundant repetition, it is only the subtle changes in sound quality that occupies our attention. Thus, in a sense, Reich has led us to quite close to a Schaefferian reduced mode of listening, which in its radicalness wouldn't hear this as voice at all. Again, to round off this section, I simplify matters by using an axis and three different levels:
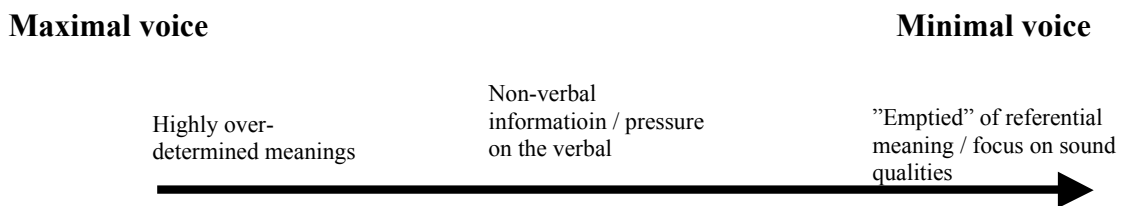
Maximal voice   **Minimal voice**

Highly over-determined meanings

Non-verbal informatioin / pressure on the verbal

"Emptied" of referential meaning / focus on sound qualities

Figure 6-Maximal to minimal voice: Information axis

Human-ness

One effect, that the cartoon industry was early in discovering, was that technologically transformed voices were well suited to match their cartoon characters, which were mostly

animals, but with many recognizable traits of human behaviour. The characters could then have voices which, through the deviation from the most important identity categories, were more open than a normal voice to being transferred onto an imaginary creature, while it still retained the ability to speak and act in a recognizably human manner.[23] Even if the voices in acousmatic music lacks the complimentary visual entities that enforce the integration of such imaginary characters, voices that are apparently marked by technology can encourage linking voices to acousmêtres with only partly recognizable human traits, that is, when qualities perceived as belonging to a voice differs more or less from our expectation of what a human vocal apparatus could possibly produce. The last dimension in my model, I have called *human-ness*, in search of a better word, to designate the degree to which the qualities of a voice are recognizably human or human-like.

The possibilities of signal processing and vocal synthesis have opened for associating voice with fields of meaning outside the norms of vocal expression. This also has opened for acousmêtres with qualities associated with animals, objects, natural phenomena or machines. As for the latter, the term *cyborg* has been used to designate hybrid identities involving human and machine, and not surprisingly it has been used to designate the combination of voice and technology.[24] Thus, as source or cause responsible for the sound of the voice it can deviate from the human in a lot of directions, which it would take to much time to exemplify in this context. This can be seen from two angles. On one side, one can focus on the non-human qualities that contribute to the voice, on the other one can focus on which human qualities that are not present. Even if these points of view are interdependent, it still makes sense to see them separately, since the non-human qualities both can be added to the human and replace the human. For instance, in Lars-Gunnar Bodin's *On speaking terms II: Poem no. 4* the cyborgian voice we hear still contains a discernable human voice, with all of its features intact. In Charles Dodge's *Speech Songs*, it seems more like machine articulation and phonation have replaced the human, but since phonemes and words still can be heard, there still are human elements present. In other words, the verbal is also a marker of human-ness, in this context. Eventually, by following these two axes to their extremes we can be lead beyond voice all together. Were a source is recognized not in any respect resemble human vocalization, and belonging entirely to another category of sources we would have passed the limits of even the minimal voice into non-voice.

---

[23] See Grant (1993) and Stamelman (1980) for accounts of the speed manipulated voices in selected cartoons.
[24] See Bosma (2003).

The trajectories along these axes passing into or fading out from voice seem to have been an attractive field for composers. It is enough to mention Jonathan Harvey's transitions between boy and bell in *Mortous Plango* (1980) and Trevor Wishart's metamorphoses between diverse animals like bees, horses and birds and human voice in pieces like *Vox V* and *Red Bird*.[25] Also in an "abstract" piece like John Chowning's *Phoné* we can find movement along these axes, going from purely synthetic tones to ones that border on voice, and to tones that approach realistically sung tones.

