Journal of New Music Research, Vol. 23 (1994), pp. 169–204 0929-8215/94/2302-0204$6.00

Schema-Based Tone Center Recognition of Musical Signals

Marc Leman

ABSTRACT

This paper presents a model of schema-based tone center recognition of musical signals. A distinction is made between passive and active schema-based recognition. Passive recognition assumes the use of a schema as a template, that is, a pattern with which the new information is correlated. Active recognition assumes an active role of the schema itself. The model is part of a theory of musical morphology whose aim is to provide an operational account of music cognition in terms of physiological acoustics (psychoacoustics) and self-organization theory (dynamic systems theory). The model is an example of nonsymbolic research in music imagination and has applications for music analysis as well as interactive music making.

INTRODUCTION

Computer simulations based on auditory models and principles of self-organization show that networks of artificial neurons (in particular Kohonen-type networks, Kohonen 1984), can develop functional organizations that are relevant for tone center perception (Leman 1989, 1990, 1991a,b, 1992a,b). The network, trained with music, develops ordered areas that reflect the circle of fifths. The response structure to chords and tone centers can be considered an aspect of its emerging behavior. The results, which fit with the mental representations for tone center perception (Krumhansl 1990), have led to a theory called "tone semantics". It explains the context-dependent semantics of pitch perception — the way in which tones get a function within a musical context — on four grounds:

i. the specific overtone structure of the tones;
ii. the tone distribution in the time and frequency domain;
iii. the auditory processing strategies; and
iv. appropriate strategies for memory organization.

Computer simulations show that a schema for tone center perception can emerge automatically — without effort. Since the schema is produced by a self-organizing neural network and the inputs are patterns that come from an auditory model, no information from higher levels or "domain specific knowledge" is needed to explain the emergence of the response structure. The schema emerges by purely

data-driven methods.

But data-driven long-term learning is just one aspect of the schema dynamics. The other aspect is about how the schema interacts with the continuous flow of information at short-term. The interaction may rely on constancy-detection with the help of a schema. It is assumed that rapid changing patterns (coming from the senses) are related to memory structures (schemata) that are less vulnerable to rapid changes. As such, new information is resolved by the existing learned knowledge frame specialized in the detection of information that is invariant over short timespans of the information flow. This allows the organism to react efficiently to the stimuli in the environment. The semantics is then operationally defined in terms of the tension between two levels of processing: short-term variant and long-term invariant.

The aim of this paper is to explore aspects of short-term schema-driven perception in a tone center perception task. A distinction is made between passive and active schema-based recognition. Passive recognition assumes the use of a schema as a template, that is, a pattern with which the new information is correlated. Active recognition assumes an active role of the schema itself. Although the physiological and psychological foundations of active schema-driven perception are not well understood (Bregman 1990), a framework is presented of how stable structures of tone perception, once established, might operate in interpreting musical signals. The model is inspired by the metaphor of the brain as a complex dynamic system.

SIGNALS, IMAGES, SCHEMATA, AND MENTAL REPRESENTATIONS

In modelling perception and cognition it makes sense to distinguish between four types of "representational categories": signals, images, schemata and mental representations.

- A *signal* is the acoustical representation of musical information. In our model, digital signals are sampled at 20000 sa/s. Figure 1 shows an example of a digital signal that contains three frequencies: 600, 800 and 1000 Hz at SPL levels of 60, 55 and 50 dB (the maximum is 80 dB).

- An *image* or *perceptive map* represents the signal at the level of the auditory periphery. The most complete auditory image occurs at the level of the auditory nerve. At further stages, starting with the cochlear nucleus, auditory processing becomes differentiated and more specialized (Young et al. 1988).

  In the computer model, an image is represented by an ordered array of numbers (a vector). Each number represents the probability of neural firing during a short time interval. The rate at which images change may depend on

the type of image. There are different kinds of images conceivable: auditory nerve images, onset images, offset images, autocorrelation images, frequency transition images,...(Cfr. Brown 1992). In tone center recognition the images that are most relevant are the autocorrelation images and related images. In the model, the auditory nerve images are updated every 0.4 ms, the autocorrelation images only every 10 ms.

- *Schemata* or *cognitive maps* store categorical information and reflect the functional organization of the auditory cortex. Neurophysiological studies (e.g. Suga, 1988) provide evidence for the existence of different kinds of schemata. Schemata are assumed to have a longer life-time than images. They are typically used for invariance-detection. As suggested by computer simulations, the *schema* for tone center perception is probably learned. Figure 2 shows an example of a learned schema. The response patterns, which are invoked by specific input patterns, may themselves be considered as images. On this map, the tone center images are organized in the circle of fifths.

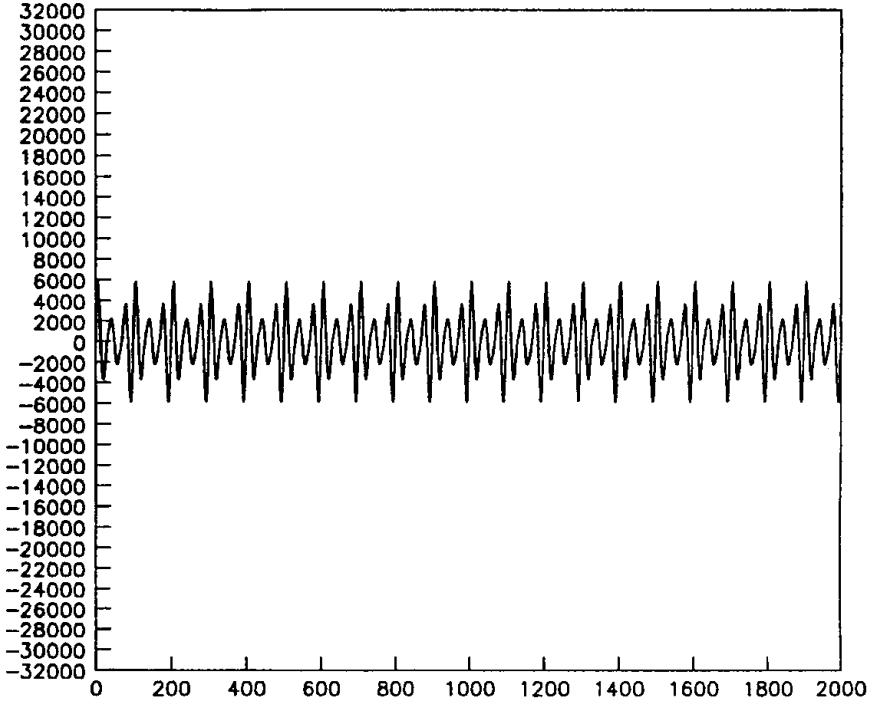

Fig. 1. A musical signal that contains three tone components: 600 Hz at 60 dB, 800 Hz at 55 dB and 1000 Hz at 50 dB. The representation of the intensity is linear (using a 16 bit resolution). The signal, which takes 100 ms, is sampled at 20000 sa/s.

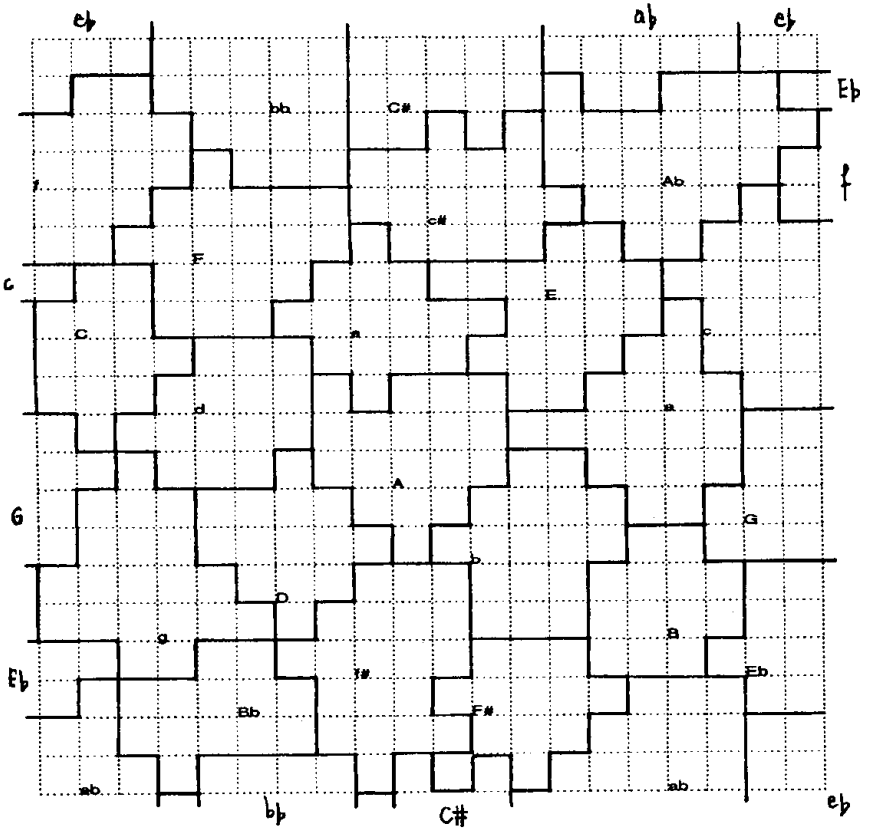

Fig. 2. Learned schema. The map is a two-dimensional array of artificial neurons (represented by boxes in a Kohouen-type network). In this example, the neurons of the upper side are connected to the neurons of the lower side, and those of the left side are connected to those of the right side, thus forming a torus structure. This schema has been trained with a short piece of music transposed in all keys, using the auditory model of Terhardt et al. (1982) (Leman 1991a). The schema is interpreted as a classifier: neurons which belong to the same tone center are grouped and labeled accordingly. The network appears to be highly ordered and the circle of fifth's can be clearly distinguished.

- *Mental representations* are knowledge structures that are supposed to be used in solving specific tasks. Mental representations are artifacts derived from psychological testing and they refer to a "mental" world, rather than a "physiological" or "brain" world. By using techniques of multi-dimensional scaling, it is possible to depict the data of the tests as mental spaces. Well-known examples are the mental spaces for phonemes, timbre (e.g. Grey 1978)

and tone center perception (Krumhansl 1990). Figure 3 shows the mental representation for tone center perception.

One of the aims of modelling perception and cognition is to show that these different representational categories are somehow related to each other. Signals are transformed into images, and images organize into schemata. By looking for correlations between the model and the data gathered by cognitive psychologists one may even try to relate the cognitive brain maps to the space of mental representations. In tone center perception, one of the aims is to show that the mental structures can be completely understood in terms of causal relations between signals, images and schemata. Contrary to the static world of mental representations, this approach offers a dynamic point of view by showing (a) how the cognitive map comes into existence, (b) how it is used in a music perception task. The computer simulations, by which the auditory strategies are implemented must rely on plausible neurophysiological grounds. They are useful to get a deeper understanding of music cognition.

