ARTIST: ADAPTIVE RESONANCE THEORY TO INTERNALIZE THE STRUCTURE OF TONALITY

(a neural net listening to music)

by

FRÉDÉRIC GEORGES PAUL PIAT, M.S.

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN HUMAN DEVELOPMENT

AND COMMUNICATION SCIENCES

THE UNIVERSITY OF TEXAS AT DALLAS

August, 1999

PREFACE

This dissertation was produced in accordance with guidelines which permit the inclusion as part of the dissertation the text of an original paper, or papers, submitted for publication. The dissertation must still conform to all other requirements explained in the "Guide for the Preparation of Master's Theses, Doctoral Dissertations, and Doctor of Chemistry Practica Reports at The University of Texas at Dallas." It must include a comprehensive abstract, a full introduction and literature review, and a final overall conclusion. Additional material (procedural and design data as well as descriptions of equipment) must be provided in sufficient detail to allow a clear and precise judgment to be made of the importance and originality of the research reported.

It is acceptable for this dissertation to include as chapters authentic copies of papers already published, provided these meet type size, magin, and legibility requirements. In such cases, connecting texts which provide logical bridges between different manuscripts are mandatory. Where the student is not the sole author of a manuscript, the student is required to make an explicit statement in the introductory material to that manuscript describing the student's contribution to the work and acknowledging the contribution of the other author(s). The signatures of the Supervising Committee which preceed all other material in the dissertation attest to the accuracy of this statement.

ACKNOWLEDGEMENTS

I wish to thank all of my dissertation committee members and particularly my advisor W.Jay Dowling for all the useful suggestions that improved the quality of the present dissertation. I am also indebted to Renaud Brochard for his help regarding statistical analyses and the construction of stimuli. Finally, I want to extend my gratitude to Boubakar Diawara of the company InovaSys without whom I may have never worked on the fascinating topic of neural networks.

ARTIST: ADAPTIVE RESONANCE THEORY TO INTERNALIZE

THE STRUCTURE OF TONALITY

(a neural net listening to music)

Publication No. ________

Frédéric Georges Paul Piat, Ph.D.
The University of Texas at Dallas, 1999

Supervising Professor: W. Jay Dowling

After sufficient exposure to music, we naturally develop a sense of which note sequences are musical and pleasant, even without being taught anything about music. This is the result of a process of acculturation that consists of extracting the temporal and tonal regularities found in the styles of music we hear.

ARTIST, an artificial neural network based on Grossberg's (1982) Adaptive Resonance Theory, is proposed to model the acculturation process. The model self-organizes its 2-layer architecture of neuron-like units through unsupervised learning: no a priori musical knowledge is provided to ARTIST, and learning is achieved through simple exposure to stimuli. The model's performance is assessed by how well it accounts for human data on several tasks, mostly involving pleasantness ratings of musical sequences.

ARTIST's responses on Krumhansl and Shepard's (1979) probe-tone technique are virtually identical to humans', showing that ARTIST successfully extracted the rules of

tonality from its environment. Thus, it distinguishes between tonal vs atonal musical sequences and can predict their exact degree of tonality or pleasantness. Moreover, as exposure to music increases, the model's responses to a variation of the probe-tone task follow the same changes as those of children as they grow up.

ARTIST can further discriminate between several kinds of musical stimuli within tonal music: its preferences for some musical modes over others resembles humans'. This resemblance seems limited by the differences between humans' and ARTIST's musical environment.

The recognition of familiar melodies is also one of ARTIST's abilities. It is impossible to identify even a very familiar melody when its notes are interleaved with distractor notes. However, a priori knowledge regarding the possible identity of the melody enables its identification, by humans as well as by ARTIST.

ARTIST shares one more feature with humans, namely the robustness regarding perturbations of the input: even large random temporal fluctuations in the cycles of presentation of the inputs do not provoke important degradation of ARTIST's performance.

All of these characteristics contribute to the plausibility of ARTIST as a model of musical learning by humans. Expanding the model by adding more layers of neurons may enable it to develop even more human-like capabilities, such as the recognition of melodies after transposition.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Why music to explore cognition?

Music provides one of the best domains for the study of basic processes underlying perception and cognition: it is simple enough to be characterized along a few basic physical dimensions, and yet it is a very complex phenomenon. The complexity of music extends well beyond the complexity of its syntax, which music theory aims to specify. Music theory defines rules concerning the relationships between musical elements (e.g. notes, groups of notes, keys, counterpoint and so on...), describing a syntactical complexity comparable to that of natural languages. This would be a sufficient reason to study the mental processes involved with music with the aim of understanding the human mind. But there is something deeper than this superficial complexity. It relates to the diversity of behaviors and experiences music evokes.

Just like other forms of art, music can induce feelings such as tension, relaxation, melancholy, joy, or even euphoria, even though words are hardly adequate to represent emotions. However, music seems to go further than most forms of art in the behaviors and the strength of the passions it elicits. The selling of millions of copies of a work in a year, the gathering of hundreds of thousands of people in one place or the report of out-of-the-body, religious or cosmic experiences (Gabrielsson and Lindstrom, 1994) all testify to the exceptional power of music. For some cultures, music is even a means to reach the state of trance.

The counterpart of this exceptional power of music that is almost never mentioned is the power to annoy people. If you gather a random sample of five persons or more in a room, you can be sure at least one of them will ask you to change the music within 30 minutes whether what is playing is lullabies, rap, new-age, arabic folk, free jazz, indian folk, trash or serial music (unless all of them are polite beyond reason). However, none of them will be made so uncomfortable by the Picasso on the wall. It is not that everybody loves Picasso, but it seems that people cope more easily with a picture or a poem they do not like than with music they do not like. Of course, part of it is because our ears are omni-directional receptors, and that ignoring sounds is more difficult than ignoring an image in the peripheral vision; but informal observations tend to confirm that music has more 'stressing power' than other arts. In this light, the reactions to music look more like a reflex than like a cognitive phenomenon. But what is 'noise pollution' to some is music to others. We can observe many instances of this, and some musical styles never cross barriers between cultures or between generations. What is adored by a particular culture or generation can remain totally incomprehensible to others. The gap does not even need to be as large as a whole generation; sometimes, being confronted with the musical tastes of a brother eight years younger is enough to make you feel two decades older.

1.2 Musical tastes and acculturation

What seems to emerge from these few informal examples is that people mostly like what they know. Radio programme certainly know this, and concluded —apparently, with success— that the more a song is 'hammered' in the listeners' minds, the more likely they will buy the record! In contrast, people are quite insensitive to musical styles foreign to their experience. Some music seems incomprehensible until we learn to appreciate it. This, however, is not at all unique to music but applies to other domains as diverse as wine tasting or face recognition. Just as the novice wine taster may mistake a Pinot for a Cabernet and the person raised in the Western world may mistake a Japanese face for a Korean one, the Western ear will perceive many Indian songs as being identical. The best illustration of this comes from the declaration of Ravi Shankar (probably the world's greatest sitar player) at Woodstock, following the impressive applause of the public after his first ramblings on the instrument: "I think you'll love the songs we are going to play, especially since you liked so much the tuning of the instruments!" Another consequence of people liking what they know is that the early works of geniuses are usually controversial, in scientific as well as artistic domains. Often, geniuses are not recognized as such until their time is over, or at least until the public and their peers got used to their works. For instance, it took time for the works of Van Gogh, Galileo, Beaudelaire, or Stravinsky to be accepted without controversy.

Thus, the issue of what kind of music people love vs the kind they hate is completely tied to the exposure they had to music. Simple exposure to music results in what Francès (1958) has called the 'musical acculturation', building 'perceptual habits':

> "We are prepared to enjoy and assimilate perfection [of masterpieces] by these daily acquisitions, these automatic reactions born of constant unreflective exposure to a mass of secondary works [...] We can conceptualize [a type of musical perception] only as a process of development, and never as simply falling under a 'stimulus-response' schema. We must distinguish between the effects of acculturation —unreflective, involuntary, and resulting from almost passive familiarity with works— and the effects of education, where perceptual development is supported by the acquisition of concepts and symbols that provide for the definition of forms, their elements and articulations."(p.1—3)

As Francès points out, even when no explicit teaching takes place, the simple 'passive' exposure to stimuli is enough to drive a gradual acculturation resulting in a wealth of implicit knowledge, without any conscious effort towards learning. Moreover, one cannot fully appreciate a piece or even a style of music unless the cognitive structures (also called mental schemata) necessary to interpret and understand it are present (Krumhansl, 1983): we need to be 'prepared to enjoy and assimilate the perfection of masterpieces'.