We are left, then, with a simplified representation of the dimension of *human-ness* that would look something like this:

Maximal voice   **Minimal voice**

All human   Hybrid / partly human   Non-human
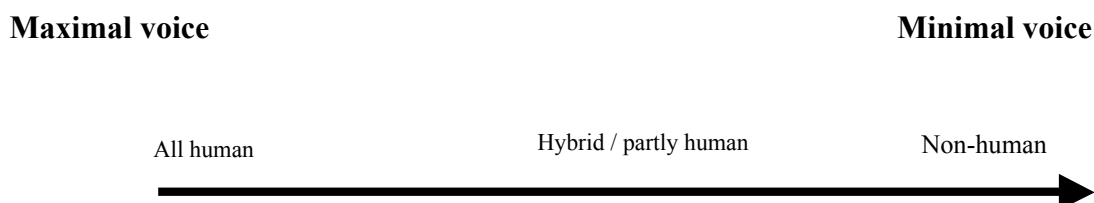
Summary and conclusion

So then, after having explicated the suggested dimensions in this model it still remains to see if it can be applied as a useful tool in analysis of EA works with vocal components. One first step can be to see the five dimensions that I have discussed so far in this article combined with the preliminary centre-periphery model (see Fig.7). Such a representation can be used to visualize how segments of works can be classified according to all of the five dimensions. This gives us a five-point plotting, and by adding lines from point to point, a graphical shape occurs that will be particular to one certain configuration of values on the five axes. (See Fig.8). It would then be possible to compare different shapes, either from different segments of a work or of general tendencies throughout works. Moreover, the size of the area confined within the same shape could then function as a measure of the degree in which a segment is generally close to the maximal or the minimal. Even if I have started out from a division of

---

[25] In Wishart's *On Sonic Art* (1996) he describes the process of metamorphosis……………….

the dimensions in three values corresponding to the maximal and minimal extremes and one middle value, it is possible to differentiate it further, e.g. in a five point scale with "maximal", "bordering to minimal" and "minimal low", "minimal high" and "bordering to non voice" as the "values".
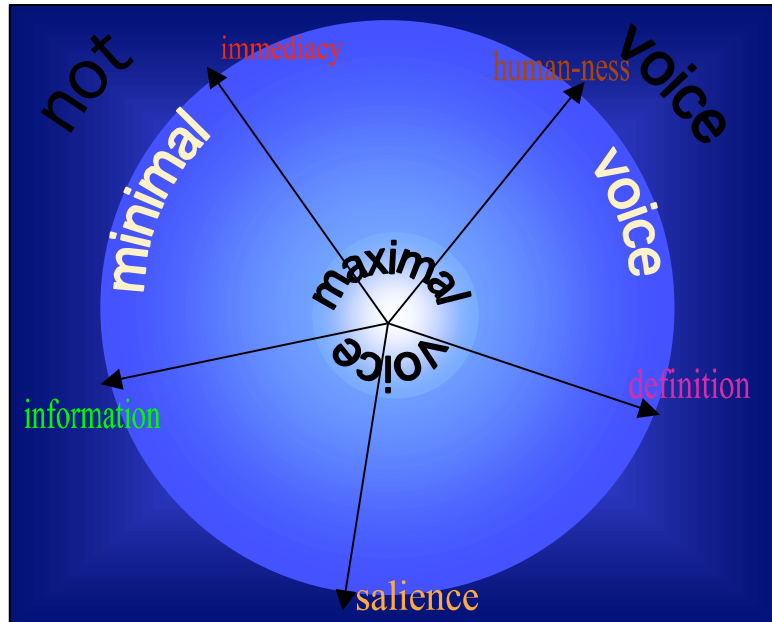