A MODEL OF TONE CENTER PERCEPTION

The computer model is depicted in Figure 4. It consists of two modules: a perception module (the preprocessor) and a cognition module. The perception module is based on an auditory model. Its function is to extract the relevant pitch information that is present in the signal and prepare it for further use. The

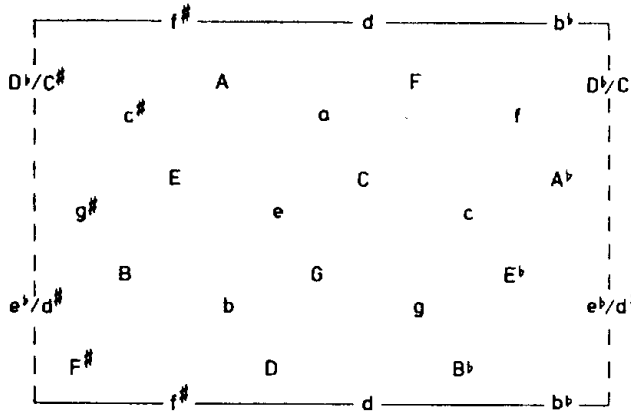

Fig. 3. Mental representation (Adapted from Krumhansl 1990). As in Figure 2, the upper and lower, and the left and right sides connect, thus forming a torus. The tone centers are organized in terms of the circle of fifth's.

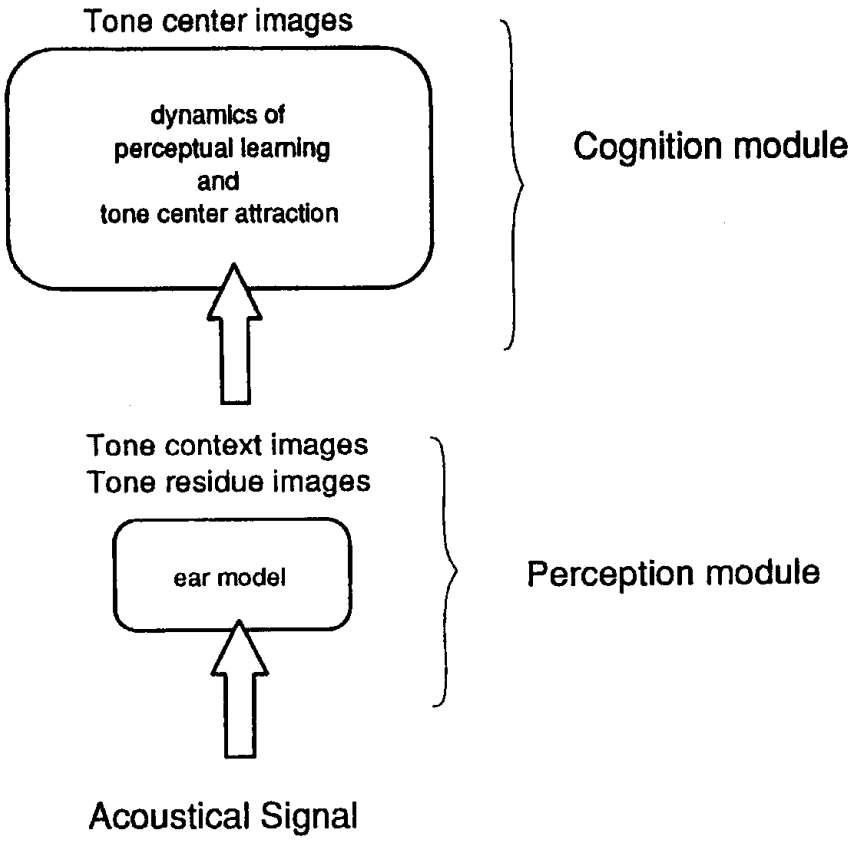

Fig. 4. The computer model. The model consists of a perception module and a cognition module. The perception module, based on an auditory model, transforms the acoustical signal into tone residue (completion) images and tone context images. The cognition module processes the tone context images into tone center images by categorization (dynamics of perceptual learning). The recognition and interpretation part is achieved by a tone center attraction mechanism.

cognition module comprises two types of dynamics. One type operates at long term and is supposed to be responsible for the organization of neuronal functions into a stable tone center perception schema. As a matter of fact, the outcome of a self-organizing learning process is highly dependent on similarities between the patterns delivered by the preprocessor. If the patterns are unrelated, there is little hope that some inherent structure can be found by learning. We may assume that the periphery "extracts" relevant information from the environment and "prepares" it for further use at the cognitive level. Trained with Western tonal music, the

response structure of the schema reflects learned relationships between tone centers (in terms of circles of fifths) and "tone center images" emerge in the schema. The other type of dynamics, called "tone center attraction", operates at short term and relates incoming images to the learned stable points of the schema.

In the past, psychoacoustical foundations of harmony have been based on auditory models of pitch perception, in particular, the perception of implied fundamentals (Terhardt et al. 1982; Parncutt 1989). The evidence is based on the fact that for a given tone, when the fundamental and some of the first harmonics are filtered out, we hear a tone with a slightly different timbre but with the same pitch. The pitch perceived is called the low pitch, the virtual pitch, or the residue pitch, and it corresponds with the fundamental that is missing in the filtered signal. The phenomenon is not just limited to performance under bad conditions (filtered signals), but works equally well under ideal listening conditions. As such it was argued that it could provide a psychoacoustic foundation of harmony (Terhardt 1974). Listening to a chord played by an acoustic piano will produce images that determine harmonic progressions. Contrary to the perception of implied fundamentals, however, the root of the chord is not necessarily heard but it plays an important role in music.

In the present model, it is assumed that completion patterns (not necessarily fusion!) provide sufficient information for tone center perception. The relevant pitch range is to be found in the residue region, that is, between about 50 and 500 Hz (Zwicker and Fastl 1990). The frequencies of this region convey the pitch information of the music at a particular point in time. This information is represented by a so-called "tone residue image" or "completion image".

The peripheral part of the model has been adopted from an auditory model developed by Van Immerseel and Martens (1992). It transduces the signal into a set of neural firing rate patterns which are assumed to represent the information conveyed by the auditory nerve. Related models have been described by van Noorden (1982), Assmann and Summerfield (1990), Meddis and Hewitt (1991), and others. Meddis and Hewitt give a detailed introduction to the pitch identification capabilities of such models, including the perception of the missing fundamental, ambiguous pitch and pitch shift. The completion images are based on a periodicity analysis of the neural firing information in the auditory nerves. In contrast to Terhardt's model, the subharmonic templates are not learned, but are supposed to have a physiological basis.

Auditory nerve images

The peripheral part takes into account the filtering of the outer and middle ear, and the hydro-mechanical bandpass filtering in the cochlea.[1] The latter is implemented as a bank of asymmetric bandpass filters at distances of one bark (one critical

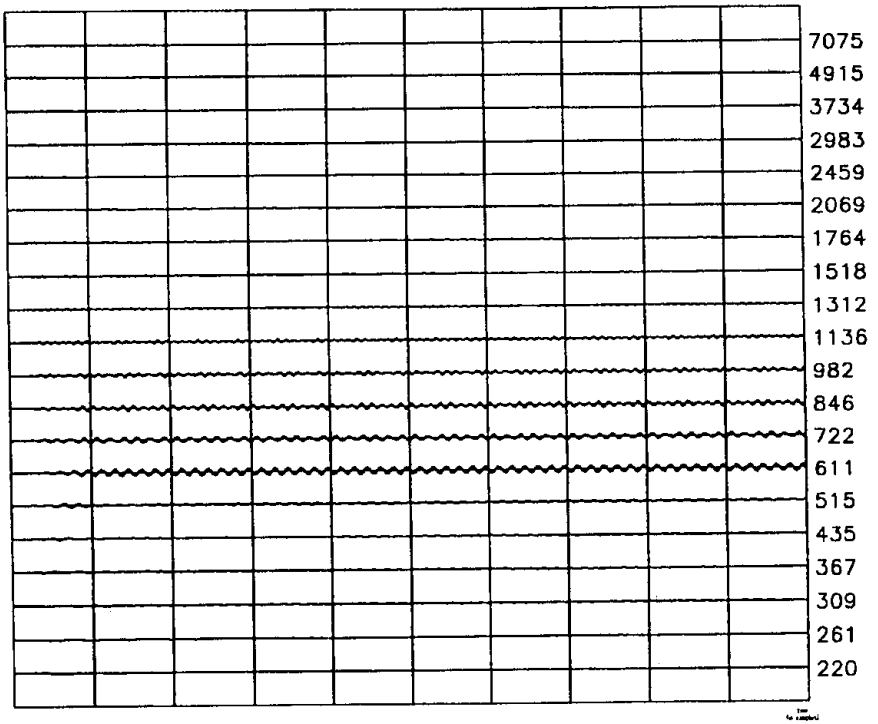

Fig. 5a. Bank of filters. The input is the signal shown in Figure 1. The vertical lines along the X-axis mark sections of 10 ms. The center frequencies (CFs) of the filters go from 220 Hz (lowest channel) to 7075 Hz (highest channel).

band). Twenty such filters are used in the range of 220 to 7075 Hz (center frequencies) (See Figure 5a). Further stages of the model perform a mechanical to neural transduction (Figure 5b). The following features are important:

- **Half-wave rectification:** due to the polarisation of the stereocilia of the inner hair cells, only the positive phase of the signal is captured by the hair cells.
- **Dynamic range compression:** intensity is coded both by the spike rate of the neurons (represented by the probability of firing during a defined time interval) and the activity at different channels. The design by Van Immerseel and Martens is based on a transition zone of maximum 50 dB and the representative values for spontaneous and saturated firing of the neurons are between 0.05 and 0.15 spikes/ms respectively.
- **Short-time adaptation:** there is an increased sensitivity after a period of nonstimulation, and a suppressed sensitivity after a period of strong stimulation.

- **Temporal information encoding:** because of their inability to synchronize with fast frequencies, neurons transfer amplitude modulations. In the original model this is implemented by a low-pass filtering (250 Hz) of the neural firing patterns, thereby extracting the envelopes of the neural firing pattern. This technique, inspired by the phase-locking capabilities of the auditory nerve patterns to the amplitude modulations in the signal, allows a considerable down-sampling of the original frequency to 500 sa/s. The aim was indeed to develop a real-time model for speech recognition. For musical data, however, 250 Hz might be too low for a fine periodicity analysis. Therefore, the model has been modified such that the neural firing patterns are filtered at 1250 Hz (sampling rate of 2500 sa/s) instead of the original 250 Hz. At higher frequencies (smaller distances between the peaks) the latencies of the synchronization stay in a larger range reflecting the fiber's refractory period. It is known that the limit of synchronization (or phase-locked discharges) of the auditory nerve fibers is about 4000—5000 Hz (Javel et al. 1988). The phase-locking to integral multiples of the period gets lost at about 1100 Hz (Gelfand 1981). The current limit is an approximation to the latter range.