The process of acculturation and its result constitute the basic issues of the present research. More specifically, how do we acquire these cognitive structures? How do they prepare us to hear music, how do they come into play to mediate our perception of music?

Experimental evidence

Most of the time, those cognitive structures have to be acquired through experience and exposure to the music, that is unless the music in question is simple enough to be interpreted by basic, hard-wired, universal structures. Experimentally, the relationship between familiarity and pleasantness was explored by Smith and Melara (1990) and North and Hargreaves (1995). The former study found that novices rated chord progressions as sounding best if they were the most familiar and prototypical. Even a slight deviation from prototypical progressions would make the pleasantness rating drop significantly. In North and Hargreaves's study, subjects were asked to rate 60 musical excerpts on complexity, familiarity, and pleasantness. As expected, there was a positive linear relationship between familiarity and pleasantness. Further, the relationship between pleasantness and complexity took the shape of an inverted 'U', indicating the existence of a range of complexity that is optimally pleasant. If we consider in turn each extremum on the complexity scale, this implies two things. First, that music subjectively too complex to be graspable by the mental schemata of the subjects is not pleasant. Second, that pleasure derives from the extensive use of those mental schemata: music so simple as to necessitate only a small part of the schemata to be interpreted is not pleasant, but rather boring.

Even with regard to musical experiences evoking physical emotional reactions that go beyond the usual pleasantness of everyday listening, Sloboda (1991) emphasizes the importance of learning: "The physical responses described are part of the innate autonomic response system of all human beings. They do not have to be learned.

However, it is clear that the ability to experience these responses in connection with specific music structures is learned." (p.119)

Smith, Kemler Nelson, Grohskopf and Appleton (1994) strongly argue that one cannot seriously study how novice listeners perceive music without referring to the specific tunes familiar to them. They review the evidence showing that for a long time, novice listeners have disappointed experimenters by failing at musical tasks that are considered as being the most basic for music perception (e.g., tasks relying on octave equivalence). It is indeed amazing that most of us cannot recognize easily all musical intervals even after having been exposed thousands upon thousands of times to all of them. Smith et al. proved experimenters' intuitions to be right, namely that novice listeners can recognize intervals between notes quite easily, as long as they are given instructions that will enable them to label correctly their answers. When given two notes and told to imagine it is the beginning of a very familiar tune, novice listeners could consistently pick the right tune out of three choices. In some cases, performance rivaled that of experts. This study suggests that in many instances, the best (and maybe only) way for novice listeners to make their musical knowledge explicit is by referring to the tunes that are very familiar to them.

Internalizing invariants

The acculturation process is commonly referred as taking place through the extraction and internalization of environmental invariants (Shepard 1984, J.Gibson 1966). In other words, the process of acculturation depends on our ability to detect consistancies in the world and to tune our perceptual or cognitive system to them. We extract invariants from the stimuli present in the environment; that is, we learn from exemplars. We have a memory for exemplars. We can recognize a classical rendition of a tune and declare, for instance, 'This is the national anthem'. Moreover, we show generalization to some extent: we can process a new stimulus never heard before, as long as its structure is similar enough to that of the exemplars learned, that is if it conforms to some internalized invariants. When Jimi Hendrix starts playing, again we can declare that it is the national anthem, even though we may never have heard it played in this key before, and certainly not on distorted electric guitar, with such glissandos and embellishment notes. We can recognize this as an interpretation (i.e., instantiation) of the tune because of the underlying structure common to both stimuli, because they share some invariants. Even though Gibson's idea (1966, 1979) seems like a good start for trying to understand acculturation, it leaves many questions unanswered, especially regarding how are the invariants actually extracted and used to mediate perception.

It is quite difficult to talk about the different invariants of diverse musical styles abstractly, without referring to any particular style. It will be useful from now on to focus on a particular style, to get a clearer idea of the issues and arguments. We should just be cautious that our reasoning stays general enough to still hold when applied to other musical styles. For practical and theoretical reasons, Western tonal music is chosen as the style of focus. First of all, living in a Western world, it is easier for us to have access to Western music stimuli. Second, the overwhelming majority of the theories and research applied to music have focused on this style, and these might be useful to assess and understand better the issues of the present research. Third, this style is complex enough to represent a challenge for any theory or model, but in the same time perfectly lends itself for the study of invariant extraction because it is the object of many constraints. These constraints result in the existence of a variety of regularities embedded in the music, which represent as many sources of invariants to internalize. The most prominent characteristic of this style is probably *tonality*, and it is impossible to explicitly understand Western music without knowing what tonality is. Therefore, this concept is introduced in the next section, before we can get a closer look at the issue of mental schemata.

1.3 Tonality...

1.3.1 ...Can be a vague concept

Tonality is a complex concept and does not lend itself easily to definition, as notes Cross (1985, Chapter 1): "Tonality, in particular, is an ambiguous concept; part of its ambiguity derives from the tendency of theorists to use the term to denote different aspects of different bodies of music. [...] Several proposals for formalising the concept of tonality have been made (see West et al., Chapter 2 of this volume). However, these tend either to suffer from an openendedness which limits their ability to generate strong hypotheses about the cognition of musical structure or to be based upon axiomatic premises which themselves should be the objects of empirical study." (p.7—8)

So tonality is not such a simple, binary notion that we could classify all musical pieces as being either tonal or not according to some precise criterion and without controversy. There is a continuity from tonal to atonal, and even within a piece some tonal passages can alternate with atonal ones. Perhaps that beyond formalizations, tonality should be primarily referred to as being a feeling, that could be the emotional/cognitive counterpart of the more physical, sensory-oriented concept of consonance. As a subjective quality, tonality is for each individual dependent on the effects of acculturation, and is greatly affected by the environment. This goes along with Francès' (1988) account of the historical evolution of the sense of tonality, when he states that following some composers' ingeniosity, some well-known patterns of notes were enriched with the addition of new notes, and that "these foreign elements were one after the other *naturalized* and integrated into the stock of sonorities admitted as musical. This did not prevent tonal feeling from being modified little by little by these successive contributions." (p.109)

1.3.2 ...Is choosing a note as center of gravity

More directly related to the stimulus itself, tonality refers to the existence, the choice of a particular tone as the reference, a center around which the other tones gravitate. Randel (1978) refers to tonality as being "a system of organizing pitch in which a single pitch (the tonic) is made central". Thus the idea of hierarchy is already present in the concept of tonality: the central tone, or tonic, occupies a privileged position relative to the others, and fulfills the role of referrant. Choosing a center around which to compose a piece means that the tonic will usually be played more often, for longer durations and at more 'strategic' times than the other notes. Therefore, it will be felt as a better completion of the preceding context, being more stable, more restful, and not needing further resolution to another note. Note that the causality stated here is not obvious. Rather, it is empirically verified through listening, and suggests that musical regularities and expectations form the core of music cognition: if a piece embeds some regularities up to a certain point in time, there is a natural and universal tendency to believe these regularities will be present in the music to come. If the expected note, the tonic, is not played but rather substituted with one of its (pitch) neighbors, a feeling of tension is created. The urge to hear the tonic will be even greater and so will be the feeling of relaxation and completeness when the tonic is actually sounded. Thus the tonic 'attracts' neighbor pitches and tonality can be viewed as a kind of gravity. This is why the word *stability* is also used when referring to tonality.

The problem with the above definitions is that none is suited to establish a criterion applicable to a model in order to determine if the model has abstracted the regularities embedded in tonality. Krumhansl's research, reviewed in Chapter 4, provides a characterization of the strength with which tonality is present in a human cognitive system, of how firmly are the tonal mental schemata established.

1.3.3 ...Implies relationships between tonic and other pitches

In summary, the prime meaning of tonality refers to whether a piece is strongly tonal, with a clearly preferred tonic, or not. The choice of a tonic has consequences on the use of the other pitches too, because some pitches are incompatible with the tonic, whereas some others are very compatible. It follows that tonality in this sense also refers to the existence of a *set of particular relationships between pitches*. Those two things, the choice of a tonic and the particular relationships between pitches, seem to be completely confounded in Western music. Thus, what was said in the previous paragraph regarding the properties of the tonic (more stable, played more often and for longer durations) also applies to a certain extent to some pitches other than the tonic (to the most acoustically similar to the tonic).

