QUALITIES AND FUNCTIONS OF MUSICAL TIMBRE

Stephen McAdams & Kaija Saariaho
I.R.C.A.M., 31, rue Saint-Merri, F-75004 Paris, France
tel: (33.1) 277.12.33, poste 48.38
uucp net address: decvax!mcvax!ircam!smc or kaija

ABSTRACT

Many questions are being raised about the ability of various dimensions of timbre to carry form in music. This paper will describe both a musical and psychological search for form-bearing elements in music. What is their nature and how do they behave? Can we describe them in general so we know what to look for with respect to timbre, and so we can advance our understanding of how to construct a language with the many new musical elements that have become available through digital synthesis? Six criteria believed to be essential for a form-bearing element are described. We investigate the function of timbre as a musical element in poetic speech and consider the implications for underlying psychological structures that allow this sonic art form to have such a high degree of sophistication. Then we discuss what we might learn from poetic speech for more general timbral considerations in music and examine the timbral derivatives of spectral form and frequency structure with respect to their musical qualities, functions and possible transformations.

1. THE NATURE OF FORM-BEARING ELEMENTS

When we speak of *Form* we are specifically referring to form as perceived or apprehended, not necessarily the *structure* created by an artist. Many experiments in the Twentieth Century have shown that structuring in itself is not enough if it cannot be apprehended or decoded for various reasons, including biological or psychological limits on the processing of optical or acoustic structures. Therefore it is important to consider the psychological constraints on carriers of form. We propose six criteria for investigating the form-bearing potential of a proposed musical dimension.

1. **Form-bearing elements are perceptually differenciated into discrete categories.** This means they are either artificially restricted to differentiable categories, as is the case with musical pitch, or a continuous acoustic dimension is broken into categories by the auditory system, as is the case with speech phonemes. One important characteristic of musical form is that it is accumulated in memory across time. This implies that if some element is easily remembered, its potential contribution to a form is greater than another element which is very difficult to remember. There is a great deal of psychological evidence that we remember discrete dimensions much better than continuous ones. This is not to say that in the case of musical pitch, for example, we cannot appreciate whether a note is out of tune or not: of course most of us can; but, if a melody is played that continually changes in pitch and only passes through the appropriate pitch at the right time, one generally has a difficult time remembering this form in order to recognize it later or to compare it with another similar form. Whereas a categorized dimension may serve the building of structure, continuous variation along this dimension may serve for expressive tranformations of the "ideal" categories as we find in expressive intonation and in micro-tonal ornamentation in Indian and Arabic music, for example. The classification of points along a single dimension or several correlated dimensions helps also with the characterization of the properites of musical entities. It is with the contrasts between these entities that musical structure is built. The notion of "property" applies as much to the quality of an entity as to relations among entities, or their function within a musical context. An entity, for us, is an auditory image whose constituent elements behave coherently as a group. This holds as much for the frequency components of a complex tone as for the different voices in homophonic orchestral writing.

2. **The perceptual categories are ordered so that the relations among them are functional.** This relates to the syntactic characterization of the properties of musical entities, and is a very strong characteristic of the pitches in modal and tonal music systems, e.g. the functional relationship between the seventh degree of the major scale and the tonic is different than that between the seventh and the sixth. These relations are not, of course, entirely independent of the context, but they certainly have strong tendencies of their own due to their position within the system that often makes them difficult to circumvent in non-tonal music. This ordering may be of different types, e.g. the hierarchical ordering of tonality versus the serial ordering of twelve-tone music. The way of ordering often places severe constraints on the possibilites of musical form that can successfully be realized within such a system. The classification schemes and their ordering must therefore reflect psychological possibilities or these structures will not be decodable by the listener and thus not contribute in themselves to an appreciation of musical form. One of the great problems of our time in music, with the demise of a common musical language in the Western world, is the invention of psychologically relevant functionalities within the increasing number of musical dimensions of compositional interest.

3. **The functional relations are of varied strengths and types which allows for the building and release of tension.** In tonality, for example, the strenth of the relation between the first and fifth degrees of the scale is much greater than that between the third and fifth, and indeed such movement is structurally less important for tension than in the former case. There are many different ways to describe the function and strength of relations among entities which include similarity/dissimilarity, consonance/dissonance, dominance/submission, etc. It is important that the nature of the relations be rela-

tively independent of the local context of the piece and yet closely bound to the "context" of the ordering system itself.

4. **Attention can be paid directly to qualities of a category, to qualities of relations among categories, or to combinations of relations.** With pitch for example one can focus on the pitch itself, to the quality of an interval between two pitches or to the quality of a chord as an assembly of intervals. In different musical contexts one or the other of these levels of attention are more important for the extraction of musical structure. This demonstrates the importance of emergent properties of perceived relations or grouping, a realm where the psychological processes of auditory organization play a strong role in what is available at a given moment as musical material emerging from the acoustic structure. There is obviously an interest in combining qualities as well as their emergent relations. This raises compositional problems concerning the interdependence of manipulating both musical entities and the properties that emerge from their interactions.

5. **The categories, functional relations and ordering within a classification system must either reflect the existing structure of the mind and world or be susceptible to learning by listeners if the structures are to be apprehended as musical form.** This is a large complicated area in the cognition of music that evokes the schemata of perceiving as stored organizations of perceptual experience. Obviously, one of the most important elements here is the categorization, structuring and then committing to memory of experience. Important work in the development of perceptual skill has shown that the basic machinery for interpreting the world exists at birth and that perceptual learning is a process of progressive discrimination, or reduction of discrimination, i.e. the elaboration of innate schemata and the further division of existing categories, or the removal of existing boundaries between categories. In essence, we seek the biological limits to the processing of structure and form. Clearly, there are forms that we cannot appreciate since we have no possibility of developing the necessary perceptual tools to discern them. This argument is a bit dangerous in the sense that it easily becomes cannon fodder for conservative attitudes that would use it to debase works they are not able to understand. But isn't art both a fight against conservatism and an exploration of the limits of the possible?