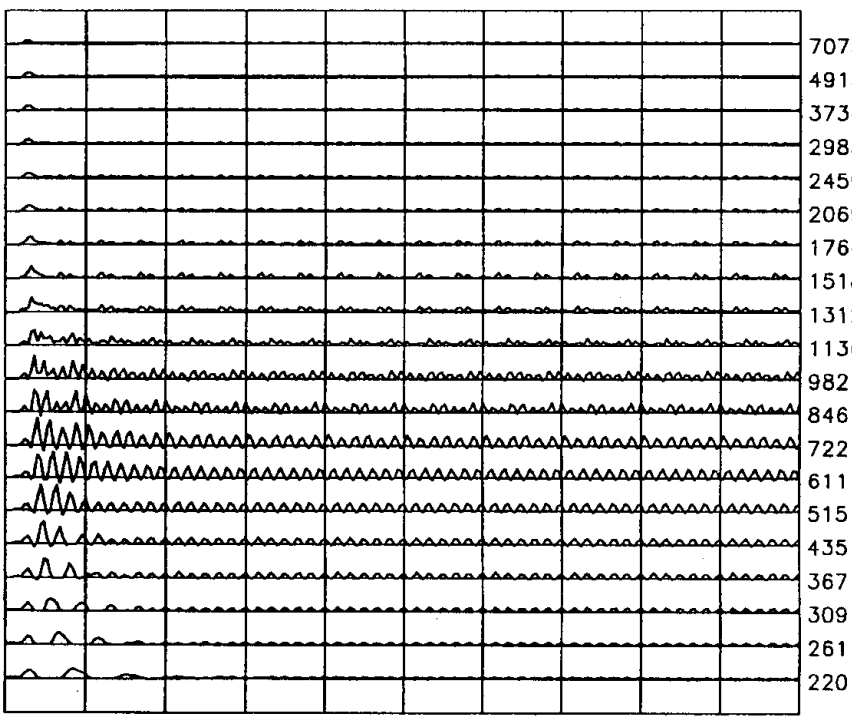

Fig. 5b. Auditory nerve patterns. This figure shows the effect of the neural transduction of the filtered signal of Figure 5a.

- **Amplitude modulation:** the neural firing patterns can be considered amplitude modulated signals. Also, if two frequencies fall within one critical band, their spectral components are not resolved but they generate a modulated signal in the corresponding auditory channel. See for example channel 9 (CF = 846 Hz) in Figure 5b. Similarly, since filters overlap, there can be an influence from one channel on the other. Pitch extraction is based on these modulated signals.

The model operates at a sampling rate of 20000 sa/s but the extraction of envelope patterns (justified by the synchronization phenomenon) allows a down-sampling to 2500 sa/s. The neural firing patterns (Figure 5b) are updated every 0.4 ms and thus the output of the analytical part of the auditory model (the "auditory nerve image") is a vector of 20 elements, each element representing the probability of neural firing within that interval.

Tone completion images

A pattern-completion module has been build on top of this peripheral part. Its function is to transform auditory nerve images into tone completion images by a periodicity analysis of the neural firing patterns in each channel (filter). The periodicity analysis is implemented by a short-term autocorrelation function and for that reason the images are also called "autocorrelation images". To sharpen the peaks in these images, the firing values are clipped to the mean value of the analyzed frame. The autocorrelation is defined in Eq. 1.

$$R(n) = a(n)\sum_{k=1}^{K-n} s(k)s(k+n)w(k) \tag{1}$$

where $R(n)$ is the autocorrelation value at lag n (in the range from 1 to K), and $s(k)$ is the signal at k. $w(k)$ takes the form of a decaying exponential:

$$w(k) = \exp\left(\frac{-k}{T}\right) \tag{2}$$

where T is the time constant. The function $a(.)$ attenuates $R(.)$ according to a parabolic function:

$$a(n) = 1-\alpha(n-\frac{K}{2})^2 \tag{3}$$

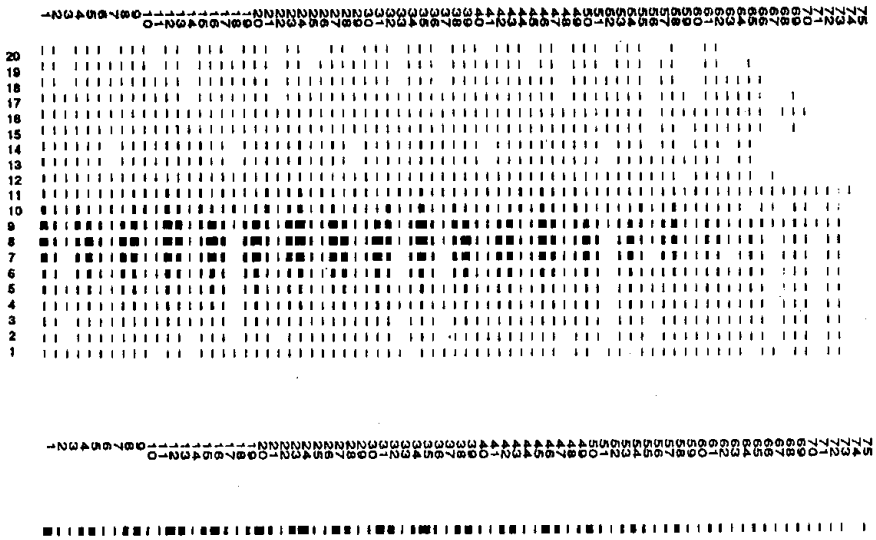

Fig. 6. Correlogram. The channels are marked by numbers from 1 (CF = 220 Hz) to 20 (CF = 7055 Hz). The time lag (0.4 ms) of the autocorrelation is indicated by the numbers 1 to 75. To find the frequencies, one should divide 2500 by the number corresponding to the time-lag. The lower part of the figure shows the sum of the autocorrelations over the channels.

In the examples that follow, K is 30 ms (75 samples), T is 16 ms (40 samples) and α is 0.5. The autocorrelation images are computed at intervals of 10 ms. The attenuator was introduced to decrease the role of those regions in the autocorrelation image that have either a too dense or too sparse resolution. For example, between 500 Hz and 250 Hz, there are but 6 units to represent the chromatic range, whereas from 250 Hz to 125 Hz, there are 11 such units. Since semitones are to be distinguished, one must be careful that the resolution is fine enough. (For other purposes, interpolation could be used). The present approach has solved slight deviations that were recurrent in the results of previous studies.

Figure 6 shows the autocorrelation functions of different channels in one single frame of 30 ms. The points on the abscissa correspond to the time-lags, from 1 to 75. The ordinate represents the channels, from low frequency (CF 220 = channel 1) to high frequency (CF 7075 = channel 20). The values of the autocorrelation are represented by the thickness of the lines. The values below 0.1% of the highest value in the figure are not represented. The lowest line shows the summary autocorrelation or completion image which, for this purpose, was normalized according to the highest value in the vector. The value at time-lag 12.5 (here represented by activity at point 12 and 13) corresponds to a frequency of 200 Hz

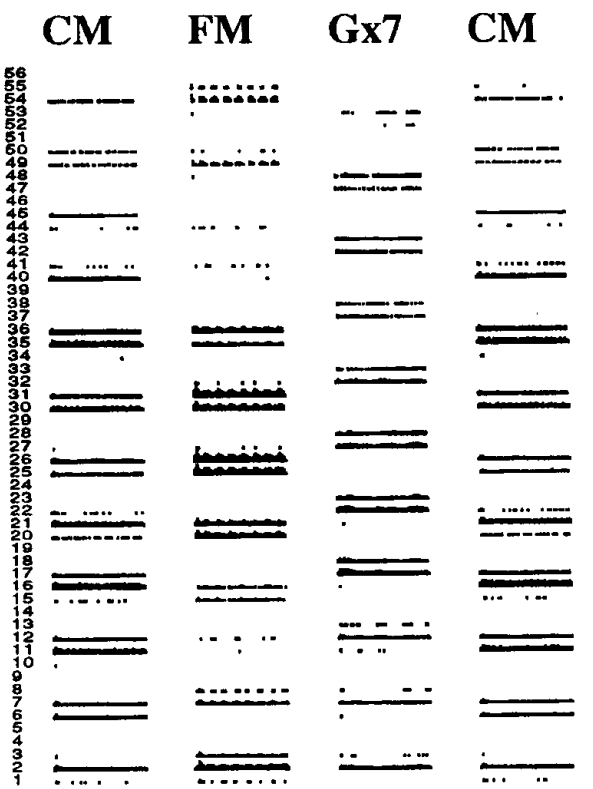

Fig. 7a. Autocorrelation images of the Shepard-tone chord sequence CM—FM—Gx7—CM at channel 6 (CF = 515 Hz). The X-axis is the time (points at intervals of 10 ms). The Y-as shows a selection of the autocorrelation, from 2 ms (=500 Hz) to 24 ms (41.6 Hz).

(= 2500/12.5). Without the attenuator, this would be the highest value — its frequency corresponds with the residue pitch of the tone complex. In tone center recognition, however, it is more important to consider the pattern as a whole rather than in terms of exact pitch.

Figure 7a, 7b, and Figure 8 show the autocorrelation analysis in a somewhat different way. The signal is a sequence of four chords: CM—FM—Gx7—CM. The tones that make up the chords are Shepard-tones. The spectrum consists of octave-components within a bell-shaped envelope which favors the region between 500 Hz and 1000 Hz. Each chord has a duration of 500 ms and has a short exponential onset and offset of 30 ms. Between the chords, there is a rest of 200 ms. Figure 7a shows the evolution of the autocorrelation function in channel 6 (CF = 515 Hz). The abscissa is the time with points at intervals of 10 ms. The ordinate shows the

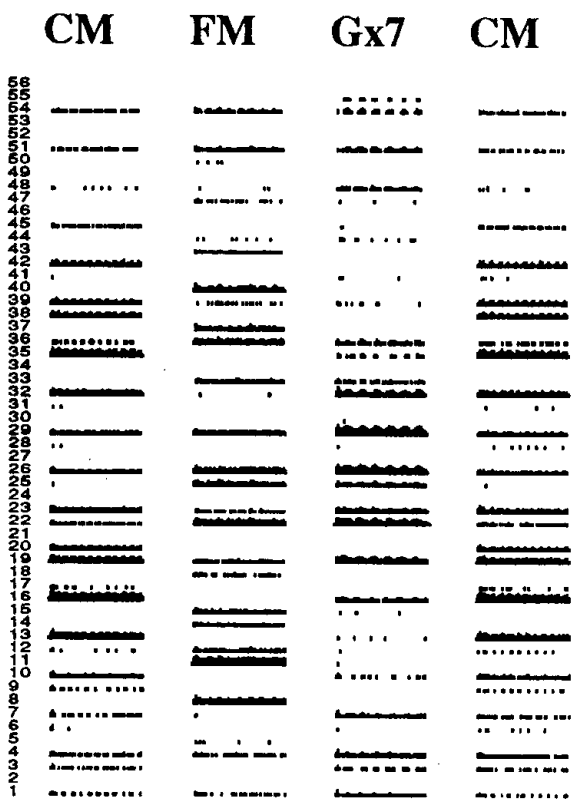

Fig. 7b. Autocorrelation images of the Shepard-tone chord sequence CM–FM–Gx7–CM at channel 9 (CF = 846 Hz). The X-axis is the time (points at intervals of 10 ms). The Y-as shows a selection of the autocorrelation, from 2 ms (=500 Hz) to 24 ms (41.6 Hz).

time-lags of the autocorrelation. The time-lags can be interpreted as an array of amplitude modulation detection neurons. The range corresponds to the residue pitch range of 500 Hz (= time-lag 1) to 41.66 Hz (= time-lag 56). To calculate the frequency at a particular time-lag i, one needs to divide 2500 by $i + 4$. The values below 20% of the highest value in this sequence (to which all values are normalized) are not represented. Figure 7b shows the evolution of the autocorrelation function in channel 9 (CF = 846 Hz). As in Figure 7a, the values are normalized to the highest value in this sequence.