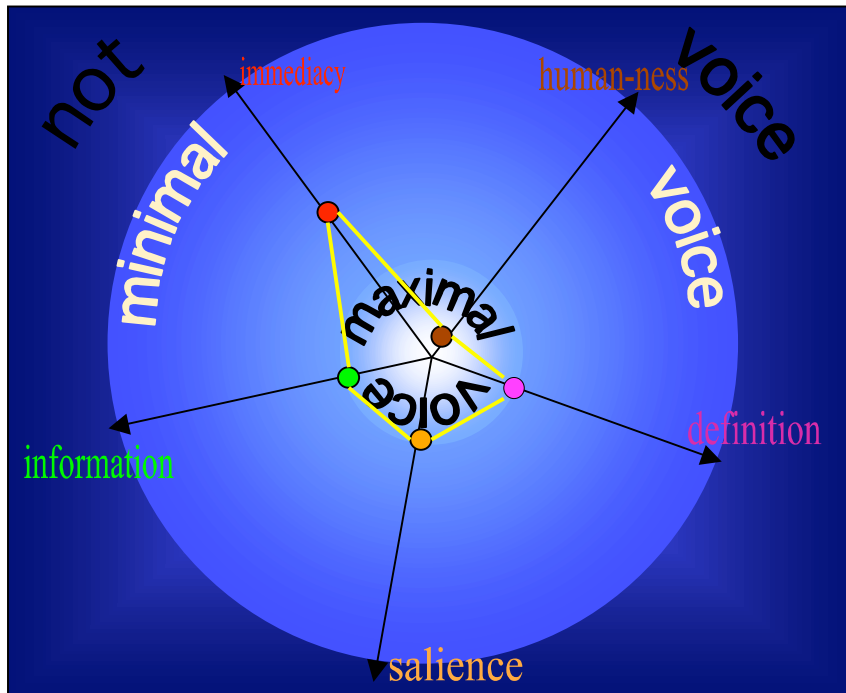**Figure 7: 5-dimensional model of maximal and minimal voice**

Figure 8: 5-dimensional model with points and lines

To give a thorough justification of the proposed model, one would need to apply it to as many different types of vocal expressions in electroacoustic music as possible, and then to see if the model seems to embrace issues that are significant to the perception and interpretation of the segments. One of the issues that has to be addressed is whether the axes are fully independent, or whether they show tendencies to co-vary. Even though the model as of today can turn out to have many shortcomings, I still believe that its main ideas can be further developed into one among other tools that can deal with the multiplicity of works that apply different forms of vocality in electroacoustic music.

Literature

Aksnes H. (1995) Det symbolske og det semiotiske: Et møte mellom Julia Kristevas poststrukturalistiske teori, kognitiv teori og Arne Nordheims Nedstigningen. *Studia Musicologica Norvegica* 21,

Belin P., Zatorre R.J., Lafaille P., Ahad P. and Pike B. (2000) Voice-selective areas in human auditory cortex. *Nature* 403, 309-312.

Bergsland A. (1999) Vævet av "Roser" : Olav Anton Thommessens "Vævet av stængler" som lesning av Obstfelders "Roser". Hovedoppgave. NTNU, Trondheim.

Bergsland A. (2000) Sangkomposisjon som lesning. *Musikklidenskapelig årbok, Musikkvitenskapelig institutt NTNU*

Bolter J.D. and Grusin R. (2000) Remediation. Understanding New Media. MIT Press, Cambridge, Mass.

Bosma H. (2003) Bodies of evidence, singing cyborgs and other gender issues in electrovocal music. *Organised Sound* 8(1), 5-17.

Chion M. (1999) The voice in cinema. Columbia Univerity Press, New York.

Emmerson, E. & Smalley, D.: 'Electro-acoustic music', *Grove Music Online* ed. L. Macy (Accessed [23.01.2006]), http://www.grovemusic.com

Emmerson S. (2000) 'Losing touch?':the human performer and electronics. In Music, Electronic Media and Culture. Edited by Emmerson S. Ashgate, Aldershot. pp 194-216.