6. **Relations among categories must be able to maintain a certain degree of invariance under various classes of tranformations.** If patterns composed within a set of dimensions are not susceptible to being varied *and* still being perceived as similar, then the set of dimensions cannot be strong contributors to musical form. Afterall, variation and transformation are some of the main features of music throughout the world. The limits of perceiving a transformed pattern as somehow related to an original one point to psychological limits of viable processes of musical transformation. Pitch patterns, for example, can be transposed (a translative transformation in the log frequency domain maintaining both interval and contour information), played in a related key (a mapping transformation that maintains contour and interval class information), expanded or contracted (a multiplicative transformation that maintains contour while losing interval information), etc. and still be perceived as related to the original motif. Psychologically this implies a mental representation of the original concept or model (motif) which maintains certain properties in its structure during these transformations and which are perceivable as such. An exploration of this domain could ultimately open up a wide territory of rich functional substitutes for the process of variation so common in musics of the world. However, it also implies limits to the possibilities of transformation through the many quality dimensions of musical entities, and limits to the possibilities of significant coherent structuring of musical material that is to be transformed.

While most of the examples cited above were drawn from pitch structure, since this is more easily understood by most people in our culture, we are interested in exploring the aspects of timbre that can satisfy these criteria and the ways in which they can be used to build musical forms. We say "aspects of timbre" because it is now very well understood that timbre is a multidimensional musical element, and it is our belief that not all of the dimensions of timbre will show themselves to be equally strong with respect to their ability to carry structure. This is not to insinuate that not all of these aspects have musical importance, but we are trying to understand which elements can be used structurally and which will necessarily remain in the realm of ornamentation and expressive coloration. There is also the question of understanding as much as possible the *a priori* "laws" and "principles" by which we can determine whether a dimension is susceptible to rich structuring. Part of the problem of evaluating the role of timbre itself in this way is that in most existing Western music of any esthetic import, the use of timbre is very heavily subordinated to pitch and rhythmic structure. To search beyond these limits we asked ourselves if there was any sonic art form in which timbre plays a much stronger role than pitch in the carrying of form. Of course, the prime candidate for such a consideration is poetry, and more specifically, poetry in which the poet has a fine ear for sonic form as well as for semantic form.

2. POETIC SPEECH AS A TIMBRE-BASED SONIC ART FORM

The human voice is the sound source for which we probably have the finest timbral discrimination and the richest structural comprehension. There is massive evidence from neuropsychology and from experimental and cognitive psychology that we have special brain mechanisms for the decoding and processing of human speech. Humans also have the most sophisticated vocalization apparatus of any known animal species. Most of the information in the speech stream is carried by acoustic elements that we normally associate with timbre in music. While pitch plays an important role in prosody (the segmentating and phrasing functions in speech) and to a certain extent in semantics (communicating aspects of meaning with inflection that might otherwise be ambiguous), its role as a structural element is relatively weak in non-tonal languages such as the Indo-european languages when compared to the role played by vowels and consonants. Chinese and many African languages use pitch as a phonological element, where the consonant-vowel complex changes meaning if the pitch pattern or relative pitch height is varied. In African song in cultures with tone languages, the fact that relative pitch height carries meaning places certain constraints on melodic contour when words are sung to music.

2.1 Sonic form-bearing elements in poetic speech

Considered very simplistically, the following are among the more important sonic form-bearing elements of poetry:

1. **Vowels** are the timbral resonance of the vocal tract and may be associated with spectral color in musical timbre. One kind of musical structure carried by vowels in poetry is rhyme. Acoustically they are described by the spectral form or envelope. They are of variable duration and often when added emphasis is desired on a certain syllable the vowel is lengthened. Other uses of vowel

structures include clustering of vowels and patterns where certain vowels are conceived as "leading tones" to other more structurally important vowels. One might conceive of the ordering of vowel structures as a timbral "harmonic progression" implying a structural ordering of the main vowels used in a poem. This structuring can serve the musical function of tension and release.

2. **Consonants** are generally due to rapid changes in the resonances of the vocal tract often accompanied with noise produced by constrictions in the mouth cavity when the tongue is placed near the roof of the mouth or the teeth are moved close to the tongue or lips. Consonants may be associated with attack and decay characteristics as well as with various transition sounds in musical timbre. Alliteration is a poetic sound structure based on patterns of similar consonants. Some consonants are of more or less fixed duration (like the stop consonants b, d, g, and p, t, k) while others can be lengthened for emphasis (like the fricatives and sibilants - f, v, th, s, z, sh). Consonants are the articulations of and transitions between vowel carriers. In this way they may be conceived as a complex pivot between various vowel "states" which has a very strong signature of its own and even changes slightly the nature of the vowel it is framing, giving the vowels form and place in the timbral line. Consonants are also time markers in the way that they induce sudden changes in the acoustic stream.

3. **Rhythm** derives are great deal from stress patterns as well as the patterns of consonants and lengthening of vowels and certain consonants. It can be extracted from changes in overall energy level, sudden pitch changes, spectral changes that define consonant boundaries, as well as changes in noisiness or harmonicity that vary with voicing (whether or not the vocal cords are vibrating). Breaks in speaking, i.e. juncture, are also an important ingredient in the perceived rhythm of speech.