It is assumed that the completion images result from an integration of the autocorrelation functions over all channels. This is shown in Figure 8. The most prominent tone of the first chord is at point 35, its frequency corresponds to: 2500/(35+4) = 64.1 Hz (= DO2).

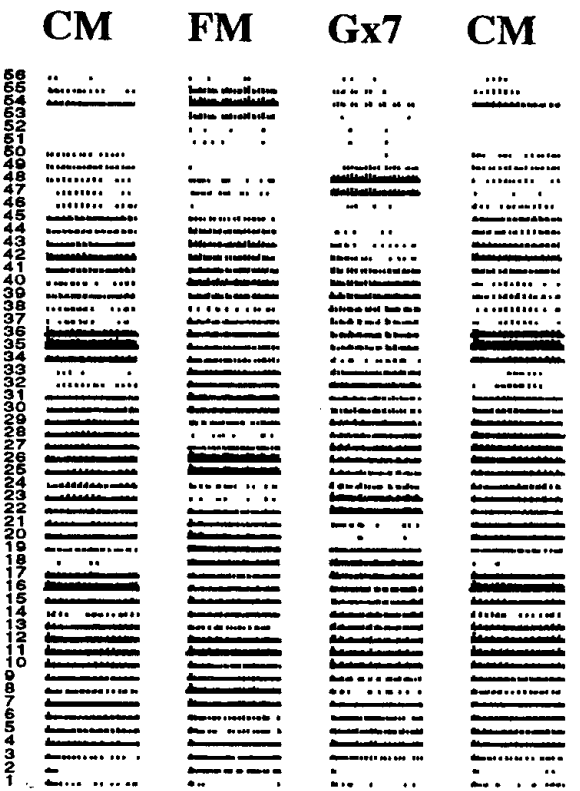

Fig. 8. Summary autocorrelation images (alternatively called: completion images or tone residue images) of the Shepard-tone chord sequence CM–FM–Gx7–CM.

Tone context images

To account for the temporal dependencies among the images, one further step is required. In music, chords are temporally related to each other and their order determines our tonal feeling (Brown 1988). There are indeed many reasons to believe that images "out of time" fail to provide sufficient grounds for explaining the complexities of harmony and tone center perception. There exists no perception "out of time" and consequently, the model should somehow account for these "large-scale" time dependencies.

The question is how we can keep track of the time-dependencies in the model. At present, the self-organizing map (Kohonen 1984), which we used as a paradigm for perceptual learning, does not consider data as ordered. One solution would be to modify the design of the neural network by allowing temporal integration in the

output of the neurons. Each neuron would display a temporal characteristic that is defined by its impulse response. This would make the neurons susceptible for temporal information (Naranjo 1993). More recently, the Kohonen network has been modified to allow the mapping of sequences of inputs without having to resort to external time delay mechanisms (Chappell and Taylor 1993). Such a modification, however plausible that might be from a biological point of view, has not been realised within the constraints of this study. The approach is restricted to a short time integration that is external to the network. As such, the network is kept as simple as possible and temporal dependencies are represented in the patterns themselves.

Time-dependencies between images can be accounted for by context values representing the recent history of the images. The basic requirements for such values have been described in (Leman 1992b). They are here summarized by means of Eq. 4.

$$v_i\ (t+1) = (1-\frac{1}{w})v_i(t) + A_i(t) \tag{4}$$

where i is a component of the vector v and A is the amplitude of the input signal. This equation resembles a Leaky integrator with a time window w. In the current simulations $w = 3$ seconds. The resulting images are called "context images". The context images based on Figure 8 are shown in Figure 9. As before, the values are normalized with respect to the highest value in the sequence and values below 20% of the highest value are not shown.

DATA-DRIVEN LEARNING

In recent years, neural networks have been used to obtain a better understanding of how schema emerge by self-organizing learning processes. The self-organizing map (Kohonen 1984) provides perhaps a initial model for such a recognition framework. It projects the similarities among images onto a map and produces (a) a two-dimensional topological organization of the tone context images (local as well as global), and (b) the emergence of responses to tone centers (Leman 1991a gives a detailed account). The network should then be extended by an attractor dynamics (or associative network) where the tone center patterns are used as stable points. Since the attractor dynamics is quite complex we have not followed this research path. At present, the attractor dynamics is separated from learning and the stable points are calculated by a short-hand method as an alternative to learning.

When one knows how to produce good representatives of class information, it

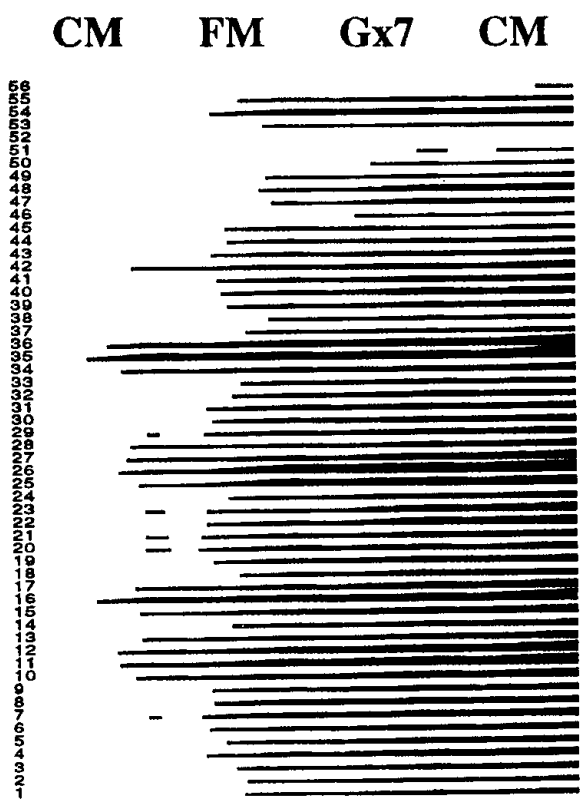

Fig. 9. Tone context images of the Shepard-tone chord sequence CM—FM—Gx7—CM using an integration time of 3 s. The values are normalized with respect to the highest value in the sequence.

is indeed not necessary to rely on neural networks. In music, we have text books and practical knowledge suggesting ways to obtain images for tone centers. Tone center images can be calculated by extracting context images from cadences. Cadences are known to produce fairly stable tone centers. In order to have an idea of their internal structure, one may then rely on simple calculations of the similarity relationships.

The procedure involves three cadences (based on the chord degrees I-IV-V7-I, I-II-V7-I and I-VI-V7-I) for each tonality, together with the major and minor scale and their transpositions over the chromatic scale. This gives a total of (3 * 2 * 12 =) 72 different cadences. The context image (See Figure 9) for each cadence is then calculated and only the last pattern (a 56-dimensional vector) is extracted. As a result, one obtains 72 vectors that represent context images — one for each cadence. This set is further reduced by taking the mean of the three types of

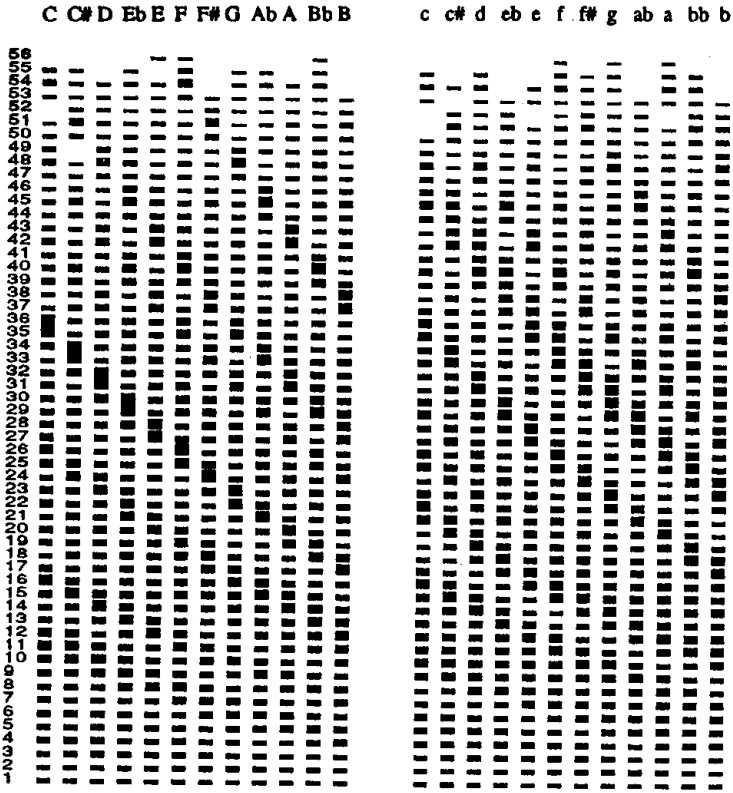

Fig. 10. Tone center images.

cadences for each tone center, which gives (72/3=) 24 tone center images. These are shown in Figure 10. The images are represented on the horizontal axis, while the vertical axis specifies the frequency content in terms of the time-lags.

The relationships between the patterns show a similarity of 0.98 and 0.96 with the relationships in the data of Krumhansl: see Figure 11a and 11b. Training a neural network with all patterns of 72 cadences (that is, 18360 56-dimensional patterns) would probably produce results that are but slightly different.

SCHEMA-DRIVEN PERCEPTION

The tone center attractor dynamics model (TCAD) describes recognition in terms of attractors, stable states and transitions towards stable states (Haken and Stadler, (eds) 1990; Amit 1989). The point of view suggests some interesting approaches

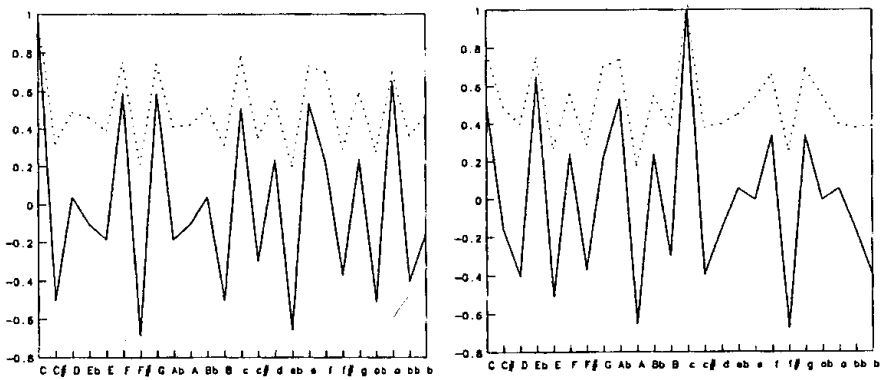

Fig. 11. The full curve shows the data of Krumhansl (1990). The dotted curve shows the correlation coefficients of the tone center images depicted in Figure 10. (a) The tone center of **C** is compared to all other tone centers; (b) the tone center of **c** is compared to all other tone centers.

to music perception and music analysis. But why do we need schema-driven perception?