Let us illustrate this with an example: if the pitch C is chosen as the center for a piece, it will most likely be the one played most often, and the one beginning and terminating the piece. Knowing this, one can be quite confident that the pitch occurring most next will be G (the fifth of C: C-D-E-F-G = five pitches); if the piece does not start with C, G would be the most likely second choice. G is the pitch class perceptually most similar to C, probably because of the harmonious acoustic relationship between the two pitch classes: the acoustic frequency of a note belonging to the G pitch class will be very close to 1.5 times that of a note belonging to C. Countless theories contend that it is the simplicity of this 3/2 ratio between frequencies that makes the relationship harmonious (see Helmoltz 1885/1954 for instance). Indeed, that this relationship is present in the raw acoustic signal must have consequences on even the earliest stage of sound processing, that is the pattern of excitation of the hair cells in the ear.

Relationships are quantified!

We have just seen that the tonic is the most stable note, and that it conveys some of this stability to closely related notes. The question is, how much exactly is conveyed, and to which notes? If the relative stabilities of the tonic vs all the other notes were known, the concept of tonality would be much clearer. To have the set of particular relationships between pitches (mentioned above) quantified would at least provide us with an operational definition of tonality.

This being known, we could compare it to the set of pitch relationships present in a musical piece or in a mental schemata, and derive the extent to which the piece or schemata in question is tonal. This would tell us the degree of conformance to the 'ideal' tonality.

In fact this is just what Krumhansl and Shepard (1979) did. Using a method called the probe-tone technique, explained in details in Chapter 4, they measured the stabilities of all the different pitches relative to the tonic. They provided the world of music research with this standard against which pieces and schemata can be evaluated, because it specifies the most salient invariants of tonal music. The result was a picture of tonality clearer than ever before, and even more importantly, a way to compute a degree of tonality.

A final note

Tonality also has a second meaning nested in the first: provided that a piece of music is tonal (organized around a central pitch), the tonality of the piece can refer to the key in which the piece is written, that is to *which* pitch is its center. For instance, C is the tonality of a piece if it is written in the key of C (if the pitch C was chosen as the center of the piece). The reader unfamiliar with further musical concepts is referred to the Appendix for an overview of the definitions relevant to the present work.

1.4 About mental schemata

Jones and Yee (1993) emphasize how little we know about mental schemata: "Many other [than time hierarchies] aspects of musical structure (e.g., tonality, pitch contour, melodic and rhythmic structure, etc.) undoubtly determine both scheme acquisition and application (e.g. Lerdahl and Jackendoff 1983; Sloboda 1985; Dowling and Harwood 1986; Krumhansl 1990). However, this area of research is relatively new, and these aspects remain to be explored. As such, it offers opportunities to study the way attending and attentional schemes change as listeners become more familiar with various musical events" (p.98).

However, the research of Dowling, of Jones and of Krumhansl started to shed some light on the intricacies of how music is mentally processed. What is known primarily relates to the way the output (response) of memory processes is affected by different characteristics of the musical input.

For instance, it is now clearly established that melodies' contours, pitch intervals, key relationships and tonal strengths (i.e., degree of tonality) all affect their recognition and discrimination (Bartlett and Dowling, 1980, 1988; Dowling and Fujitani, 1971; Dowling and Bartlett, 1981). Moreover, those features act in different ways on memory performance depending on the delay after presentation (Dowling, 1991; Dowling, Kwak and Andrews, 1995). They were also shown to interact so much with the melody's rhythm that Jones (1993) tested the hypothesis that what serve as anchor points in remembering a melody are the places where melodic and rhythmic accents coincidence. She concluded that melodic and rhythmic structure are psychologically inseparable.

We also know that we can aim our attention to particular pitch-time windows to recognize a familiar melody even if it is interleaved with distractor notes (Dowling 1973, 1990; Dowling, Lung and Herbold, 1987; Andrews and Dowling, 1991). This may relate to the greater difficulty in detecting mistunings for unfamiliar scales (e.g., pelog scales from javanese music) relative to familiar ones (Lynch, Eilers, Oleer and Urbano, 1990). Sloboda and Edworthy (1981) showed that it is easier to attend to several simultaneous melodic lines if they are in harmony with each other.

Some studies have illustrated Rosch's (1975) idea of cognitive reference points with musical phenomena. For instance, tonal melodies are good reference points, in that they are easier to remember than atonal ones. Furthermore, those two types of melodies are more easily confused with one another if the atonal melody is presented first rather than second (Krumhansl, 1979; Bartlett and Dowling, 1988). This asymmetry suggests that atonal stimuli are encoded in memory by reference to existing, tonal mental schemata. Apparently, when accessed later, the memory trace for atonal stimuli is quite different from what was really played. This might be true for tonal stimuli too, but to a smaller extent. Hence, the memory trace for atonal melodies is said to be unstable.

All these results point out the pervasiveness and automaticity of the action of mental schemata: most of the tasks used in these experiments could be carried out efficiently if the required decision was made on a purely sensory basis, preventing context effects. Apparently, as soon as more than one note is played, the stimulus is interpreted as music and correponding processes and schemata are engaged, taking into account the context for processing. That infants outperform adults in some discrimination tasks perfectly illustrates this (Lynch et al., 1990). However, once mental schemata are acquired, they seem impossible to 'turn off'. They become fully integrated to our perceptual system.

So how is it that two different notes or intervals are not equally remembered? How is it that the same note is remembered differently depending on the preceding context? Familiarity certainly plays a role in all this, but the remarkable thing is that it applies even to stimuli never heard before, with which we are not familiar at all. The musical style is what is familiar, but how do we recognize this style as familiar? How do we recognize the invariants as being the same? And how did we derive an idea of style just from being exposed to many particular melodies?

1.5 Overview

Those are the kind of questions explored in the present work. The quote from Jones and Yee (1993) in the previous section perfectly summarizes the context of the research presented here. Many 'traditional' models of music perception have addressed the questions above. They are often built on assumptions borrowing from Gestalt or music theory, usually depending on whether the focus of the model is on rhythm or on tonality. Most of them offer good insights regarding analyses of particular pieces of music, or can predict the key in which a piece is written, or even try to predict the final state of a memory representation of the piece (Deutsch and Feroe, 1981; Lerdahl and Jackendoff, 1983; Winograd, 1968; Simon and Sumner, 1968; Narmour, 1991; Jones, 1981). All this relates to certain aspects of our perception of music, and accounts for some features of memory for music. However, none of these models seems to address the issue of how mental processes actually carry out such basic tasks as recognizing a tune or mistaking one note or chord for some others (but Bharucha 1987 does). Moreover, the issue of the process of acquisition of mental schemata seems virtually untouched. It is a great challenge for a theory or a model to account for a unitary system that can develop through exposure to exemplars, memorize them and generalize from them. It seems like anything less specific than a model would be quite vague or abstract, and could hardly account for all this in a clear and convincing way.

Building a model could help us understand how some of the musical perceptual phenomena mentioned above arise, after simple exposure to musical pieces has resulted in the learning of a musical style, in the extraction of the regularities of its musical environment. It can help us understand how memory for exemplars and generalization can co-occur. Ideally, a good model of music learning should be able to become familiar with any style of music it is exposed to. But for a start, as has been argued before, we will focus on Western tonal music to test the model and explore in depth if the regularities present in this style are internalized by the model. Only after the model has been shown to 'understand' one style is it appropriate to test whether the model can learn any style.

A class of models that seems very appropriate to address this issue is the class of connectionist models, because they adapt to their environment. Also called Artificial Neural Networks (ANNs), these models not only learn the specific stimuli they are exposed to, but can generalize their behavior (or output) to novel stimuli they have never 'seen' or 'heard'. This is a highly desirable feature for our model, since we want it to be sensitive to the familiarity of both the specific pieces it was exposed to *and* the musical style they define together. Chapter 2 introduces the general principles of ANNs and the proposed model ARTIST.

Chapter 3 illustrates the basic functioning of ARTIST with some simple simulations showing how familiar melodies can be recognized. Then the model's ability to focus on pitch-time windows to recognize a familiar melody among distractor notes like humans (Dowling, 1973) is tested. Chapter 4 surveys Krumhansl's research on the characterization of tonality in mental schemata. This will provide us with a reference against which the model's schemata can be compared, telling us whether ARTIST was able to internalize tonal invariants from its environment. Chapter 5 will implement a simple Markov model of music perception so that comparison with ARTIST will help us understand the latter better. Chapter 6 provides a review of the development and acquisition of tonal schemata. Then the developmental process followed by ARTIST will be examined and compared to humans'. Finally, Chapter 7 will use ARTIST to make predictions about music perception, and an experiment with humans will be proposed to test these predictions.