4. As mentioned above, **intonational pitch** is also an element that helps with the segmentation and phrasing of speech. This is an element that is totally absent in the notation of poetry on the page and which poets use more or less well in reading their own and others' poems. Again, our feeling is that this is more expressive than structural when compared with vowels and consonants.

Given our orientation toward timbre structures, the main focus of this paper will be on the production, perception and organization of vowels and consonants.

2.2 Production and encoding of phonemes

The linguistically significant features of speech sounds can be accounted for by describing speech as a concatenation of discrete phonemes (Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1972). The phoneme is, above all, an abstract class of elemental sounds which are sufficiently similar among themselves and distinct enough from those of another class to be clearly perceived and identified in spite of idiosyncratic variations due to speaker and content. For example, we need to be able to group together all of the versions of a given vowel like /a/ as spoken by people with different accents, and consider them as instances of the same functional element, and distinguish this group from the group of instances of another vowel like /o/. These different versions of the same functional entity are called allophones and in music we might consider using them as expressive variations of the same timbre group.

Speech science has delineated a system of elementary "distinctive features" of phonemes whose individual properties and combination rules account for the apparently invariant characteristics distinguishing one phoneme from another. Each phoneme can be considered as a constellation of these distinctive features, and the features themselves are very closely bound to the way in which the phoneme is produced. This implies that we may have mechanisms for deducing the physical causality behind the sounds we hear and then for organizing and interpreting them accordingly. Vowels, for example, might be classified according to how open the mouth cavity is and how close to the front of the mouth the tongue is raised toward the pallette, and how much the mouth is rounded. And consonants can be classified according to whether vocal cord vibration starts with a delay or not, by place in the mouth where the air cavity is constricted by the tongue or lips, by whether air is let pass at the point of constriction before the mouth cavity moves to form the succeeding vowel, etc. All of these subphonemic features have their acoustic correlates, but one thing that is fascinating about speech is that there is a high degree of overlap in time between the features that are present for a consonant phoneme and its succeeding vowel. This means the acoustic information that carries these features is highly encoded and special perceptual mechanisms are needed to decode this information and recover the original phonemes that were intended. This also means that the information from phonemes are not simply sequenced one after the other but are heavily embedded in one another such that each phoneme exhibits a significant restructuring of its acoustic cues according to the context.

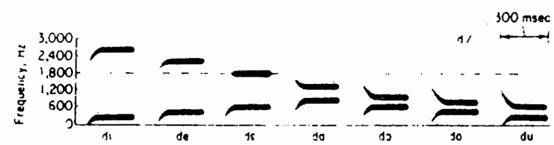

Figure 1. Spectrographic patterns of the first two formant frequencies for the synthesis of /d/ in front of various vowels. The relative frequencies of the formants' steady-states specify the vowel. Note that the initial transition portion of the higher formant changes direction depending on its frequency position: ascending when above 1800 Hz, constant at 1800 Hz, and descending below 1800 Hz. Essentially, the formant trajectory is moving away from 1800 Hz toward the vowel's formant position. This reflects the place of articulation of the consonant /d/ and also shows how the acoustic structure depends both on the consonant and its succeeding vowel. Despite these significant acoustic differences, these are all heard as /d/. (from Liberman *et al.*, 1972, Fig. 2.2, p. 20)

2.3 Decoding and perception of phonemes

The speech code, then, must make use of special perceptual mechanisms that serve as its decoder. And the functionality of the phoneme must be intuited as well. Considering the mental organization of perceptual relations among phonemes can be a useful step toward organizing similar sounds for musical purposes.

When studies of the perceived similarities among phonemes have been conducted in a manner similar to the timbre space studies of Grey (1977) and Wessel (1979), the multidimensional perceptual structure appears to have dimensions that correspond to the mode of articulation of the phoneme. That is, phonemes with similar modes of production, like closed, back vowels (/i/, /I/, /ε/) or like voiced stop consonants (/b/, /d/, /g/) tend to be grouped together in the "phoneme space". These diagrams also show which aspects of production are the strongest with respect to the perceptual organization of the phonemes.

Vowels can be fairly well represented in three dimensions which portray height (openness of the mouth cavity), advancement (placing of the tongue - front/back) and retroflexion (curving of the tongue toward the roof and drawing the tongue toward the rear of the mouth). (see Figure 2) A good deal of investigation of the structure of vowel space and its import for timbre organization has been conducted by Slawson (1985).

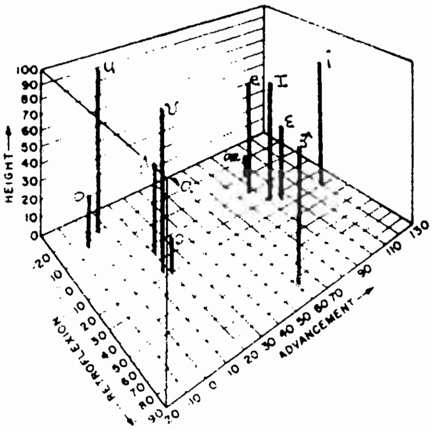

Figure 2. 3-D multidimensional scaling solution representing the similarity structure of 12 American English vowels in terms of 3 distinctive features. The units on the axes are arbitrary. The vowels are /o/ - go, /u/ - glue, /ʊ/ - hood, /ʌ/ - mud, /a/ - father, /ɔ/ - awe, /ae/ - had, /e/ - great, /ɪ/ - hit, /ɛ/ - bet, /i/ - sheet, /ɝ/ - nerd. (from Singh & Woods, 1971, Fig. 3, p. 1864)

Consonants need to be represented in 3 or 4 dimensions which portray manner (stopped, fricative, etc.), place (of constriction by tongue or lips), orality (nasal or oral), and voicing (presence of vocal cord vibration at the beginning of consonant production). The interest in these spaces obtained by perceptual comparison is that they reflect the modes of production, which would lead one *a priori* to consider that a sense of physical causality is important behind the categorization of timbral features. The implications for synthesis are that there are well-defined dimensions that can be fruitfully extrapolated into musical material and along which this material can be organized.