First of all, in order to recognize tone centers, we need a schema to match the incoming images. This is the classical template-matching approach which is best conceived in terms of distributed representations. Tone center perception is often ambiguous in the sense that the perceived key is related to different tone centers at the same time. In Jazz music, for example, tone context patterns often have no pronounced tone center and part of the art is just to avoid the attraction of tone centers. The tone context images are localized somewhere in the middle of different tone centers, and the forces of attraction have a relatively weak influence on the position of the percept — which gives more freedom to the performer. Symbolic approaches often base the analysis on conceptual fixations in terms of key labels. One is in the key of C or F, but there is no specific information about the position in a space of tone centers, nor about the degree of similarity to one or the other key. Here, the tone center is specified with respect to all tone center images (templates) in terms correlation coefficients. In that sense, the classical approach is local and qualitative, while this one is distributed and quantitative. It is possible, however, to reduce the distributed account to the fixations of the linguistic-based paradigm by extracting at each time the tone center whose correlation is the highest over all tone centers[2].

A second answer to the question why we need a schema is more subtile, and much less easier to model. The idea is that a schema may actively determine the meaning of a particular incoming image on the basis of contextual information. In

that sense, we speak of interpretation rather than recognition. The role of an active schema is particularly relevant in cases where previous perceptual images are reconsidered in the light of new evidence. Consider a chord sequence with degrees IV—V—I. After hearing the first chord, the tone center will point to the tonic of the chord. It is only after hearing the rest of the sequence that the first chord can be reconsidered in function of new evidence. In this case the schema can be thought of as an active operator which is able to process contextual information in order to interpret the images of a near past.

It is possible to interpret such behavior in terms of an attractor dynamics. A state is an image at a certain point in time. Tone center images are interpreted as stable states that attract other less stable states — in particular the tone context images. The properties of attraction (strong or weak) may then depend on the distance between the unstable states and the stable states. When an unstable state is near a stable state, it will be attracted so that it will come closer to the stable state. This mechanism is believed to play a central role in meaning formation: the meaning of an object perceived is then measured by its distance to the attractors.

In what follows, passive and active schema-based perception are handled in more detail. The first is called recognition, the second interpretation.

Recognition

Recognition is associated with the determination of the position of a tone context state in a framework of tone center images. This is done by computing the distance of the tone context state to the tone center images. This yields a vector of correlation coefficients whose dimensionality is identical to the amount of tone centers, e.g. twenty-four.

This idea was applied to the musical signal of the Arabesque no. 1 (Cl. Debussy) played by Tamas Vasary (Deutsche Grammophon 429 517—2). The score excerpt is shown in Figure 12. The signal was obtained by digital sampling of the audio signal at 44.1 kHz. It was then downsampled to 20 kHz in order to fit with the sampling rate of the auditory model. The auditory nerve images, completion images and context images were computed with the auditory model using the same parameters as described above. Figure 13 shows the completion images at the beginning of the example. Figure 14 shows the tone context images (normalized to the highest value in the sequence). These patterns (normalized according to the Euclidian norm) are compared with the 24 patterns (also normalized according to the Euclidian norm) of Figure 10.

Figure 15 shows the result of the pattern-matching. The horizontal axis is the time. There are marks every 7.5 s. The vertical axis shows the correlation coefficients for each tone center. The black strips indicate the tone center which has the highest value at that moment. A important point to keep in mind is that the

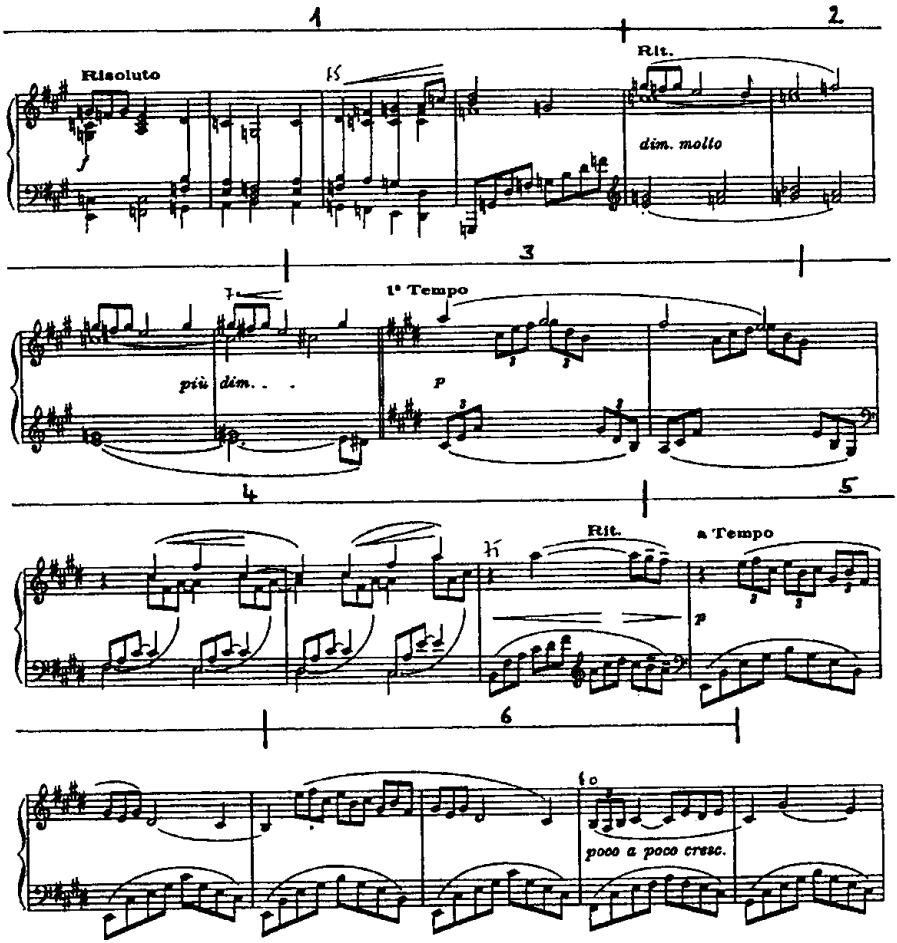

Fig. 12. Arabesque no. 1 by Claude Debussy, measures 62 to 81. The marks indicate sections of 7.5 s.

templates have been constructed with Shepard-tones, while this musical signal contains piano tones. Further, it may be interesting to remark that the analysis is not very different from an analysis of the same piece, based on the auditory model of E. Terhardt et al. (1982) (See Leman 1992a).

Globally speaking, the analysis suggests that the first half of the piece is in the tone center of C, while the second half is somewhere between c#, E and A, and then in f# and in E. The short black strips in the middle reflect the modulation.

A more detailed analysis of the first section (from 0 to 7.5 s) shows that in

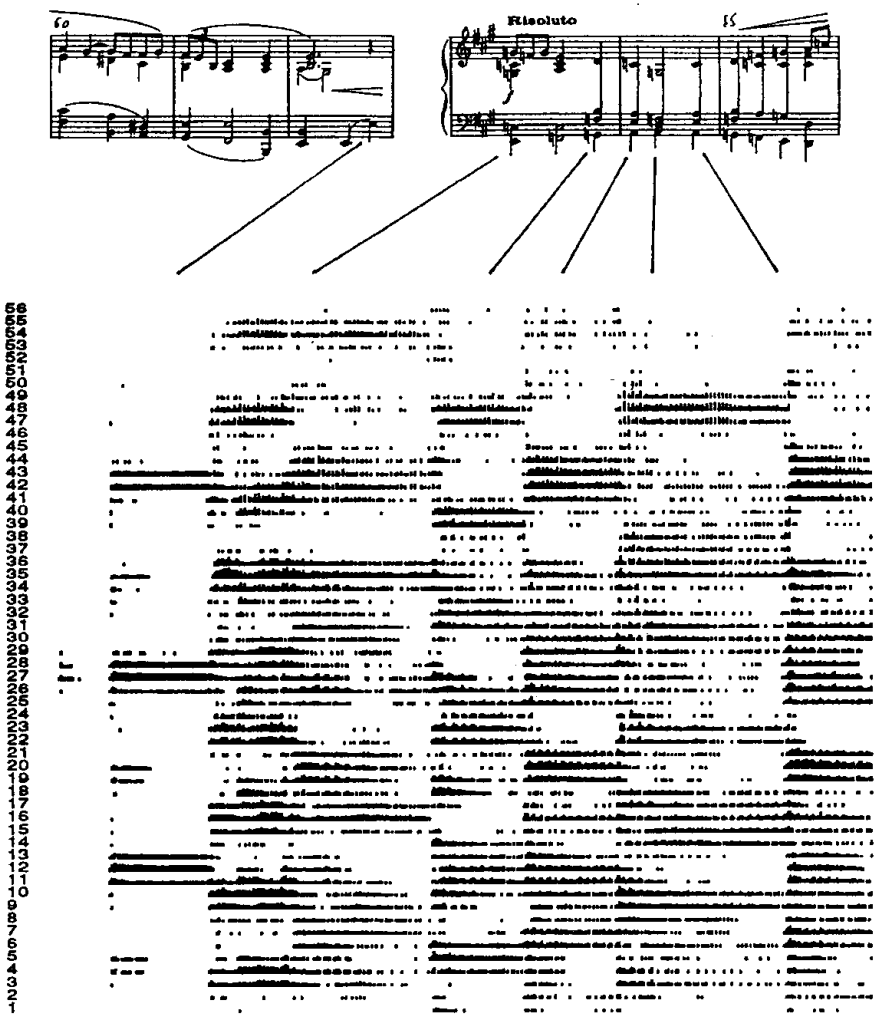

Fig. 13. Completion images of the first three measures of the Arabesque no. 1 in a recording by Tamas Vasary.

measure 65 the evidence for F is higher than for C. Most musicologists would probably argue that the first section is in C, without any modulation to F. The model is quite sensible to the occurrence of the chords AM, FM, CM, Dm7. In the second section (7.5 to 15 s) the black strip suggests the tone center of d for a very short period, but the main part of this section is in C. This continues to the third section (15 to 22.5 s) although there is a lot of movement in the tonal space. At the end of the first bar (which corresponds with in the middle of this section) there

Fig. 14. Tone context images of the first three measures of the Arabesque no. 1. The values are normalized with respect to the highest value in this sequence.

is a high evidence for c#. At the beginning of the Primo Tempo, there is no stable high value. In the fourth section there is evidence for E, A and f#. In the fifth section (30 to 37.5 s), there is a clear shift from f# to E at the beginning of measure 76 (a Tempo).