CHAPTER 2

ANNs, ADAPTIVE RESONANCE THEORY AND ARTIST

After a brief summary of the possibilities demonstrated by ANNs, this chapter will explain the basic principles on which they are built. It will follow with the presentation of a particular class of ANNs, the ART networks, based on Grossberg's Adaptive Resonance Theory. Finally, a model instantiating this theory, baptised ARTIST (Adaptive Resonance Theory for the Internalization of the Structure of Tonality), will be proposed.

2.1 What ANNs can do and their applications to music

In the last years, it seems that connectionist research has undergone an exponential increase in popularity, for a plethora of reasons. Amongst them, let us mention for instance connectionist models' abilities to learn from examples. This means that such models are sensitive to their environment, they develop according to which stimuli are presented to them. This is a great advantage over models born from more classical approaches of artificial intelligence, for which the rules governing behaviour have to be known and made explicit. Moreover, the latter perform poorly in case of change in the environment, because they were not programmed to respond to such a change. In contrast, the former only need a new period of learning (through exposure to the new environment) until they can perform well again. Other interesting capabilities of ANNs include their ability to generalize their behaviour to new situations, to behave consistently even in the presence of noise, in case of incomplete information or of degradation of the system itself, and the ability to explicit a metaphoric view of some neurophysiological processes found in the nervous system.

Connectionist models have been used to simulate a variety of processes involved with music perception and cognition. Most of them focused on processes primarily working on the pitch dimension of music, such as the perception of pitch (Sano and Jenkins, 1989), of chords (Laden and Keefe, 1989) or of tonality (Bharucha, 1991; Leman, 1991), or the generation of expectancies (Bharucha and Todd, 1989). Some dealt with the purely temporal dimension of music (rhythm), being applied to the segmentation (Carpinteiro, 1996) or the quantization of musical time (Desain and Honing, 1989), and a few integrated both dimensions to address cognitively higher level problems such as string instruments fingering (Sayegh, 1989), Jazz improvisation (Toiviainen, 1995), the categorization of musical patterns (Gjerdingen 1990) or the composition of melodies (Todd, 1989; Mozer, 1991; Lewis, 1991).

Probably the most popular amongst psychomusicologists is Bharucha's model MUSACT (1987), even though it did not involve any learning but was rather handcrafted according to the principles of music theory. Specifically, three levels were postulated to build the model, one for pitch classes, one for notes and one for keys. The relationships between those three levels were also postulated according to music theory. MUSACT was able to predict the performance of humans in detecting changes in a sequence of chords, as well as the perceived relatedness of two chords. The general contribution of this model is that it showed that even with a fairly simple connectionist model, it is possible to simulate human data fundamentally linked to the perception of tonality.

The simplicity of the model is a great advantage in that it makes it relatively easy to understand how it works. Moreover, controlling the knowledge built into the structure of the model gives us a good insight regarding the information sufficient to perform a certain task. Laden (1995) extended this model so it could handle a greater variety of keys and of chords. The resulting neural network gives a good account of some human data regarding tonality.

2.2 Basic principles of ANNs

A neural network is composed of units, also called neurons because of their similarities with brain cells, and of links connecting them to form a network. Every neuron fulfills a role, integrating information coming from other neurons and distributing the result of this integration to other neurons (McCulloch and Pitts, 1943); unless they are input units, in which case they transmit information from the outer world into the network.

The integration of information consists of summing all incoming signals or input intensities. Most of the time, the resulting sum is rescaled through a transfer function so the activation of the neuron will fit within a given range (usually $[0, 1]$ or $[-1, 1]$). The resulting activation is propagated to other units through more or less efficient connections or synaptic weights. For instance, a synaptic weight equal to 1 will carry 100% of the activation a, whereas a weight equal to 0.5 will transmit half of it ($0.5 \times a$). This is illustrated in Figure 2.1.

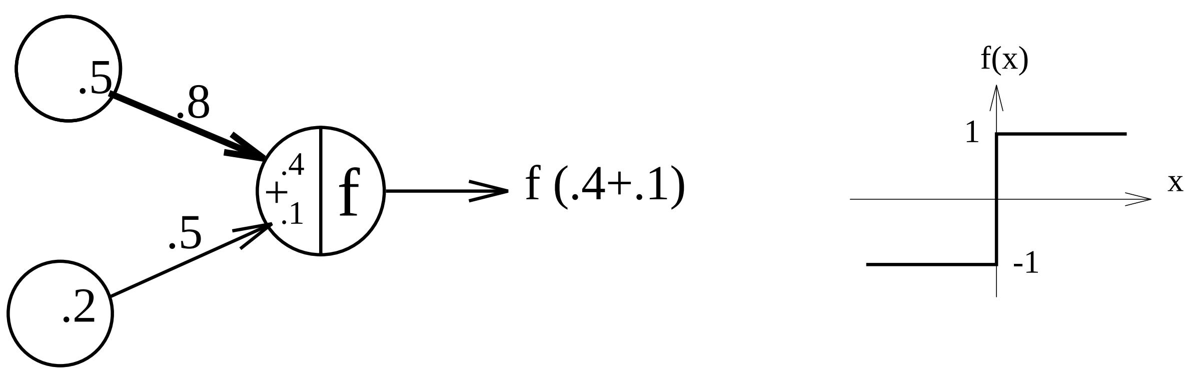


Figure 2.1: The formal neuron.

Several learning rules are usually available when teaching a task to an ANN. They are algorithms by which the network automatically updates its connection strengths to reach a stable internal structure, allowing internal representations to develop (see for instance Rumelhart and McClelland 1986, Rumelhart, Hinton and Williams 1986). Thus, depending on the type of ANN implemented and the position of a neuron in the network, the function of this neuron can be imposed *a priori*, as it has to be the case for input units, or it can be acquired through learning, that will dictate what information the neuron will receive through its incoming connections.

Therefore, the possibility of teaching an ANN to perform a certain task mostly depends on the information it is fed in input, and on the learning rule used to modify the connections. The following section presents a particular type of ANNs called Adaptive Resonance Theory (ART) networks, as it will serve as a basis for the model presented in this thesis. This type of ANN is specified by the choice of a particular architecture (the general pattern of organization neurons and connections follow) and of a set of rules governing the learning and the spreading of activation in the network.

2.3 The Adaptive Resonance Theory

2.3.1 Why the ART ?

The Adaptive Resonance Theory was chosen as the framework to implement the new ANN model presented in this chapter because it integrates many features desirable to meaningfully process musical information and to achieve the kind of performance we want it to be able to simulate. Specifically, it integrates unsupervised learning/self-organization, intrinsic hierarchical structure, top-down activation, account of attention, biological plausibility, and both learning of categories and of examplars. Moreover, Gjerdingen (1990) built an ART-based model that "demonstrate[s] how untrained listeners might be able to sort their perceptions of dozens of diverse musical features into stable, meaningful schematas" (p.339), with promising results. Griffith (1993) also succesfully applied ART networks to the identification of tonality in music.

Concerning the supervised/unsupervised aspect of the learning involved in an ANN, one needs to be cautious not to generalize too fast. It is easy to mistakenly infer supervised learning from back-propagation learning. Even though this is often true, back-propagation learning can be really unsupervised like in Bharucha and Todd's (1991) recurrent network model used to learn sequences of musical input, where the target used to compute the error signal back-propagated is in fact the next stimulus in a series, and is not provided by an explicit teacher. So ANNs learning with back-propagation are not excluded *a priori* for this reason, but because they typically lack the attentional component and the hierarchical architecture of ART networks.

The hierarchical structure is not only important to mirror musical pieces' structures (Lerdahl and Jackendoff, 1983), thus predisposing the neural network to extract this structure. It also allows the learning of arbitrarily long sequences of notes without requiring an incredible amount of resources, as is the case in humans: After sufficient exposure, we can easily remember every single musical event (or musical chunk) of an hour long symphony, which means a sequence of the order of tens of thousand events. By recursively putting together 'chunks' of notes, the hierarchical network can memorize sequences of an exponentially growing length as the number of levels increases linearly! This performance may be possible for a simple feed-forward neural network, but the memorization may not be so perfect because of overlap and interferences between learned patterns.