Evidently a sense of the mental relationships of these phonemes in isolation or attached to simple syllables does not suffice to predict their utility or interest in a musical or poetic situation. Where they become interesting and get closer to carrying meaning and form is when one develops a language which organizes and constrains their combination into morphemes or syllables in language and into sonic events in music. One notes in language that vowels (the carriers or "steady states") can mostly be combined and arranged into sequential complexes with few constraints, though some combinations are found much more seldom in a given language than other combinations. For example the main vowels /a/, /o/, /i/, and /u/ are much freer in their combinatoric capacity than the American retroflexive vowel /ɝ/ (as at the end of father). This vowel when in combination with other vowels is always in a succeeding postion. The constraints on combinations of consonants into complexes appear much more extensive: one never finds the combination /lb/ at the beginning of a syllable in English for example, while /bl/ is not uncommon. In the construction of a language of timbres the rules of phonetic combinations, and particulary the constraints on these, could serve as a level of organization preceding a timbral syntax.

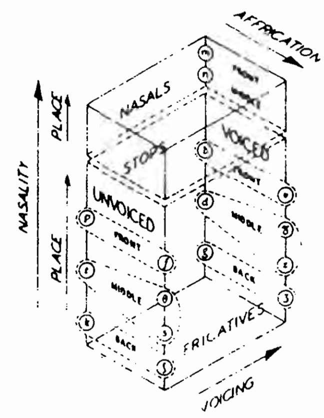

Figure 3. Representation of 16 English consonants in terms of 4 distinctive features. The unfamiliar symbols are /θ/ - *th*ink, /ʃ/ - *sh*eet, /ð/ - *th*at, /ʒ/ - gara*g*e. (from Shepard, 1972, Fig. 4.8, p. 88)

2.4 Towards meaning and form

The development of our thought on this is clearly incomplete since it focuses too heavily on the phonetic aspect of speech without considering the complexity of meaning and function which, of course, determine the use of one phoneme rather than another and also determine the relative functionalities among various phonemes. We have these aspects in mind but our initial interest in directing this line of thinking toward digital synthesis has biased us toward a more bottom-up approach.

Work by the structural linguists, among which Roman Jakobson is important for his considerations of the linguistics of poetry, has focused a great deal on the fact that the existence of a phoneme category does not mean that it has functional, phonological significance in a given language (cf. the essays in Jakobson, 1976/1978). The value of a phoneme is in how it *functionally* distinguishes itself from other phonemes. That is, if you change a phoneme at a particularly place in a word and that does not change the meaning of the word, then these are more likely two instances of the same functional class. Classic examples are the lack of distinction between /r/ and /l/ in some Oriental languages, /b/ and /v/ in Spanish, /v/ and /w/ in Hungarian, etc. While newborn babies seem capable of hearing these as different categories they learn that the difference is not significant within their mother tongue and subsequently ignore such distinctions until such time as they learn another language where these distinctions are important (Eimas, 1985). The problem in music would be one of building such a musical timbral language where these functionalities become apparent in the listening to the music itself, since we do not have a common language to draw from in this situation. The development of a distinctive functionality is the realm of timbral phonology which distinguishes itself from phonetics in concentrating on function and contribution to meaning rather than on acoustic structure.

The harmonious arrangement in time of the complexes of such timbral elements is the realm of syntax. This is where the compositional process really begins, since it is at this level that the complexity of relationships can begin to take on musical form. Below we will discuss some thoughts and preliminary excursions in this direction with different kinds of sound material. Ultimately, the construction of the myriad relations and references among the elements of the sound material built into an auto-referential network is where musical meaning, the realm of semantics, begins to grow.

3. TRANSLATIONS TOWARD TIMBRE IN MUSICAL CONTEXTS

We must at some point step away from considerations of language proper and return to music. But is the distance as great as all that with respect to timbre? Imagine the timbral complexity that is used with great expressive appeal in Indian tabla playing. A good tabla player essentially makes the drums speak to us with an uncanny vocal quality. It is no great surprise to find out that initiates learning to play the tabla start by learning a vocabulary of syllables with which they recite their tabla exercises ***before ever touching the drum!*** One senses strongly that there is a very sophisticated syntax behind the combinations of these elemental sounds that gives tabla playing its richness with respect to musical meaning. A great deal of this richness resides as well in the sophistication of the rhythmic system, but the timbre is an essential element. Can we consider that perception of this richness and a sense of its vocality somehow implies that part of our musical appreciation processes are "squatters" in the domain of speech, at least where timbre is concerned? The fact that appreciation of speech structures is so rich and that voice has always been a very important musical instrument is reason enough to at least consider the possibilities of invoking speech processes in the perceptual organization of timbre.