The sensitivity to changes in tone centers is due to the integration time, which is 3 s. If this integration time would be larger, then the judgments would be less susceptible to chord changes. On the other hand, we feel that a short integration time, and thus high sensitivity of the chords, reflects a common practice in Jazz performance.

Interpretation

The recognition part is limited to a registration of the position of the tone context in a framework of tone centers. It does not involve any active participation of the schema. Therefore, problems such as those related to the interpretation of the IV-

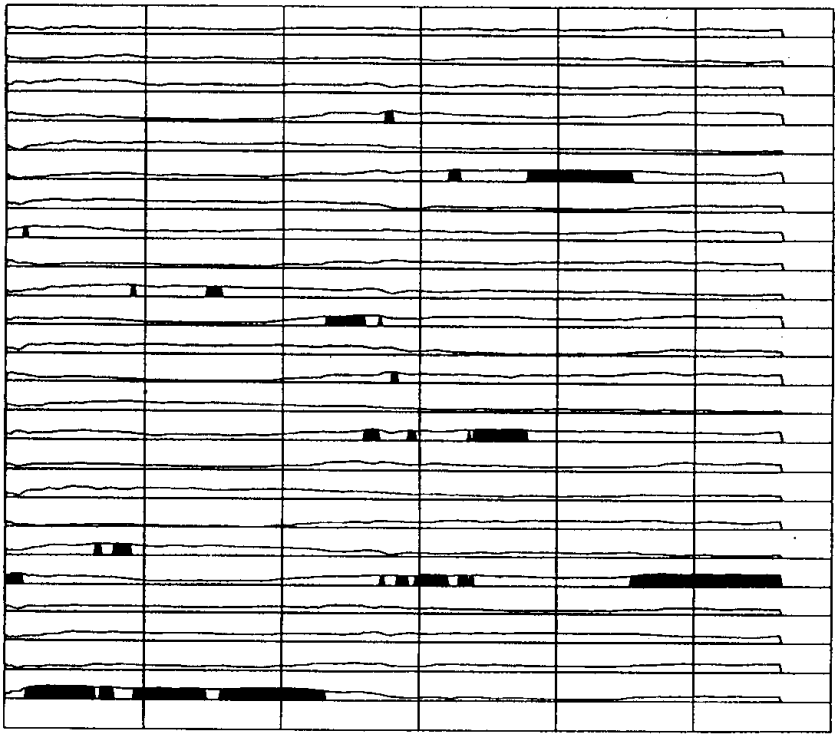

Fig. 15. Passive schema-based tone center analysis of Arabesque no. 1. The horizontal axis represents the time, with marks at intervals of 7.5 s. The vertical axis shows the correlation coefficients (between 1 and -1) of the tone context images with respect to all tone center images. The black strips mark the closest tone center.

V-I sequence, are not resolved. Also, in the beginning of the example, the MI is not recognized as a part of the tone center of C.

A first approach to TCAD was conceived as follows: When a tone center image is near to a particular tone context image, it exerts an attraction to it. The force of attraction depends on the distance between both images. When the distance is small, there is a great similarity between tone context and tone center, and the attraction will be great. When the distance is large, there is a little similarity between tone context and tone center, and the attraction will be weak. A strong attraction will have the effect that the context state will move towards the tone center to which it is most similar.

Although this approach seems appealing, the implementation does not produce good results. It is easy to see why: when a percept is close to a tone center image, called A, then the internal dynamics will force the interpretation state towards A

so that it will become more close to it. Without any data-driven information, the interpretation state would further move towards the attractor A. Such a dynamics produces two side-effects. The first is a sharpening of the percept by the attraction of the tone center so that the meaning of the percept will become more clear: attraction towards A implies that the perceived object is interpreted as A. This is desirable, because that is what can be expected from an interpretation. But the second effect, however, is a delay that is caused by the attraction. If the interpretation state is close to A, say, and the percept moves from A to B, then the interpretation will follow, but with a certain delay, because A will still exert force on the interpretation and keep it similar. This effect could be suppressed by a parameter that scales the force of both the interpretation and the percept. But this would be at the cost of the first effect and finally one would end with an interpretation path that is identical to the perception path. The conclusion is that a correct interpretation can never be found if the interpretation follows the time index of the percept.

The delay is similar to the well-known *hysteresis* effect (Stadler and Kruse 1990; Kelso 1990). Hysteresis occurs at phase-transitions of complex system behavior. In particular, the points at which the transitions occur are delayed by the forces of an attractor. This may be interpreted as a cohesive tendency to stable states. Hysteresis processes are probably compensated by an interpretation process in the sense that the interpretation of a percept at a certain moment in time involves a reconsideration of past interpretations — up to a certain time in the past — in the light of new information.

A more useful metaphor for TCAD is perhaps that of an *elastic snail-like* moving object. The head follows the time index of the music and the tail corresponds to a time-limited past. The position of each point P of the snail is given by an *interpretation state*. The dimensionality of this state is equal to the number of stable states so that this vector contains the distances of P to the tone center images and thus determines its position in the framework of stable points. The path followed by the head of our snail corresponds with the recognition process described in the previous paragraph. The tail, however, corresponds to a frame that keeps track of a reinterpreted past. Thus, it consists of a limited array of interpretation-vectors, one for each P-state in the memory frame.

In this implementation, the determination of the position of the head is important because the interpretation of the tail partly depends on it. Otherwise stated: the position of the tail is reconsidered in the light of the new position of the head. By changing from one attractor to the other, the interpretations of the past will change too. Competition is involved because the tail itself is susceptible to attractor-forces. So it may happen that a part of the tail remains near one attractor, while the head and the another part are near to a different attractor. That is what is meant by "elasticity": the snail might be influenced by the forces of different attractors.

Dynamics

In what follows, aspects of active schema-based tone center perception are simulated by a computational method. The process of adaptation depends on the following factors:

1. P(t,0), the tone context image (the head) at time t:
   The double time index refers to the absolute time t and an offset towards the time-limited past, which in this case is 0. The interpretation of the past may depend on the newly encountered tone context image $P(t,0)$ such that when $P(t,0)$ is close to an attractor T, then T will not only attract $P(t,0)$, but all other states $P(t,\tau)$ (τ=1,...,L-1) of the past as well, so that they become more similar to the attractor T.

2. The $P(t,0)$-attractors:
   When the similarity of a state $P(t,0)$ with any of the tone center images T^k (for k=1, ..., 24) is above a threshold h, then these T^k are considered attractors of $P(t,0)$. This set is labeled $A(t,0)$:

$$A(t,0)=\{T^k \mid cor[P(t,0),T^k]>h\} \\ =\{T^k \mid i_k \in I(t,0)>h\} \tag{5}$$

   where i_k is the k-th element of the interpretation vector $I(t,0)$. This vector contains the similarities measured as correlation coefficients cor (.) for all k.

3. $P(t,\tau)$, the tone context image at time $t-\tau$:
   The adaptation of a past state will also depend on its proper history. For example, when $P(t,\tau)$ is close to an attractor T — possibly a different one than the attractor of $P(t,0)$ — then T will attract the state $P(t,\tau)$. In other words, the proper past of $P(t,\tau)$ will have an influence on its new position in the state space as well. The past of the state $P(t,\tau)$ is per definition given by the state $P(t-1,\tau-1)$.

4. The $P(t,\tau)$-attractors:
   In general, each element $P(t,\tau)$ of the time-limited past has a interpretation-vector $I(t,\tau)$ in the state space which contains the information about the position with respect to the tone center images.The interpretation-vector $I(t,\tau)$ is associated with each $P(t,\tau)$ such that $A(t,\tau)$ is:

$$A(t,\tau) = \{T^k \mid cor[P(t,\tau),T^k] > h\}$$
$$= \{T^k \mid i_k \in I(t,\tau) > h\} \quad (6)$$

where i_k is the k-th element of the vector $I(t,\tau)$. When the similarity of $P(t,\tau)$ with any of the tone centers is above a certain threshold h, then these tone centers are considered attractors of $P(t,\tau)$. $P(t,\tau)$ is then adapted so that it becomes more similar to the $P(t,\tau)$-attractors.

5. The integrated past:
Instead of the proper history of each past state, one may also consider the attraction in terms of the integrated past. This set, called $A(t,\Sigma)$ is defined as:

$$A(t,\sum) = \left\{ T^k \mid cor\left[\frac{1}{L-1} \sum_{t=1}^{L-1} P(t,\tau), T^k \right] > h \right\} \quad (7)$$
$$= \{T^k \mid i_k \in I(t,\sum) > h\}$$

This equation accounts for the mean tone center attraction of the whole "past". It may be used as an alternative or additional constraint to the individual attractor-sets.

Another constraint could be a decreasing influence of $P(t,0)$ in function of increasing τ. At present, this constraint is not taken into account.

Obviously, the attractor sets $A(t,\tau)$, and $A(t,\Sigma)$ introduce a factor of competition. They prevent the interpretation from becoming too dependent on the current percept state $P(t,0)$ and its associated attractor set $A(t,0)$. The competition accounts for the "elasticity" but it has the property that when a tone center was easily recognized in the past, its interpretation will be more difficult to change, even in the light of new evidence. On the other hand, when an object was ambiguous in the past, its interpretation will be easier to change in the light of new evidence.

Given this background, it is possible to formulate the adaptation law. The adaptation of the ith element of a state $P(t,\tau)$ in the light of the new encountered state $P(t,0)$ is:

$$
\begin{aligned}
&P(t,\tau)_i = P(t-1,\tau-1)_i + \\
&\alpha * \sum_{k \in A(t,0)} \left[cor\left[P(t,0), T^k\right] * T^{k_i} \right] + \\
&\beta * \sum_{k \in A(t-1,\tau-1)} \left[cor\left[P(t-1,\tau-1), T^k\right] * T^{k_i} \right] + \\
&\gamma * \sum_{k \in A(t-1,\Sigma)} \left[cor\left[P(t-1,\tau-1), T^k\right] * T^{k_i} \right]
\end{aligned}
\tag{8}
$$

The summations run over all T^k that satisfy the above conditions for Eq. 5, Eq. 6, and Eq. 7, respectively. The parameters α,β and γ are scaling factors that define the rate of adaptation. The equation is applied to all elements i of the vector P.

Equation 8 says that the adaptation of element i of the vector associated with $P(t,\tau)$ is based on the previous value plus a change based on the competition of three attractors:

1. $P(t,0)$-attractors
2. $P(t-1,\tau-1)$-attractors
3. $P(t-1,\Sigma)$-attractors

The rate of adaptation depends on the correlation between P and those T that belong to the A-set. When the correlation is high, then the influence of T will be greater, otherwise it will be less great. Obviously, Eq. 8 applies only to those P-states that belong to the time-limited past. Therefore, the range of τ is from 1 to L-1. $P(t,0)$ by itself is not adapted.

