Chapter 8

General Discussion

The applicability of the proposed model to two different problems, voice separation and tempo tracking, supports the validity of the hypothesis about the existence of correlative perception in humans. Moreover, the existence of correlative perception is indirectly substantiated by the consistency of the proposed model with music theory. So many coincidences are very unlikely by chance alone.

It is reasonable too that the perception reduces the redundancy of the input information in order to achieve a compact representation. Most likely, the brain tends to save the memory store and facilitates access to accumulated knowledge by representing information in an aggregate form.

As mentioned in Chapter 1 and Section 7.2, musical signal is a carrier of some semantic information. The task of music recognition is therefore extracting this information which can be realized by data aggregation. This information falls into several categories. A listener perceives structurally organized sound, fine execution nuances, emotions of the performer, and acoustic characteristics of instruments and environment. For example, performer's emotions (joy, anger, etc.) are transmitted by so-called essentic curves in loudness and tempo changes (Clynes 1977; 1983). A remarkable peculiarity of timbre to represent physical phenomena (force, tension, etc.) is described by Cadoz (1991). We can say that most semantic musical information relates to different forms of causality in sound.

Our study is devoted to the recognition of special cases of causality, structural causality. We adduce arguments in favor of the fact that an appropriate representation contributes to the recognition of audio structure by identity and by similarity, both in pitch and time domains.

In particular, we deal with the decomposition of chords into tones. From a physical standpoint, all sounds are decomposable into pure tones which correspond to resonances of vibrating bodies. Consequently, from a physical point of view a chord cannot be regarded as more complex than a monolithic complex tone. Indeed, both can be produced by a single vibrating body (loudspeaker, piano board), whereas tones are perceived as entireties, and chords are perceived as compounds. On the other hand, sounds from several physical sources can fuse into one, as in case of voice synthesis in pipe-organ and symphony orchestra (see Sections 1.4 and 7.5).

Therefore, the difference between monolithic sounds and sound complexes relates to psychology, information theory, and computer science rather than to physical acoustics. In fact, our goal is to find groups of audio data whose complexity is intermediate between that of sinusoidal tones and spectra of complex sounds, i.e. we deal with representations of data.

In our study a sound is said to be compound if its spectrum can be structured into independent groups of partials. The grouping is performed with respect to the similarity of groups and with respect to the criterion of least complex total representation. In dynamics this representation links similar spectra into acoustical trajectories, contributing to segregation of acoustical processes (polyphonic voices). In statics we find constituents of sound complexes (notes). We prove that optimal representation of a tone reveals its monolithic nature but optimal representation of a chord does reveal its compound structure.

We see that optimal data representation reconstructs physical causality in sound generation, revealing several excitation sources which result in a sensation of a chord. It is remarkable that the two different matters, physical causality and optimality in data representation, correspond to each other.

We can adduce some general arguments in favor of this correspondence. Since most physical processes evolve continuously, their successive states are not very much different from each other. Usually, these successive states have certain trends which are determined by some causes. On the other hand, these causes are usually not numerous, implying the effects to be not numerous too, so that the trends cannot mix chaotically. Therefore, the classification of data with respect to different trends results in the data representation where the segregated trends correspond to certain causes. Since the reaction of a physical system to an excitation is in a sense "optimal", the corresponding description should be "optimal" too. Therefore, finding the optimal description of a process is a way to recognize causal relations.

The next question is how the trends mentioned can be recognized. If we assume the continuity of physical processes, the trends can be revealed as corresponding to minor changes in the successive states of the phenomenon. To detect these minor changes, correlation analysis of slightly distorted data together with the method of variable resolution are quite suitable.

As mentioned in Section 2.4, both correlation analysis and method of variable resolution are realizable on neuron nets with parallel computing. This is consistent with the hypothesis that similar functions can be performed by the brain (Rossing 1990, p. 164), indirectly justifying our approach to perception modeling.

Thus we formulate the following hypothesis.