One shouldn't try to drive such analogies too far in a literal sense. Speech sounds, as we have said, have special processing mechanisms and it appears that non-speech sounds are processed differently than speech sounds. Where the anaogy may be useful is in considering that timbre perception is in some way dependent on a decoding of the physical behavior of the sound source. Certainly, we are not limited to sounds produced by physical models or those that mimic physical sound sources, but more and more we are beginning to ask ourselves why it is that many intricate synthetic sounds used in computer music are embedded in musical structures that so often *sound* very simple, even simplistic. It seems entirely likely that the resulting sound structures, while being perhaps acoustically rich, lack necessary coherence in their behavior as interpretable by the musical ear. It is this behavioral coherence that allows the ear to follow a sound source, to identify it sufficiently to be able to put it into relation with other sounds. This perceptual coherence also allows the sound source a greater range of flexibility as far as movement through its different musical dimensions, particularly those giving rise to timbre. It is a common experience in computer music that a given sound simply transposed to another frequency range often acquires a new identity and any attempt at creating more complex musical structures, like polyphony, are very difficult since the ear cannot maintain the integrity of the hypothetical source carrying the musical form. What we are searching for is a way of creating complex sound sources which maintain a perceptual coherence. This coherence might be obtained by creating a believable pseudo-causality in the behavior of the sound source's timbre by its covariation with register and dynamics, for instance. A pseudo-physical model that describes a source in terms of resonances and a source of excitation allows this possibility.

This notion of coherence of behavior does not have to be limited to sources of sound, though. It can also be extended into the behavior of groups of sounds such as chords or composite timbres derived from the grouping of several sources into a single musical image.

The idea for our present work is to attempt to draw from the psychological structures that underly speech perception in terms of the ideas of carriers (steady-states, functions, resonances) and transitions (attacks/decays, operators, articulators, pivots). We have focused on two main aspects of sound sources, i.e. spectral form and frequency content, and have begun investigations into the timbral qualities and musical applications that can be derived from these. An important point in this choice is that the two aspects of sound can be controlled and shaped independently, depending on the synthesis methods chosen. Also, these elements can be considered to correspond more or less closely with the notions of resonance structure and source of excitation, respectively, in many kinds of physical sound sources.

3.1 Spectral Form

This is the acoustic domain most closely related to that of vowels and to many of the consonants in speech. In the CHANT-FORMES environment at IRCAM there exists a growing library of phonemes that are primarily based on vocal phonemes and which were used extensively in the second author's composition *Jardin Secret I* (ICMC '85). Work by Rodet & Dépalle (cf. these Proceedings) has begun to develop systematically a library of procedures that can be used in the same sense of carriers and transitions that we envision here. In the realm of carriers, we might consider more slowly evolving spectral forms whose qualities would be identified by the non-temporal distinctive features of a multi-resonance system. These qualities would include the classic notions of brightness and spectral color as well as the many aspects of vowel identity such as advancement (acuteness), compactness (smallness), closure, tenseness and nasality (Fant, 1973, chap. 8; Slawson, 1985). The development of simple ways of moving between these states in such a way as to not evoke the presence of a transitional phoneme is relatively straightforward in FORMES.
Transitional phonemes are really a way of getting between two states by a more circuitous route, moving in a more or less rapid way that evokes the presence of an articulatory element. This can serve to demark more clearly the change between carriers and, as shown by the structural possibilities of alliteration in poetry, can introduce additional material for the elaboration of a musical form. Their qualities are identified by temporal distinctive features such as the way in which each resonator changes over short periods of time with respect to the others. In the voice these formant transitions usually last less than about 100 msec in some consonants and would be induced by changing place of articulation and nasalization. In musical instruments these transitions last less than about 50 msec and are seen in the relative onset or offset times of different frequency regions in the attack and decay of a tone. In synthesis, of course, the specification of these time periods is completely open and can be explored at will. Many consonant transitions, when stretched out in time, lose their original phonemic identity and acquire an entirely new one: a simple transformation yielding new timbral forms.

3.2 Frequency Content

We consider the area of frequency content to comprise two main parts: noisy signals and "tonal" signals, i.e. those with definable frequency components. The latter category may also be divided into harmonic and inharmonic sounds. In a sense, the passage from harmonic sounds through inharmonic sounds of increasing complexity and then to noise presents a field of many possible paths between noisy and clearly pitched sounds, since inharmonic sounds have either multiple or dispersed pitch sensations and with increasing complexity become more and more grainy or rough sounding moving toward noise as the spectral density increases beyond the abilities of the auditory system to resolve the individual partials or to extract periodicities or near-periodicities in the waveform.

As carriers, the contrast between noisy and clear sounds can be a strong structuring element as witnessed in the second author's piece *Verblendungen* (ICMC '84). The extent to which degree of roughness of sounds can be used for structural purposes is not clear though, since it seems difficult to create any sense of categorized scale of roughness. Nonetheless, this quality can be of great importance at the level of texture and surface in a piece.