2.3.2 Basic principles of ART

Grossberg (1982) developed with Carpenter (1989) an ANN model for pattern recognition, based on fuzzy adaptive resonance theory (ART; see also Carpenter and Grossberg 1987), and taking into account a variable they call vigilance. A single parameter to modulate the effects of attention sounds at first very limited, but the network is not claimed to be a model of attention, but rather a model of learning incorporating an attentional component through a parameter of vigilance, that influences the network's learning behavior in a major way. ART networks are used to classify the elements of a set of inputs into categories. Hence the networks comprise 2 levels exchanging information: F_1, in which the activity pattern encodes the stimuli, receives the inputs given to the system; and F_2, in which the nodes represent one category each, gives the response of the network to the categorization task. The learning is unsupervised, meaning that the network is never exposed to a right answer. Rather, it has to derive categories by itself, from the similarities between inputs. We can visualize the learning problem the network has to solve as covering the input space with boxes, a box figuring a category (Figure 2.2). That is, for any input I, an F_2 node will respond with a high activation.

2.3.3 Learning and resonance

In the learning phase as well as during the categorization task, the definite assignment of a stimulus S to a particular category requires two steps. First, the best fitting (most activated) candidate category Y is computed. Second, the choice of Y has to be confirmed, by verification of the matching criterion. This depends on the vigilance parameter and guarantees that S possesses enough features defining the category Y to belong to it. Thus, the activation of an F_2 node cannot be directly interpreted as the input I belonging to this category, because the activation of this node may not be very significant. The activation of a category node has to be high enough before we can validate and interpret the response. This means that competitive learning has to take place to insure that in the end only one node is active within F_2 (the categories do not overlap). Furthermore, it means that new units can be added during learning to learn new input patterns if it does not fit well in even the best matching category; whether the input fits well or not in the category depends on the vigilance. This is sketched in Figure 2.2. Here are the steps followed by the network after presentation of a stimulus S:

a) S creates a pattern of activity I across level F_1.

b) This activity propagates through the weights (where Long Term Memory traces are stored) and in turn creates a pattern of activation across level F_2.

c) A 'winner takes all' strategy is applied (lateral inhibition) to level F_2, so that the most highly activated node Y is selected, representing the most probable category. Thus the activation pattern in F_2 becomes composed of one '1' and the rest are '0's. This can be understood as 'making an hypothesis' about S's category.

d) Activation is sent back through the same weights to level F_1 to create another activation pattern X. Because of the way it is obtained, X would be the stimulus most activating the designated category, so it plays the role of prototype for the category Y.

e) Then I and X are matched, to see if the stimulus S can fit into the category designated by being close enough to its prototype. This can be seen as 'testing the hypothesis'.

If a match is found (resonance occurs), S is recognized as belonging to the category. If we want to use S as a learning pattern, then the matching of I and X (their intersection $X*$, that can be interpreted as the features of S the network pays attention to) is used as the pattern of activation at level F_1 to modify the weights from there to the category node.

If S does not match the activated category, *mismatch reset* occurs: The category is disabled and the search for another category starts. Once all categories have been tried and if S does not belong to any, a new node is created in F_2 which weights are set up to perfectly resonate with S (the vector of weights is set equal to the activation pattern vector I).

Through all those steps the idea of vigilance appeared once, in the criterion used to decide whether resonance occurs, determining in fact if I is considered close enough to the prototype X to fit in its category. Expressed mathematically, a match is found if:

$$\frac{|I \wedge w|}{|I|} \geq \varrho$$

ϱ is the vigilance parameter, between 0 and 1. The fuzzy AND operator $\wedge$ gives a vector in which each component is the smallest between the corresponding components of I and w:

$(I \wedge w)_i = min(I_i, w_i)$ (In the binary case, if I and w are binary vectors, the operator reduces to the logical AND)

Therefore, this quantity is always smaller than I and w (or equal). So the ratio has to be smaller than 1, and this justifies that vigilance cannot be greater than 1. The fuzzy AND operation can be understood as removing from S the features that do not match with the category (this occurs for all i verifying $I_i > w_i$; what is left over is the recognized part of the stimulus, what the network pays attention to: what is common to the stimulus and the prototype). Then, dividing by the norm of the input $|I|$ normalizes, so the ratio value is:

1 if the patterns are identical, or if all the features of S are present in the defining features of the category (if $I_i < w_i$ for all i)

close to 1 if they are alike, sharing many features

close to 0 if they are very different, sharing few features

0 if they are orthogonal

Then it becomes clear that setting a threshold to be compared with this ratio is equivalent to giving a tolerance whithin which the pattern I is to be accepted in the category. So, a low vigilance leads to a broad generalization and to the abstraction of prototypes. If the vigilance is close to 0, the criterion will be easily matched, so that a stimulus will be accepted in a category even if it just holds a vague resemblance with the elements of the category. We can indeed in this case say that the network does not pay much attention to the stimulus. If the vigilance equals 0, then any stimulus will be accepted in the category, therefore only one category will be created, grouping all the stimuli together: one box covers the whole space, and the network does not show any discrimination.

Conversely, if the vigilance is close to 1, a stimulus needs to be very close to the prototype in order to be accepted into the category, the network pays a lot of attention in the sense that it shows sharp discrimination. In the extreme case, with a value equal to 1, each stimulus creates its own category since it cannot fit in any other, and exemplar learning is thus reached (see Figure 2.2). This kind of learning is to avoid when the goal is generalization. The problems caused by a too high attention are well shown in Luriia's "The mind of a mnemonist" (1968). This book shows how this person with photographic memory had such difficulties to recognize anybody, since we look slightly different from one time to the other. All the problems associated with the lack of forgetting and the consequent lack of analytical abilities are also presented, pointing out that perfect memorisation can be undesirable.

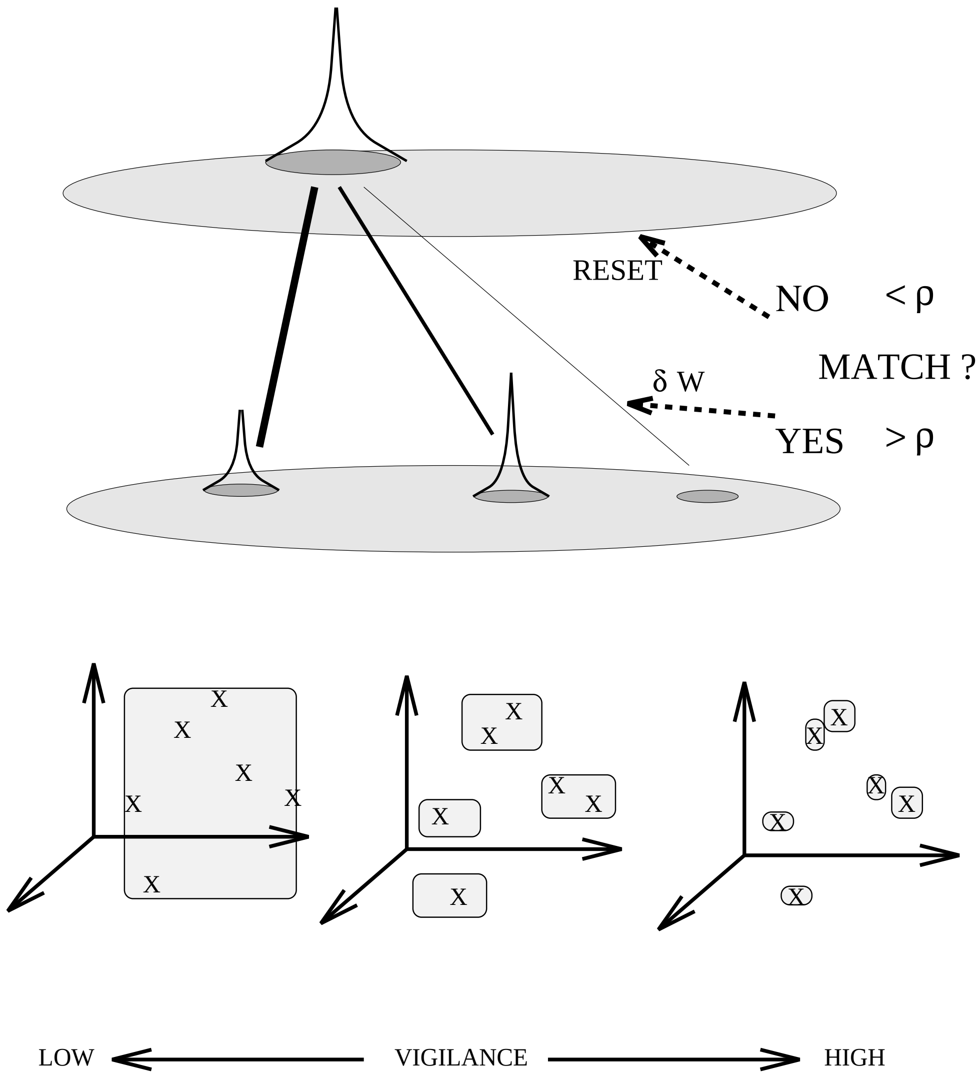


Figure 2.2: The learning process of L'ART pour l'art and the covering of the input space as a function of the vigilance criterion.