To allow a better concurrence between the members of the A-sets, the values (correlation coefficients) have been normalized with respect to h.

Active schema-based recognition of tone centers

TCAD has been applied to the Arabesque no.1. One of the problems concerns the representation of the output. For each point in the running time, there is a frame of the past time, so that normally, the output should be conceived in four dimensions: (i) tone center, (ii) similarity to tone center, (iii) running time frame, (iv) and offset to time-limited past. The dynamics could be visualized with the aid of a movie but for practical reasons this is here reduced to an output at fixed intervals of 0.1 s. In particular, the states are shown just before they leave the memory buffer. This buffer (L) is taken to be three seconds.

Figure 16 shows the output of TCAD with the following parameter settings: $\alpha = 1$, $\beta = 0.5$, $\gamma = 0$, the threshold $h = 0.73$. The backward adaptation effect is seen clearly in the beginning. The single note MI is now mainly interpreted to be in C. The reference to the tone center of F, however, has remained, although the

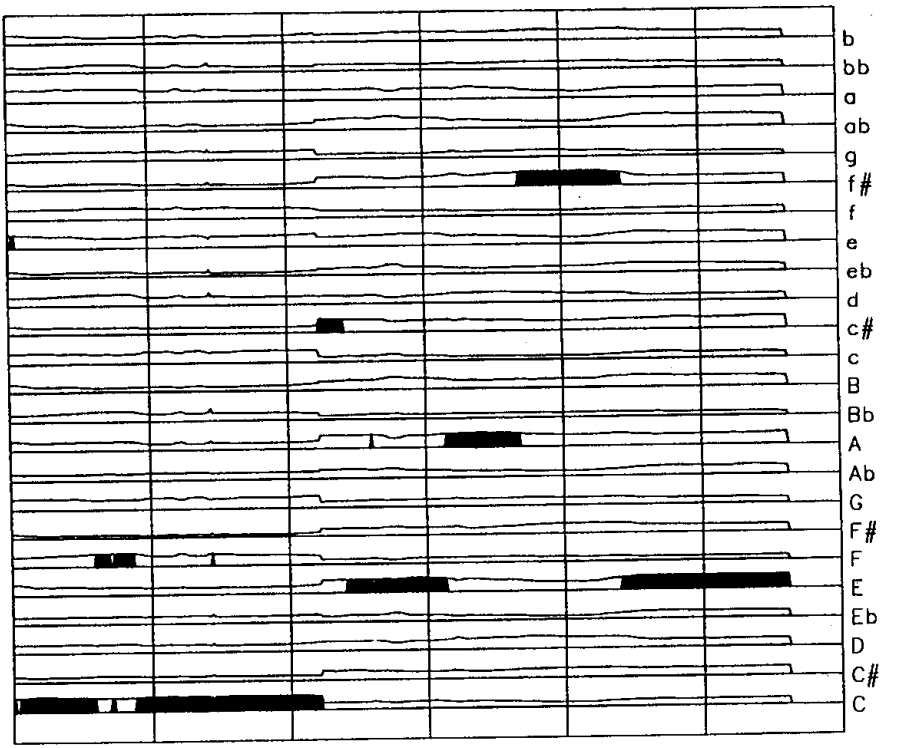

Fig. 16. Active schema-based tone center analysis of Arabesque no. 1. The horizontal axis represents the time, with marks at intervals of 7.5 s. The vertical axis shows the correlation coefficients (between 1 and -1) of the tone context images with respect to all tone center images. The black strips mark the closest tone center.

difference with C is very small. The difference in correlation coefficient is about 0.02. In the second section, the reference to d has gone. In the third section, there is now a clear demarcation between the first part (in C) and the second part. This effect is explained by the so-called elasticity of the snail. In addition, the hesitations in the second part of section 3 (Figure 15) make place for a pronounced decision in favour of E, although also A has a high value. Delay effects, which were due to the integration, are better accounted for and a demarcation (the jump from C to c#) is clearly visible. Sections four, five and six confirm the above observations.

The introduction of the integrator has thus far been less fruitful. Figure 17 shows the effect with $\gamma = 0.25$. Due to the integration, the reference to F in the first section has gone, but integration seems to have a negative effect of the

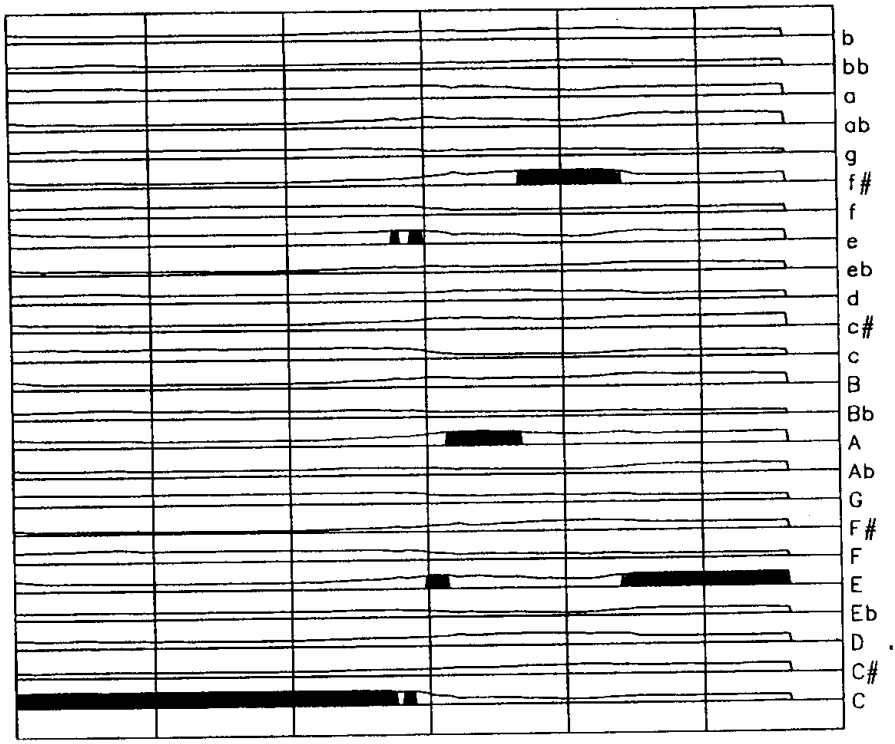

Fig. 17. Active schema-based tone center analysis of Arabesque no. 1. The horizontal axis represents the time, with marks at intervals of 7.5 s. The vertical axis shows the correlation coefficients (between 1 and -1) of the tone context images with respect to all tone center images. The black strips mark the closest tone center.

recognition of the modulation (section 3). Further experiments with parameters might point out that the influence of the integrated past should be less important.

NEUROPHYSIOLOGICAL FOUNDATIONS OF SCHEMA-BASED TONE PERCEPTION

The neurophysiological basis for schema-based tone perception rests on two grounds: evidence for a representation of tone completion images, and evidence for brain maps.

The neurophysiological evidence for tone completion

Schreiner and Langner (Schreiner and Langner 1988) have studied the temporal properties of neurons responding to envelope variations. They found that neurons in the IC (Inferior Colliculus) of the cat exhibit particular response characteristics to amplitude modulated tones. The properties of the neurons can be described by two functions: (a) The tuning curves, which define the filter characteristics in the frequency domain. The best frequency, that is the frequency to which the particular neuron is most sensitive, corresponds to the carrier frequency which in the model is equivalent to the center frequency of the auditory channel. (b) The temporal modulation transfer functions (tMTFs), which define the filter characteristics in the temporal domain. These functions are defined as the amount of neural activity evoked by amplitude modulated tones as a function of modulation frequency. Of particular relevance is the synchronicity of the neuronal activity to the envelope of the stimulus. The modulation frequency to which the neuron is most sensitive is called the best modulation frequency. Schreiner and Langner found evidence that the IC has layers (laminae) of neurons which are tuned to the same carrier frequency but which differ from each other with regard to the best modulation frequency. In fact, neurons tuned to the same best modulation frequency have a spatial concentric arrangement that is orthogonal to the spatial organization of the neurons that have identical center frequencies.

A bit simplified one could say that the autocorrelation analysis comes down to the transformation of temporal aspects of the signal into a spatial organization. For each carrier frequency, there is a set of neurons that is sensitive to different time-lags in the envelope (or amplitude) variations of the signal. A physiological model for such a neuronal correlation mechanism has been proposed by Langner (Langner 1983).

Neurophysiological evidence for a tone schema

The evidence for a spatial organization of the temporal aspects of the signal suggests further processing in terms of spatial arrangements.

The term "functional organization" of neurons means that neurons that belong to a certain area can have (or develop) a particular functionality or response characteristic to a given stimulus. There is evidence that neuronal functions of different nuclei in the auditory brain are ordered according to a specific frequency axis. These quasi *tonotopic maps* of neuronal functions (frequency filters) are logarithmically distributed and correspond to the place coding of frequency along the basilar membrane. Alternatively, this is called a *cochleotopic organization* of neurons (Schreiner and Merzenich 1988).

The organization of the cerebral cortex is such that neuronal cells with common

specificities are grouped together and are separated from cells with other specificities (Zeki 1993, p. 288). As Zeki shows for vision, connections in the cortex are commonly of the "like-with-like" type, one group of specific cells in one area connecting with their counterparts in another area. There is indeed evidence for a number of different types of maps, beyond the tonotopic or cochleotopic representation. Among different species, auditory maps have been found that are very specific for e.g., amplitopic representation, odotopic ("echo delay") representation, Doppler-shift (frequency-frequency) representation, representation of binaural data, space maps, amplitude modulation rate (Cf. Schreiner and Langner 1988) and other (Suga 1988). According to Suga, the size and topographic environment of these maps is an indicator of the importance of the parameters for the species. His work provides an abundance of evidence for the existence of cortical maps for auditory imaging in the mustached bat (the Pteromotus parnellii). The mustached bat emits complex biosonar signals and listens to echos for orienting itself and for hunting flying insects. These signals get localized somewhere on an internal map in the cortex. The map functions as a kind of resonance system in responding to the environmental stimuli. Signals acquire meaning because they are relevant for the action of the organism in the environment.

Maps that belong to a certain sub-modality (as Zeki shows for colour vision) may also differ with respect to the level of integration. Some maps are specialized in low level responses while others operate on a higher level. In auditory processing one may assume that the cochleotopic maps, which represent the cochlea much like the brain area V1 represents the retina, are low level maps. Maps, such as those for form or colour vision rely on these maps but are on a "higher level" in that they are responsive to more complex features of the signal. It is possible that similar "high level" brain maps may exist for tone center recognition. Both the preprocessor and the organizational principles of the present model are inspired by this plausible neurophysiological foundation, and the simulations give reasons to believe that the response structure to chords and tone centers is a working hypothesis. The self-organizing neural networks, which develop a functional organization by learning, indeed suggest that certain cells are responsive to general categories, such as chords and tone centers. Brain research is needed to verify the hypothesis.