The whole realm of possible "tonal" sounds, particularly with respect to degree of harmonicity, is only beginning to be explored. Ideally one would like to be able to characterize inharmonic sounds on the basis of interval color, deriving from a global sense of the quality of pitch intervals present in the sound, in a manner similar to the perception of chord qualities. Earlier attempts at making an *a priori* theoretical approach to the search for and characterization of these sounds were not overly successful (McAdams, Gladkoff & Keller, 1985), mostly due to the inadequacy of the perceptual models used. However, working from a base of existing inharmonic instrumental sounds, i.e. multiphonics, Assayag, Castellengo & Malherbe (cf. these Proceedings) have made an important step toward finding ways of grouping these sounds according to possible musical functions and relations. Therefore we have opted for a more empirical method to search for and evaluate inharmonic sounds that are pleasing to the ear, and will eventually follow out the process of characterization in a manner similar to Assayag *et al.*

We have developed a system on the 4X synthesizer with the score language 4XY developed by Robert Rowe which allows one to build inharmonic complexes by ear from scratch, to progressively add frequency components to a pre-existing complex, or to modify existing components. All of this occurs with the help of an interactive graphics and auditory feedback. Then this or any other pre-existing sound can be transposed to other registers or loudness values and compared with the original, adjusting either its spectral envelope or frequency content in order to construct a timbral identity that maintains its coherence with changes in register and dynamics. This also proves useful for finding scales within which the melodic and harmonic intervals are custom-tuned for this particular family of inharmonic sounds. We envision the development of tools that allow us then to describe the necessary interpolations between tunings when "modulation" from one family of inharmonic sounds to another occurs. Once several such families are assembled we will investigate the possible sequential and harmonic relations among families in order to develop a language of sequencing and progression for a given composition. One drawback to the present system is that the system on the 4X does not yet have the flexibility of spectral evolution that has been developed with the phoneme libaries in FORMES. A discontinuity is introduced between exploration and musical application that is not very convenient.

Noise and inharmonicity are also important elements in transitions. The presence of noise is characteristic in many consonants with a high degree of affrication (unvoiced fricatives like f, sh, s; voiced fricatives like v, j, z) or in unvoiced stop consonants (p, t, k). The timing (relative to voicing) and spectral character (place of articulation) of the noise are very important in the identification of these sounds, thus the nature of transition between or coarticulation of noise and harmonic sound can provide important articulatory information. In instrument sounds, one often finds an instability in the setting into vibration of a string or a column of air, which contains both noise and inharmonic elements. Psychological studies have shown these transitory elements to play a very strong role in the identification of instruments. All of these elements must eventually be characterized in terms of temporal distinctive features in the realm of frequency content.

3.3 Musical Functions

We will consider three areas of the functional use of these ideas and material in music: the classification of the material, shaping it to achieve directed motion, and transformations and interactions between the sets of poles carrier/transition and resonance/excitation.

It seems possible in carefully considering the implied nature of production of sounds, in terms both of resonance and excitation, that a vast extension of the notions of timbre, vowel and consonant spaces can be made and can be useful in guiding the development of synthesis-by-rule in the creation of complexly, but coherently, behaving sound structures. With an ear toward the possibilities of the physical intuitions embedded in these representations, they can be made to be more clear with respect to the implied possibilities of musical organization. Within such a framework, classification of the qualities and relations among the items would be more easily arrived at. The classification of spectral form related qualities is somehow more intuitively achieved especially for sounds similar to the voice. For arbitrary sounds removed from those of the voice or other highly familiar sounds the ability to discriminate among them decreases. As the work of Vandenheede & Harvey (cf. these Proceedings) has shown, this balance between familiarity and ambiguity of identity is a fragile realm to deal with. The same holds for the classification of frequency content related qualities, particularly inharmonic sounds, and as mentioned above we have taken an empirical approach to the establishing of qualities and functional relations among such sounds and among families of sounds.

According to Lerdahl (1985) the kind of organization that is most likely to generate rich musical structures is a hierarchical one. He has begun to investigate the creation of timbral hierarchies of spectral color by developing families of similar sounds and then sub-families and super-families, etc. In a way partially similar to the approach of Assayag *et al* with multiphonic instrument sounds this allows the expression of varying degrees of contrast, cognitive "distance", constraints on the acceptable paths from one family to another, as well as relations that vaguely resemble the functional dissonance and dominance one finds in tonal pitch structures which seem especially rich in their ability to bear intricate forms of this nature. The criteria for form-bearing elements that seem somewhat less obvious with these kinds of musical material are the evaluation of strengths of relation (unless one simply considers cognitive "distance" from a member of one family of sounds to another member of another family along acceptable paths of "modulation") and the evaluation of the relations among perceptual properties that emerge from the combinations of the members of these families. The development of a real-time interactive synthesis environment biased toward these kinds of explorations will, however, aid immensely in the clarification of the possibilities and pitfalls in this area. It will also allow a more thorough and less frustrating exploration of the kinds of transformations of material that still allow perception of similarity or at least musical relatedness.

Of primary importance for the construction of musical form is the ability to develop a sense of movement that is directed. Direction creates expectation. The building, prolonging and satisfying of expectation, of musical tension, is a fundamental element of musical experience. Directed motion implies both a sense of the direction and of the magnitude or strength of the movement. The notion of harmonic interval in pitch for example is relatively neutral in this respect compared to melodic interval, where the force of the motion depends as much on the direction of the interval as on the size. The notion of a directed timbre interval as a sort of multidimensional vector in time is implied by the work of Wessel (1979) on timbral analogies where the sense of directedness is recovered in the choice of a pair of timbres perceived to be analogous to another pair of timbres when played in sequence. A similar sense of timbral vector is well developed in the poetry of Robert Duncan and Michael Palmer who both work with ideas of vowel progressions as tone-leading and harmonic progression in the music of the poem. Since in the space of vowels, though, the relations are not *a priori* hierarchical, the building of structures that allow this directedness to create musical meaning depends on the constraints on the relationships that are established within a given poem. The choice of key vowels and their sequencing is not arbitrary, nor is the choice of which vowels are strengthened structurally by the patterns of stress.