Handel S. (1989) Listening. An Introduction to the Perception of Auditory Events. MIT Press, Cambridge, Mass.

Godøy R.-I. (1997) Foramalization and Epistemology. Universitetsforlaget, Oslo.

Grant J. (1993) Encyclopedia of Walt Disney's animated characters. Form Mickey Mouse to Aladdin. Hyperion, New York.

Laver, J. (1991) *The Gift of Speech*, Edinburgh University Press, Edinburgh, UK.

Lee K.M. (2004) Why Presence Occurs: Evolutionary Psychology, Media Equation, and Presence. *Presence: Teleoperators and Virtual Environments* 13(4), 494-505.

McAdams S. and Bigand E (1993) In Thinking in Sound - The Cognitive Psychology of Human Audition.. Oxford University Press, Oxford.

Nass C. and Brave S. (2005) Wired for speech. How Voice Activates and Advances the Human-Computer Relationship. MIT Press, Cambridge, Mass.

Norman K. (2004) Sounding Art. Eight Literary Excursions through Electronic Music. Ashgate, Aldershot.

Schaeffer P. (2002) Traité des objets musicaux. Le Seuil, Paris.

Smalley D. (1997) Spectromorphology: explaining sound-shapes. *Organised Sound* 2(2), 107-126.

Stacey P. (1989) Contemporary tendencies in the relationship of music and text with special reference to Pli selon pli (Boulez) and Laborintus II (Berio). Garland Publ., New York.

Stamelman P. (1980) A chat with Mel Blanc. In The American animated cartoon - a critical antology. Edited by Peary D. and Peary G. E.P.Dutton, New York.

Sterne J. (2003) The audible past. Duke University Press, Durham & London.

Wesling D. and Slawek T. (1995) Literary Voice - The Calling of Jonah. State University of New York Press, Albany.

Wishart T. (1996) On Sonic Art. A new and revised edition. S.Emmerson (ed.). Routledge, London.

Young J. (1996) Imagening the source: The Interplay of Realism and Abstraction in Electroacoustic Music. *Contemporary Music Review* 15, 73-93.

Music references

Bodin, Lars-Gunnar	*On Speaking Terms II, Poem No.4*, from CD *The Pioneers – Five text-sound artists*, Phono Suecia PSCD 63
Chowning, John	*Phoné* (1980-81), from CD *Turenas; Stria; Phoné;Sabelithe* Wergo WER 2012-50
Coulter, John	*Shifting Ground* (2005), DVD. Griffith University
Dodge, Charles	*Speech Songs*, from CD *Any Resemblance is Purely Coincidental*, NA043

Harvey, Jonathan	*Mortuous Plango, Vivos Voco* (1980), from CD *Computer Music Currents 5*, Wergo WER 2025-2
Lewin-Richter, Andres	*Caminando... (recuerdo de Nicolás Guillén)* (2002), radiophonic creation presented at XVIII Festival Internacional de Música Contemporánea de Alicante
Parmerud, Åke	*Les Objets Obscures*, from CD *Invisible Music*, Phono Suecia , PSCD 72
Rapisarda, Giuseppe	*The Day Before*, Electroacoustic piece presented at Zeppelin 2004, Zaragoza, Spain
Ratkje, Maja	*Voices*, Rune Grammofon RCD 2028
Reich, Steve	*Come Out* (1966) , from CD *Early Music*, Elektra Nonesuch 79169-2
Wishart, Trevor	*Red Bird* (1977), from CD *Red Bird – Anticredos*, EMF Media EM 122
Wishart, Trevor	Vox 5 (1989), from CD *Computer Music Currents 4*, Wergo WE113
Wishart, Trevor	*Blue Tulips* (1994), from CD *Voiceprints*, EMF CD 029