Data representation in terms of generative elements and their transformations reveals certain aspects of physical causality in sound generation. Such a representation is inherent in human perception, enabling source separation and tracking simultaneous audio processes.

The capacity to separate sounds and track simultaneous audio processes is extremely important for orientation in the environment and for semantic organization of information. From our point of view, the importance of these tasks explains the predominance of related audio mechanisms in audio perception (predominance of relative hearing over absolute hearing).

We suppose that our approach to perception modeling is quite general. In fact, we have already implemented models of chord recognition and tempo tracking, and we expect that the same model can be adapted to some other purposes.

In particular, the model may be applied to speech recognition. To recognize phonemes it is often proposed to recognize the contribution of different parts of the voice tract to the resulting speech signal. Since the parts of the voice tract has their own acoustical characteristics, we can pose the problem as a separation of "polyphonic lines" in the speech signal, where each line is associated with a certain part of the voice tract. Therefore, to recognize a phoneme, one needs to recognize the "chord" produced by the activated parts of the voice tract. Our model for polyphony tracking seems to be adaptable for that purpose.

Our approach can be extended from unidimensional data arrays (as audio spectra or sequences of time events) to two-dimensional data arrays with possible applications to image processing. It can be used for object separation in dynamics (by analogy with voice separation) and contour recognition (by analogy with chord recognition) in visual scene analysis.

For example, to segregate a moving object in dynamics, one can perform correlation analysis of successive instant images. In order to find both the object and its trajectory, it is necessary to find the deformations of the images which provide their high correlation. This can be done by the method of variable resolution (see Sections 2.3 and 2.4).

Further applications concern modeling of abstract thinking; see Giunchiglia & Walsh (1992) for a review. Note that our approach is based on revealing stable relationships between data blocks. Regarding stable relationships as new data, one can reveal stable relationships between stable relationships, etc. This way we come to abstract concepts which correspond to stable invariants of data representation.

Note that meaning can be understood as identifiable associations between memory patterns. Then an aggregate form of data which reveals stable relationships is the first step towards semantic organization of information.

A multi-level generalization of the model is imaginable where every next level of patterns is formed by stable relationships between patterns of lower level. To recognize meaning, the proposed model should be provided with associations between perceptual patterns (configurations of different levels) and memory patterns acquired in a previous experience. By interfacing such a model to a data base and extending to this base the methods for discovering correlations, a cognitive model can be obtained. In this model, abstract concepts are formed from associations between the patterns of high and higher levels. In other words, semantic analysis is understood as information analysis based on data aggregation and discovering the correlations of the aggregates.

In dealing with interactions of the model with a data base, the most promising solution is to use them jointly. Memory patterns (configurations of the data base) may compete with perceptual patterns (configurations of the artificial perception model), as when current patterns are compared immediately with those from a previous experience.

In this case the optimal representation of current data (in artificial perception model) may not be optimal in the joint model, i.e. the interpretation of a current message can be simpler in terms of memory patterns. Obviously, the set of memory patterns influences the system performance; it is to be expected that the different responses of different individual humans are conditioned by different experience, learning histories, or expectancies.

This means that the model of correlative perception complements the methodology of artificial intelligence with a stage of *"artificial perception"* which operates at the data input. If used in a proper artificial intelligence domain, the present model can achieve self-organization of knowledge. In pattern recognition, the model can be used to separate patterns, thereby making their identification easier. Therefore, artificial perception can interact with artificial intelligence in various modes. This corresponds to the interaction of human perception and human intelligence which complement each other and influence reciprocally.

In conclusion we point out that we may exaggerate the role of correlative perception, which operates side-by-side with many other perception mechanisms. The proposed mathematical model of correlative perception which is based on correlation analysis should not be regarded as universal or complete. To us, it is the idea of representing data optimally in terms of generative elements and their transformations which seems rather general.

Finally, we point out that neither the experiments, nor the theoretical rationale which have been put forth here should be seen as final ends. Rather this work is less a summary of theoretical applications and the obtained results than it is a posing of new problems.