When inharmonic sounds are heard as chords one might use one's intuitions of the functions of musical harmony. A big question, and one we are just beginning to understand, is whether a similar sense of direction is experienced when such sounds are heard as fused timbres as when they are heard as chords. Does the interval color under conditions of fusion function the same as under conditions of defusion? The answer to this question is unclear for many reasons, most of which have to do with problems of auditory organization. One reason is that not all inharmonic sounds can be made to fuse and for these such a comparison cannot be made. What is clear to us now is that for a composer interested in travelling between harmony and timbre, the field of acceptable musical material is relatively limited. (see Figure 4) One of the key factors limiting fusion of inharmonic sounds (aside from the fact that they are already perceived as "multiple" from a pitch standpoint) is spectral density. Sparsely populated spectra with partials that are easily resolved by the auditory system, i.e. the components are all in separate critical bands, are *very* difficult to fuse into a single object, even more so if one desires to sustain them over any period of time. The moment they are not heard as percussive sounds, they are heard as chords. Fusion is more easily achieved with denser sounds, but these are then more difficult to deal with as chords when they are defused. A way of getting around this problem (by leaving aside all ideas of procedural purity) is to find relations between more and less dense spectra that are acceptable and then densify when work in timbre is called for and rarefy when work in harmony is called for. This kind of many-to-few mapping in the frequency domain leaves the field open to a lot of flexibility, but also makes finding appropriate material more difficult.

Other kinds of transformations that we have experimented with to give a sense of progression are gradual deformations of either frequency structures or formant structures by stretching or compressing the relations between component frequencies or formant frequencies (Saariaho, 1985). The results with spectral envelope deformation have given the most interesting results, particularly when combined with inharmonic frequency structures. If the deformation takes place in time, there is a change in the overall sense of spectral color while the material is still recognized as being related to the original material. These transformations work the best when the formant trajectories are continuous. For frequency stretching techniques, though, an annoying sense of incoherent glissandi is obtained with continuous change, so the second author has developed a technique for generating micro-scale steps between the starting frequency and the target frequency. This turns the transformation between structures into a kind of micropolyphony with a temporal texture.

With either of these kinds of spectral transformations or with sequencing of spectral form-based phonemes some very interesting results have been obtained with inharmonic sounds. With spectrally dense sounds, a sense of changing timbre results. However, with sparser sounds one hears the amplitudes of individual frequency components changing over time and this has the effect of playing melodies on these partials as the resonance peaks pass over them. The effect is of creating a wall of either stable or moving frequency structures that is illuminated in various ways by different spectral envelopes which clarify particular parts of the spectrum and create melodic highlighting of the chordal field. The compositional problem is one of finding inharmonic sounds that give rise to interesting melodies in the presence of desired spectral forms.

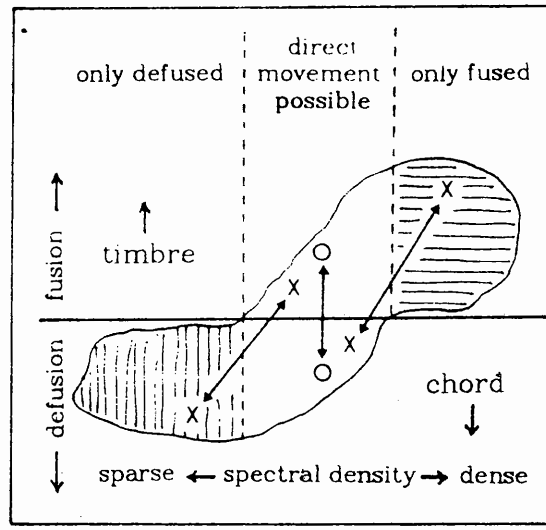

Figure 4. The realm of useful inharmonic sounds in moving between timbre and harmony as a function of degree of fusion and spectral density. Since dense sounds are difficult to deal with as chords, and sparse sounds are difficult to fuse into timbres, only the central region allows moving from timbre to chord (O <-> O) by simply operating on the parameters of fusion (onset asynchrony, vibrato, etc. cf. McAdams, 1984). For other cases an oblique path (X <-> X) requiring a change in spectral density as well as in the fusion parameters is necessary to move from chord to timbre or vice versa.

Another kind of transformation between the domain of spectral color and inharmonic interval color involves the progressive reduction of the bandwidth of the formants, thus increasing the resonance, until eventually they behave like single frequency components. The effect depends strongly on the manner in which the resonances are stimulated. With very low frequency, grainy stimulation (due to a pulse rate near the threshold of pitch sensation at or below 20-30 Hz) and with modulations between spectral forms of different vowels, for example, one hears both the texture of the pulse train and a weakly pitched sound that changes from vowel to vowel and is due to the first formant. This pitch sensation is not so noticeable when the spectral form is stable, but when it moves the comparison is heard almost as a pitch interval. Then, as the bandwidth narrows, this pitch becomes stronger and clearer and is accompanied by the pitches of the partials at frequencies that were previously formant frequencies. This gives a very direct sense of the comparison between the spectral form and its inharmonic chord equivalent. If, on the other hand, the resonances are stimulated at audio frequencies and thus give rise to a strong pitch, the timbral quality of the spectral form is more vocal and one integrates all of the formants into one percept. The pitch due to the harmonics completely overpowers any pitch or timbral height sensation due to the first formant. When the bandwidths are narrowed and this evolves into a chord the relation between the timbre and the chord is much less evident going from a single pitched vocal timbre to an inharmonic chord. In this latter class of cases, the structures that give both interesting spectral forms and chords *and* an interesting relationship between them exist in a limited region as in Figure 4. Here the limiting dimension is related to the strength of the pitch sensation. Again we need to develop a system of finding instances of each that are interesting and related and then finding a musically interesting way to get from one to the other such that the intermediate stages do not cause a loss of musical coherence.