Both the fundamental role of lateral inhibition and the discard of the irrelevant features of a stimulus present in the model can explain why some psychological models see attention as a filter (Broadbent, 1958), bringing the idea of a limitation to what can be activated or attended.

2.3.4 Top-Down Activation

Another remarkable aspect of ART-based models, completely related to the point made above, is that it accounts for a role of the top-down activation: the input I is confronted with the prototype X of the most probable category (the most activated), or 'expected prototype', thus implementing the testing of an hypothesis. This confrontation occurs in F1, the level receiving inputs from the external world, which can be analogous to the Layer I of the cortex. This is consistent with Cauller's (1995) view that top-down activation is essential in directing attention to critical features in order to probe the external world (testing an hypothesis) through adjusted behaviour (eye movements or probing finger movements), and that this processing occurs at the convergence of bottom-up and top-down activities, namely in the Layer I of the cortex. Moreover, a consequence of the convergence of both signals is the filtering out of irrelevant features present in the stimuli, a kind of noise filtering, which translates into an increased stability, constancy of the way the external world is perceived, another likely function of top-down activity. It is this same filtering responsible for the non-optimal processing of unfamiliar stimuli, those inconsistent with our mental schematas (e.g., the low memorizability of atonal musical sequences). This phenomenon even extends up to the semantic level, as shown by studies about the memorization of causal relationship in short stories, exhibiting that in long term memory, people tend to 'normalize' or 'rationalize' details of stories incompatible with their belief systems because taking roots outside of their culture. I allude here to stories in the style of "The war of the ghosts", studied by Bartlett (1932) among others.

2.4 ARTIST

The general principle that has guided the development of this model was simplicity. Imposing as few constraints as possible on the model has the benefit of making it very general and realistic. For instance, using some knowledge of music theory to build the model would reduce its chances to work with other musical styles, or may defeat the whole purpose of having it learn by itself: any property emerging from learning could be attributed to the presence of the 'built-in' knowledge.

Nevertheless, imposing some knowledge into a model can be useful to test how important is that information in order to solve a particular problem. For instance, Gjerdingen's (1990) model L'art pour l'art (described below in Section 2.4.1) has two input neurons devoted to the coding of the contour of the melodies it is presented. It appears that the model makes extensive use of this information in order to classify its inputs into different categories. Therefore, we know that melodic contour is an important feature to determine the similarity of melodies. In contrast, one goal in building ARTIST is to see what can really emerge from simple, passive exposure to music, and which invariants can be extracted from the stimuli alone. Therefore no knowledge is a priori imposed on the model. With this approach, we cannot know which specific information is used by the system to solve a problem; but on the other hand we can find out whether the input data fed to the model is sufficient to solve a problem or not, i.e. if it contains all the information necessary to find a solution.

Another advantage of simplicity is that it will be easier for us to understand the model. Not that this would be a pre-requisite for a model to be good. After all, we are far from understanding the brain very well, and making a model as powerful, even if as ununderstandable as the brain, would be a feat. But understanding the model can give us insights regarding why it works or not, so we can develop it further or fix it. It can help us exploit it to its full potential, for instance to make predictions. And most important of all, it can help us know why our mental schemata develop a certain way and how they are put at use.

To illustrate how the principle of simplicity applies to the present model, it would be good to briefly present a very similar but much more complex model, Gjerdingen's (1990) *L'art pour l'art*, also based on ART theory. Contrasting some aspects of both models should help the reader appreciate the merits of simplicity. Therefore, the next section is dedicated to an overview of L'art pour l'art, with special emphasis on features that will differ from ARTIST's.

2.4.1 L'art pour l'art

Essentially, L'art pour l'art is a series of ART networks organized sequentially. There are only two such networks for now, but some more can in principle be added. Each network being composed of two levels (F1 being the lower levels and F2 the upper levels), this brings the total to four levels overall in the model. The principle underlying the choice for such an architecture is that a long sequence of events has to be segmented into chunks to be memorized, and that the length of those chunks is constrained by the capacity of our short-term memory (STM). A multilevel hierarchy of layers of neurons then permits the concatenation of the chunks. A sketch of the architecture of L'art pour l'art is presented in Figure 2.3 to show how it can learn sequences of symbols by hierarchical chunking.

The two F1 levels implement a STM, where a node represents one item. If a node is activated, this means the corresponding item is present in short-term memory. In agreement with psychological data (Miller, 1956), these memory capacities are limited to the simultaneous activations of about six elements. The few activated nodes thus represent the sequence of the few last items activated in memory. Therefore, to make sure F1 levels do not end up to be so unrealistic as to contain dozens of items simultaneously active in memory, some gain control was implemented through the use of non-specific lateral inhibition within F1.

More specifically, the F1 layer of the bottom network is the model's input, where the notes of the piece presented are coded on a total of 35 nodes. Pieces made of up to three different voices (melody, bass and inner voice) can be presented. Thirty-three nodes are needed to code them, each voice's input being coded on eleven nodes. Nine of them are used to code the note, in a distributed way: seven nodes for the diatonic notes (C,D,E,F,G,A,B) plus two nodes specifying the alterations (# or b) can designate any of the twelve pitches (C,C# or Db, D , D# or Eb, E , F , F# or Gb , G , G# or Ab , A , A# or Bb, and B) when taken together. The two last nodes are used to code the contour defined by the present note relative to the previous one, up or down. Two more nodes are added to F1 to signal the presence of particular musical relationships between voices (harmonic tritone and contrapuntal dissonance). Those relationships of contour, tritone and dissonance could be derived by the model from the rest of the input, but having those features already extracted makes its task easier. As was mentioned in the previous section, the other merit of imposing this knowledge into the model is that it is like formulating an hypothesis regarding the importance of this information. Observing that the model grows strong connections from those nodes indicates that this information has a strong influence on the model's behavior and therefore lends some support to the hypothesis that the information is important.