Apart from this, it will remain quite difficult to find neurophysiological correlates of the attractor dynamics that is supposed to underly the active schema-driven recognition model. The neurophysiological foundation of the TCAD model is less evident, partly because the model is in a primary stage of research, partly also because the neurophysiological foundations of brain dynamics are not very well known.

APPLICATIONS

The present approach has two direct applications, one in music analysis, and another one in interactive music making.

An approach to psychoacoustic-based harmonic analysis

A TCAD-analysis outputs the position of a particular tone context image with reference to a frame of tone center images. Apart from its interest as a "fine-grained" tonal analysis, it is to be expected that this information can be helpful in defining functional roles of tones and chords.

If the integration window for context patterns is small enough (about 0.5 s), then it is possible to follow the individual chords in the piece in stead of the tone centers (Leman 1991a). The approach then assumes that chords themselves can function as stable perception points. The representation of chords on a map is indeed well-ordered and related to the tone centers (Leman 1991a, 1991b, 1992b).

With two time levels (one for chords and one for tone centers) it thus becomes possible to compare objects of two types and to reason about them. This path is not explored here, but the application is straightforward. Chord images can be related to tone context images (possibly larger integration windows could be used then) and the relations can be expressed in symbolic code.

This application is currently investigated in the framework of HARP (Hybrid Action Representation and Planning) (Camurri, Canepa, Frixione and Zaccaria 1992; Camurri, Innocenti, Massucco and Zaccaria 1992). HARP is a hybrid model that combines aspects of symbolic and subsymbolic processing. A key feature is that an object at the analogical level can be "hooked" to a corresponding concept in a semantic network where it is further processed by rules or domain-specific "experts". For example, a module that looks for chords will try to add concepts for chords to the symbolic engine. Chord recognition can be done on the basis of activation bubbles in the Kohonen map. Parallel to this, there is a module that tries to match the tone context patterns (integrated over a period of 3 s) with tone center images. The resulting information is also added to the symbolic engine. With the help of domain-specific experts it is then possible to "reason" about these objects, for example, to infer the functionality of the chords from their relationship to the tone centers.

Tone center information in interactive systems

In interactive music environments — that is: systems whose real-time sound generation is based on a sound environment that includes its own acoustical signals — the model of tone center recognition might provide information for the

navigation in the tone center space. The determination of the position of the music in the space of the tone centers can be done in real-time by using the template-method. The output is sufficiently accurate to allow tonal sensitive responses to what is heard.

The realization of such a module with the current technology will mainly depend on an efficient implementation of the analytical part of the auditory model and an optimization of the periodicity analysis.

CONCLUSION

A framework for schema-based tone center recognition and interpretation has been presented. Passive schema-based tone center recognition was associated with the localization of short-term auditory images within a space of tone center images. Active schema-based tone center recognition was associated with an adaptive dynamics of the position of the percept within this schema. The model assumes that interpretation involves the reconsideration of past interpretations in view of new contextual evidences.

ACKNOWLEDGEMENTS

We wish to thank H. Sabbe, the Belgian National Science Foundation (F.K.F.O) and the University of Ghent (Onderzoeksraad) for support. We are also grateful to J.-P. Martens and L. Van Immerseel for their help on the auditory model and to A. Camurri and N. Cufaro Petroni for discussion. The scientific responsibility is assumed by the author.

NOTES

1. For a detailed mathematical description of the model we refer to Van Immerseel and Martens (1992). We limit ourselves here to those aspects that are relevant for the cognitive model.
2. One should be careful, however, in comparing the "classical" musicological approach with the one presented here. The notion of tone context image has a definition based on auditory principles while the traditional notions of "key" and "tonality" have music theoretical foundations. It will become clear in the sequel that tone center recognition points to a fine-grained tonal analysis, without explicit "reasoning" in terms of harmonic functions or "tonalities".

REFERENCES

Amit, D.J. (1989). *Modeling Brain Function: The World of Attractor Neural Networks.* Cambridge, MA: Cambridge University Press.

Assmann, P. and Summerfield, Q. (1990). Modeling the perception of concurrent vowels: vowels with different fundamental frequencies. *The Journal of the Acoustical Society of America*, 88, 680–697.

Bregman, A.S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound.* London: The MIT Press.

Brown, G. (1992). *Computational Auditory Scene Analysis.* TR CS-92-22. University of Sheffield: Department of Computing Science.

Brown, H. (1988). The interplay of set content and temporal context in a functional theory of tonality perception. *Music Perception*, 5(3), 219–249.

Camurri, A., Canepa, C., Frixione, M. and Zaccaria, R. (1992).HARP: A framework and a system for intelligent composer's assistance. In: D. Baggi (ed.), *Readings in computer generated music.* Los Almitos, CA: IEEE Computer Society Press.

Camurri, A., Innocenti, C., Massucco, C. and Zaccaria, R. (1992). A software architecture for sound and music processing, *Microprocessing and Microprogramming (Proc. EUROMICRO-92, Paris)*, 35 (1–5).

Chappell, G.J. and Taylor, J.G. (1993). The temporal Kohonen map. *Neural Networks*, 6 (3), 441–445.

Gelfand, S.A. (1981). *Hearing: An Introduction to Psychological and Physiological Acoustics.* New York: Marcel Dekker.

Grey, J.M. (1978). Timbre discrimination in musical patterns. *The Journal of the Acoustical Society of America*, 64, 467–472.

Haken, H. and Stadler, M. (eds) (1990). *Synergetics of Cognition.* Berlin: Springer-Verlag.

Javel, E., McGee, J., Horst, J.W. and Farley, G.R. (1988). Temporal mechanisms in auditory stimulus coding. In: G.M. Edelman, W. Gall and W. Cowan (eds.), *Auditory Function: Neurobiological Bases of Hearing.* New York: John Wiley and Sons.

Kelso, J.A.S. (1990). Phase transitions: foundations of behavior. In: H. Haken and M. Stadler (eds.), *Synergetics of Cognition.* Berlin: Springer-Verlag.

Kohonen, T. (1984). *Self-organization and Associative Memory.* Berlin: Springer-Verlag.

Krumhansl, C.L. (1990). *Cognitive Foundations of Musical Pitch.* New York: Oxford University Press.

Langner, G. (1983). Neuronal mechanisms for a periodicity analysis in the time domain. In: R. Klinke and R. Hartmann (eds.), *Hearing – Physiological Bases and Psychophysics.* Berlin: Springer.

Leman, M. (1989). Symbolic and subsymbolic information processing in models of musical communication and cognition, *Interface – Journal of New Music Research*, 18(1–2), 141–160.

Leman, M. (1990). Emergent properties of tonality functions by self-organization, *Interface – Journal of New Music Research*, 19(2–3), 85–106.

Leman, M. (1991a). Een model van toonsemantiek: naar een theorie en discipline van de muzikale verbeelding. Doctoral Dissertation. Gent: University of Ghent.

Leman, M. (1991b). The ontogenesis of tonal semantics: results of a computer study. In: P. Todd and G. Loy (eds.), *Music and Connectionism.* Cambridge, MA: The MIT Press.

Leman, M. (1992a). The theory of tone semantics: concept, foundation and application. *Minds and Machines*, (2), 345–363.

Leman, M. (1992b). Tone context by pattern-integration over time. In: D. Baggi (ed.), *Readings in Computer Generated Music.* Los Almitos, CA: IEEE Computer Society Press.

Meddis, R. and Hewitt, M.J. (1991). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *The Journal of the Acoustical Society of America*, 89(6), 2866–2894.

Naranjo, M. (1993). *Apprentissage et Reconnaissance de Sequences Musicales.* Proceedings of the 13th International Congress on Cybernetics, 1992. Namur: International Association for Cybernetics.

Parncutt, R. (1989). *Harmony: A Psychoacoustical Approach.* Berlin: Springer-Verlag.

Schreiner, C.E. and Langner, G. (1988). Coding of temporal patterns in the central auditory nervous system. In: G.M. Edelman, W. Gall and W. Cowan (eds.), *Auditory Function:*

Neurobiological Bases of Hearing. New York: John Wiley and Sons.
Schreiner, C.E. and Merzenich, M.M. (1988). Elements of signal coding in the auditory nervous system. In: W. von Seelen, G. Shaw and U.M. Leinhos (eds.), *Organization of Neural Networks: Structures and Models*. Weinheim: VCH Verlagsgesellschaft.
Stadler, M. and Kruse, P. (1990). The self-organization perspective in cognition research: historical remarks and new experimental approaches. In: H. Haken and M. Stadler (eds.), *Synergetics of Cognition*. Berlin: Springer-Verlag.
Suga, N. (1988). Auditory neuroethology and speech processing: complex-sound processing by combination-sensitive neurons. In: G.M. Edelman, W. Gall and W. Cowan (eds.), *Auditory Function: Neurobiological Bases of Hearing*. New York: John Wiley and Sons.
Terhardt, E. (1974). Pitch, consonance and harmony. *The Journal of the Acoustical Society of America*, 55(5), 1061–1069.
Terhardt, E., Stoll, G. and Seewann, M. (1982). Algorithm for extraction of pitch and pitch salience from complex tonal signals. *The Journal of the Acoustical Society of America*, 71(3), 679–688.
Van Immerseel, L. and Martens, J.-P. (1992). Pitch and voiced/unvoiced determination with an auditory model. *The Journal of the Acoustical Society of America*, 91(6), 3511–3526.
Van Noorden, L. (1982). Two channel pitch perception. In: M. Clynes (ed.), *Music, Mind and Brain: The Neuropsychology of Music*. London: Plenum Press.
Young, E.D., Shofner, W.P., White, J.A., Robert, J.-M. and Voigt, H.F. (1988). Response properties of cochlear nucleus neurons in relationships to physiological mechanisms. In: G.M. Edelman, W. Gall and W. Cowan (eds), *Auditory Function: Neurobiological Bases of Hearing*. New York: John Wiley and Sons.
Zeki, S. (1993). *A vision of the brain*. Oxford: Blackwell Scientific Publ.
Zwicker, E. and Fastl, H. (1990). *Psychoacoustics: Facts and Models*. Berlin: Springer-Verlag.

Marc Leman
University of Ghent
Institute for Psychoacoustics and Electronic Music
Blandijnberg 2
B-9000 Ghent
Belgium

Tel: +32-9-2644125
Email: marc.leman@rug.ac.be
Fax: +32-9-2644196

Marc Leman holds a doctoral degree from the University of Ghent and is doing research at the Institute for Psychoacoustics and Electronic Music.
His research activities focus on the epistemological and methodological foundations of a theory and discipline of musical imagination based on physiological acoustics (psychophysics) and gestalt theory (self-organisation theory). Since 1987 he has been active in (co-)organizing several workshops and conferences on Cognitive Musicology and Artificial Intelligence applied to Music. Marc Leman is editor of Journal of New Music Research (formerly Interface).