One of the most intriguing problems in dealing with a mixture of familiar and unfamiliar sound processes is the imbalance of their strengths with respect to musical memory. Many listeners upon hearing *Jardin Secret I*, in which there are sections that move from harmonic sounds very close to the human voice, then to floating inharmonic sounds and back again, remarked mostly the vocal sections and seem to have a finer detailed memory of them. Memory for the other sections was more global and the sense of structure, of memorable landmarks in the piece, was apparently provided mostly by the sections with familiar sounds. Again the problem of schemata of habitual perception returns to haunt the composer wishing to push into new areas of perceptual experience (cf. McAdams, 1982). We do not have much clear to say on this subject except to remark its ubiquity and note that it is a socio-perceptual force of great magnitude that must be grappled with.

4. TOWARD PROJECTIVE TIMBRE COMPOSITION

We have made some significant steps in the direction of understanding the nature of form-bearing elements in music and in beginning to evaluate the ability of some dimensions of timbre to satisfy the demands of such criteria. Smaller steps have been made in developing both the tools necessary to explore these dimensions thoroughly and systematically and the environment necessary to synthesize such complex structures with any degree of flexibility.

But technicalities aside, what are we striving for?

Translating from the poet Charles Olson (1950), putting music where he considered the poem:

> *A poem is energy transferred from where the poet got it (he will have some several causations), by way of the poem itself, to, all the way over to, the listener. Okay. The poem itself must, at all points, be a high energy-construct and, at all points, an energy-discharge. So: how is the poet to accomplish same energy, how is he, what is the process by which a poet gets in, at all points energy at least the eqivalent of the energy which propelled him in the first place, yet an energy which is peculiar to poetry alone and which will be, obviously, also different from the energy which the reader, because he is a third term, will take away? This is the problem which any poet who departs from closed form is specially confronted by. And it involves a whole series of new recognitions. From the moment he ventures into FIELD COMPOSITION – puts himself in the open – he can go by no track other than the one the poem under hand declares, for itself. Thus he has to behave, and be, instant by instant, aware of some several forces just now beginning to be examined.* ... **right form, in any given poem, is the only and exclusively possible extension of the content under hand.** (p. 148)

We are searching for a connectivity, at surface and in structure, that projects us, continuously or abruptly, from one moment of musical experience to the next and from the quality of the material into the nature of its organization. The many paths through such a field reveal Form to us. But we need, in exploring *new* material rather than just new relations among *familiar* material, to find a sense of the vocabulary of those substances, a sense that we can imbibe, take into ourselves and live with *away from the computer, for example*, that our musical imaginations can find the forms that are possible within them. Imagination is not the job of the computer program, it is the work of the human spirit. And yet, since the computer has presented us the possibility to explore these new substances, we are obliged by curiosity to explore them and organize them and learn their nature that they may serve our creative drives. This organization involves the development of ways of thinking about and imagining these substances, and until we have taken that step and can walk away from the terminal and *imagine freely*, all of this technology serves no artistic purpose.

REFERENCES

Eimas, P. (1985) The perception of speech in early infancy. *Scientific American*, **252(1)**, 34-47.

Fant, G. (1973) *Speech Sounds and Features*. Cambridge, Mass.:MIT Press.

Grey, J.M. (1977) Multidimensional perceptual scaling of musical timbres. *J. Acous. Soc. Am.*, **61**, 1270-1277.

Jakobson, R. (1976/1978) *Six Lectures on Sound and Meaning*. Cambridge, Mass.:MIT Press (1978), trans. from the French (1976) by J. Mepham.

Lerdahl, F. (1985) Timbral hierarchies and their psychological implications. Paper presented at the 5th Workshop on the Physical and Neuropsychological Foundations of Music, Ossiach, Austria, August, 1985.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P. & Studdert-Kennedy, M. (1972) Perception of the speech code. In: E.E. David & P.B. Denes (eds.), *Human Communication: A Unified View*, pp. 13-50, New York:McGraw-Hill.

McAdams, S. (1982) Spectral fusion and the creation of auditory images. In: M. Clynes (ed.), *Music, Mind and Brain: The Neuropsychology of Music*, pp. 279-298, New York:Plenum.

McAdams, S. (1984) The auditory image: A metaphor for musical and psychological research on auditory organization. In: W.R. Crozier & A.J. Chapman (eds.), *Cognitive Processes in the Perception of Art*, pp. 289-324, Amsterdam:North-Holland.

McAdams, S., Gladkoff, S. & Keller, J.P. (1985) AISE: A prototype laboratory for musical research and the development of conceptual tools. *Proc. 1984 ICMC*.

Olson, C. (1950) Projective Verse. Reprinted in: D. Allen & W. Tallman (eds.) (1973) *The Poetics of the New American Poetry*, pp. 147-158, New York:Grove.

Saariaho, K. (1985) Shaping a compositional network with computer. *Proc. 1984 ICMC*.

Singh, S. & Woods, D.R. (1971) Perceptual structure of 12 American English vowels. *J. Acous. Soc. Am.*, **49**, 1861-1866.

Slawson, W. (1985) Operations on timbre - perspectives and problems. *Proc. 1984 ICMC*.

Wessel, D. (1979) Timbre space as a musical control structure. *Comp. Mus. J.*, **3(2)**, 45-52.