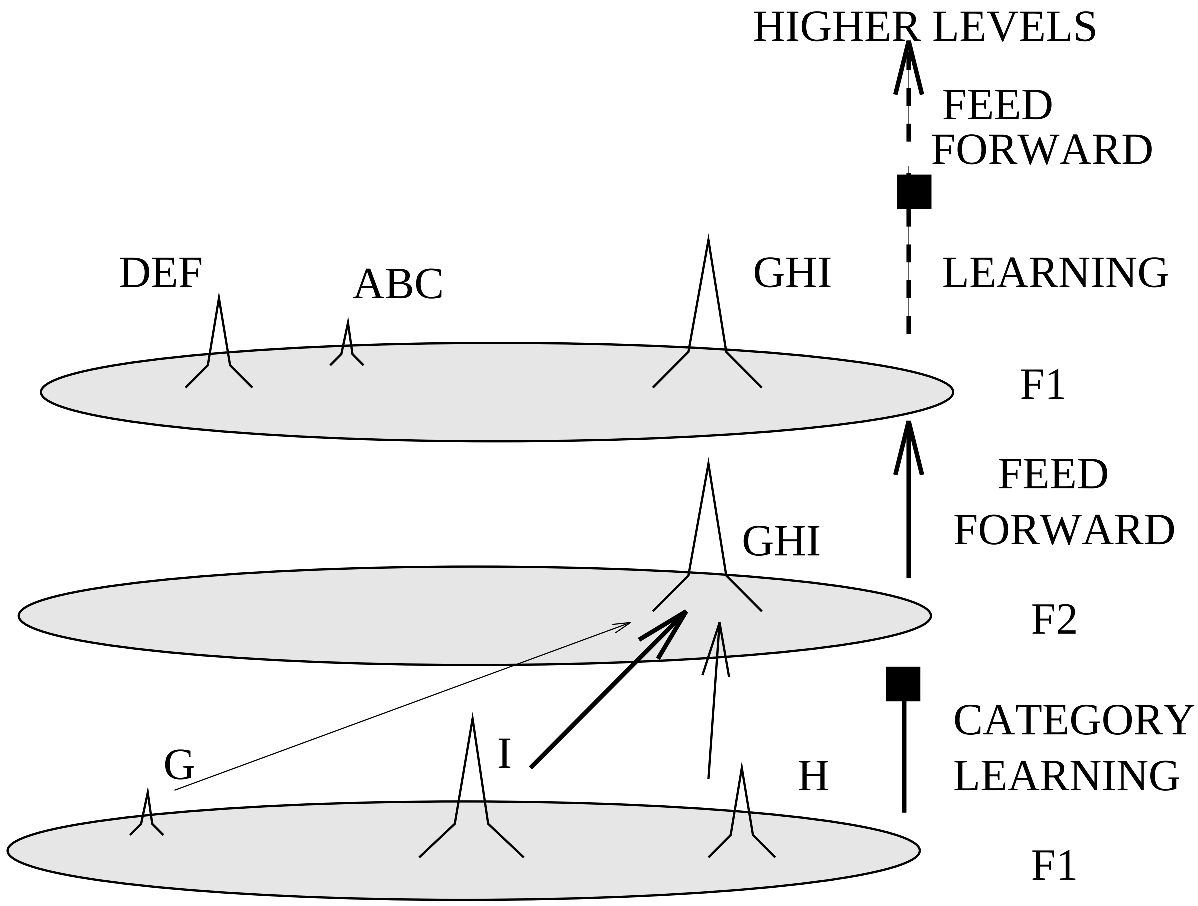


Figure 2.3: The learning of a sequence of symbols by L'art pour l'art.

Within F2 levels also, lateral inhibition is of prime importance. F2 neurons represent categories, activated by the co-occurrence of activated features at the F1 levels. They can be considered as a memory 'chunk', or as a concept. The bottom network's

F2 is made of neurons representing basic musical concept that can be defined from note sequences up to 6 notes (the limit of the F1 memory transmitting activation to it). The top network's F2 is made of neurons representing more elaborate musical concepts, defined from sequences up to 6 basic concepts (the limit of the upper F1 memory keeping trace of the few last basic concepts activated).

To classify a stimulus in only one category at a time, F2 neurons are considered on-center, off-surround cells. That is, a 'winner-takes-all' mechanism is realized at F2 levels: a newly activated neuron will turn off the activation of all the others within that same level. Thus this is equivalent to lateral inhibition too.

The benefits for using lateral inhibition at all levels are many. Mostly, it prevents the amplification of noise in the system, by inhibiting nodes that are only slightly activated and not very relevant to the information being processed. It also makes the model more realistic when viewed from a psychological perspective, by imposing a limit on STM capacity. Finally, it puts the model in a situation of forced choice at every level: the winner-takes-all strategy used at every F2 level insures that a bundle of features is only classified as belonging to one category. This makes the output of the model clearer and more interpretable. Whether this is psychologically relevant probably depends on what task the model is assigned, on what kind of output is expected from it. Lateral inhibition is further discussed in Section 2.4.3.1.

2.4.2 Coding scheme

To build a model that is based on the Adaptive Resonance Theory and capable of internalizing the structure of tonality, a choice has to be made first regarding the interface between the system and its environment. In other words, some information from the outer world has to be made available to the system. Let us note right now that this step is of crucial important, for several reasons. First, obviously, any system's learning potential is bounded by the amount and content of information it is provided with. Second, for a given architecture, the learning potential may be further restricted by the particular way in which the information is coded. The solvability of the 'XOR' problem illustrates this perfectly: coding a problem with redundant bits of information —a priori useless— enables a single neuron to learn to calculate the logical operation XOR. Third, in the particular case of auto-organizing networks such as ART, it is the *whole architecture of the system* that depends mostly on the coding of the information, and on its interaction with the learning parameters. This section explains the coding scheme used to code the information from the environment in a form that is suitable for presentation to the model.

2.4.2.1 Pitch dimension

Ideally, stimuli could be fed to the system through a microphone directly picking up the airborne acoustic information. However, extracting the notes underlying music from such a signal constitutes a whole area of research of its own. Outside of this area of research, almost all the models applied to the perception of music, especially the theoretical ones, assume a pitch-class based representation of music. That is, they take the phenomenon of octave equivalence for granted and work on sets or alphabets of 12 elements, or have 12 input units. Such a simplification is convenient because it somewhat forces the model to focus on the tonal aspect of music: the relationships between these 12 elements. However, in our case this would go against the principle of simplicity and would undermine the generality, and meaningfulness of learning and the realistic aspect of the model.

Moreover, assuming octave simplification has several drawbacks. First, octave equivalence is not always the absolute, universal principle we tend to believe it is, even though it is a well established phenomenon, and very compelling for whom hears it. According to Smith, Kemler Nelson, Grohskopf and Appleton (1994), this phenomenon is not robust at all in novice listeners, and generations of them have annoyed researchers by failing to exhibit octave equivalence. Even more important, coding every note according to its pitch class would imply that the contour of the melody is lost because of the circularity of pitches' organization. For example, if all we know is that a G follows a C, whether the interval was up by 7 semitones or down by 5 semitones is unknown. Thus a pitch class coding of musical input is not a realistic coding of humans' auditory input (even though it is probably relevant if we consider the result of the first few processing stages occurring in the brain). The fundamental importance of contour in music is established by many of Dowling's studies (1978, 1991; Dowling and Bartlett, 1981; Dowling and Fujitani 1971; Dowling, Kwak, and Andrews, 1995), and compellingly confirmed by our musical perceptions: We usually do not hear the music in terms of different arrangements of just 12 notes, but rather in terms of notes going up or down in small or big leaps. Shower singers rarely hit the right notes, but usually get the melodic contour right. Moreover, the fact that this is often enough to make the melody recognizable points out the primary salience of contour. Dowling and Hollombe (1977) examines in details the relationship between contour and octave equivalence.

It follows that in order to make the input realistic and enable it to include contour information, accompaniment, chords, different chord inversions, and maybe even different timbres, we need a one-to-one correspondence between input node (neuron) and note played. The corpus of musical pieces (described in Section 2.4.4) makes use of 6 octaves. Thus 72 ($= 6 \times 12$) neurons were required to code the musical input in terms of relative activations of notes. To have a properly called unsupervised learning, no other unit should be necessary. Unlike in L'art pour l'art, no preprocessing of the inputs was implemented other than computing the exact activation level of a unit. This is detailed in the next section.

2.4.2.2 Time dimension

When a note is played, the activation of the corresponding input node is set to one. Thereafter, the activation of this node passively decays, thus simulating the effects of diverse short-term memory processes such as echoic memory or the phonological loop (Baddeley, 1990). The activation of an input unit is strictly a function of the time elapsed since its corresponding note was sounded. The activation is assumed to decay exponentially with time t:

$$a(t) = \frac{a(t-1)}{log_2(t/T+2)}$$

where T is half the half-life of the note. In other words, the activation is divided by 2 every time $t = 2T$.

To make the simulations computationally tractable, the time dimension had to be discretized. This means that the activation state of the input and of the whole network was refreshed at regular time intervals. This was determined by the parameter T and thus could be varied in different simulation studies. This is explored in more details in Chapter 5. For now and almost all of the simulations to follow, the time interval was chosen to match the duration of one measure. This choice was natural considering both musical and psychological theories. Musically, the measure is a division of the time dimension that usually contains a few notes, and this is in agreement with how the cycles of rhythmic attention modulate our perception of music (Jones, 1986; Jones, Boltz and Kidd, 1982).

In summary, at the end of every measure, the activation of the input layer is updated. First, the residual activation resulting from the decay of the activation of previous measures is computed. Second, the activations for the notes played within the last measure are computed depending on how long before the end of the measure each note was played. The velocity, strength with which the note was played, is also taken into account, in that the original activation of the corresponding input node is proportional to the velocity.

2.4.3 Assumptions

After a coding scheme has been determined, the architecture of the network and the rules governing the learning (modification of synaptic weights) and spread of activation in the network have to be chosen. To start simple, ARTIST is first limited to the basic structure of ART networks, with only two layers or fields of neurons, F1 and F2 (as opposed to the two networks/four layers of L'art pour l'art). Some more can be added later if the complexity of the tasks we submit ARTIST to require it.

The rules used for ARTIST are essentially the same as those for ART2 networks, that were summarized in Section 2.3.3. However, a few simplifications were made (mostly two important ones), at least for a start. If the model's behavior appears to be inefficient or unreliable, unstable, it is always possible to add some constraints to match all those applying to ART2 networks.

2.4.3.1 Lateral inhibition

Input layer F1

First, lateral inhibition was not used in the input layer F1. The passive decay is expected to limit the number of items activated in STM. However, if a musical piece is played at a very fast tempo, the note events may appear in STM at a much faster pace than they vanish due to activation decay. As a result, the STM could be overloaded with many more items activated than can be handled. This is not necessarily unrealistic. The limits on STM capacity concern the number of items that can be processed efficiently. It is always possible to have a number of items exceeding this limit activated in memory, only some of them will be lost at a further processing stage or will influence their outcomes in an insignificant way.

This is probably what happens when one tries to follow the music of Charlie Parker at the climax of a solo, or even on the first bars of "Leap frog" (1950). He commonly plays several consecutive measures entirely made of 16th or 32nd notes, which means that he plays 16 or 32 notes within the duration of one measure. In a normal attending mode, trying to catch and process every note, one gets completely overwhelmed by the number of events. The sequence of notes is perceived as a continuous stream which is very difficult to parse into elementary events. Unless you are one of the best musicians in the world, many of these events must be somewhat lost to optimal processing, since it is almost impossible to sing back the phrase just played or to count the number of notes it contains, even if it fits within the temporal span of the phonological loop. This probably explains why the public was not very receptive to his music at first. However, one gets a better impression of what is played and of the whole *Gestalt* by switching to a more rhythmic mode of attending (Jones, 1982). Then, the listener can focus his attention on the most salient notes and hears the other notes as interpolations from these anchor points, filling out the background. Practically, this means that if you have internalized the 'swing' (property of where the accentuated notes are), it will be much easier to follow Charlie Parker's solos, even if you do not catch every single note played. To make a crude parallel, one can hardly get an appreciation of impressionist paintings by looking at it through a magnifying glass. Rather, a certain distance from the painting is needed in order to have an overall impression.

In summary, not imposing any limit on STM is not necessarily unrealistic. The real issue here is to find a value for T (time window governing the decay rate) that will be consistent with the limits of human memory processing. For a start, the T value is set at half the duration of a measure, which means that the activation of items in memory is halved every measure.

Abstract layer F2

Lateral inhibition is not always used in F2. As mentioned before, it forces the choice of the model to activate only one category. Thus, whether inhibition is appropriate depends on the task at hand. During learning and top-down activation propagation, it is preferable to have only the F2 winner active, to provide some stability and avoid the building up of noise in the system. However, for some tasks, we may want to know about the relative activations of all the categories, so the output of the network will be recorded before a winner is ever chosen. This could mean also that the process of inhibition itself plays an important role in our perception of music. For instance, which category came with the second highest activation and got inhibited by the winner may be of crucial importance. This information is available before lateral inhibition takes place, but is lost thereafter. Considering the importance of the inhibition process itself is in line with Meyer's theory of emotion (1956) stating that affect will arise each time a positive response is inhibited. This principle forms the basis for the claim that emotions triggered by music are mostly a consequence of preparing, and then fulfilling or denying the listeners' expectations.

In fact that will be the case most of the time that ARTIST's raw output is recorded before a winner emerges. My intuition was that allowing the same stimulus to activate many different categories at different degrees would preserve the complexity and richness inherent to music, which could be responsible for many musical phenomena. Music would certainly be much less interesting to listen and to study if any piece had one and only one way of being mentally interpreted (or 'heard'). Relaxing the constraint of inhibition will result in a much more complex pattern of activation in the abstract level F2, possibly making it uniterpretable. However, looking at the incredible complexity of real patterns of activation in the brain, we can believe that we will be closer to reality without the tidiness afforded by inhibition at the abstract level, which guarantees the response of only one category. This is taking Gjerdingen's argument one step further, given that he does use lateral inhibition but justifies the use of self-organizing networks because of their untidiness in contrast to the tidiness of music theory or even of other classes of ANNs: "the structures of even so fastidious a composer as Mozart appear as dense tangles of contrapuntal lines, thorough-bass patterns, harmonic and melodic schemata, metrical frameworks, rhythmic gestures, and a host of more ineffable features. [...] These networks are capable of independently arriving at untidy but nevertheless quite interesting categorizations of musical events." (p.340).

2.4.3.2 Winning and learning

The second important simplification concerns the learning rule. Even though no winner is chosen when the output activation is recorded, it is probably not desirable to have many categories simultaneously learning the same input pattern. That would probably result in unstable learning and in the proliferation of categories beyond several thousands. Therefore ARTIST's learning followed the ART specifications in that only one category is updated for a given input.

The ART algorithm (detailed in Section 2.3.3) goes through a search process to determine which category is the winner and should consequently undergo learning by modification of the synaptic weights. The search process is such that the F2 nodes (categories) should be checked for matching with the input in the order of decreasing *absolute* activation, until one satisfying (matching) category is found. Which means that in some cases, it is not the best fitting category that is chosen as the winner, but only a 'good enough' category with high absolute activation. I thought that learning could be more stable if the winner was always the best fitting category. This also has the advantage of reducing absolute activation and matching degree to a single idea, corresponding to the concept of matching activation, thus making the functioning of the network easier to understand. From now on, matching activation $\frac{|I \wedge w|}{|I|}$ will simply be referred to as activation. Moreover, the sequential aspect of the search process proved to be a hindrance during the first simulations. Dealing with less than thirty abstract categories, L'art pour l'art never ran into this problem, but ARTIST was getting slower and slower to find a winner as the number of categories reached several hundreds. So the implementation of the serial search for a winner was abandoned and replaced by the decision rule that the best matching category is the winner. The rest of the learning rule remained unchanged, that is, synaptic weights were modified if there was resonance, and a new node was created if there was not. Note that even though the idea of a serial search may seem unrealistic, a system as complex as the brain could probably implement quite easily something equivalent, like a parallel search based on both absolute activation and matching degree.

2.4.3.3 Other options

Carpenter, Grossberg and Rosen (1991) made a few suggestions to accelerate and stabilize learning in ART networks. Two of them were implemented in ARTIST.

Fast-commit slow-recode option: When a category commits itself by being activated for the first time, the learning rate is set equal to $\beta = 1$. This allows a kind of 'one-shot learning'. That is, the weights of the new category are set to match *exactly* the input that triggered its activation. Only upon subsequent winnings of this category we use $\beta < 1$ to update the synaptic weights, thus only allowing the category to get closer to matching the input patern, but not allowing a perfect learning.

Input normalization option:

In a complex environment, ART networks can encounter a problem of proliferation of categories due to 'synaptic erosion'. This happens when a succession of different inputs activate the same category for learning at different points in time. As a consequence of using the logical AND in the learning rule, the synaptic weights can only decrease from 1 and never increase. That is, the only way a category adapts to a new input pattern is by removing the features of the category which are irrelevant to the new input. Thus, it happens that an unfortunate coincidence of different input patterns activating the same category erode the synaptic weights down to 0 or close. This category then falls into oblivion and the input patterns it once recognized are now categorized by younger, newly committed categories. Plasticity is desirable up to some point, but the perpetual unstability of such a classification dynamics is not. Especially for the computational aspect of the simulations, having thousands of categories with only a small proportion of them useful is a hindrance.

Carpenter et al. (1991) propose several solutions to this problem. For instance, this can be solved by normalizing all input vectors. Thus the activation in F1 was normalized each time a new input pattern was presented and each time the top-down activation was propagated from F2 to F1. Normalizing inputs may in turn create another problem, because it suppresses the relative differences between amplitudes of different patterns. It is possible to keep the amplitude information while normalizing inputs, by doubling the number of input units and using half of them to code the complement of the input. This solution was not implemented because it must be computationally very demanding to use 144 input units instead of 72 (the number of synapses roughly goes from 50,000 to over 100,000). Rather, the synaptic weights incoming each category were also normalized, transforming the phenomenon of erosion into a mere reshaping of the categories.

2.4.4 ARTIST meets its environment

Stimuli

A musical corpus is needed to train ARTIST. It defines ARTIST's musical universe. Musical Instruments Digital Interface (MIDI) files were downloaded from a web-site ("The classical midi connection, http://www.dtx.net/ raborn/"). In spite of the lack of originality of this choice, Bach's 24 preludes of the Well-Tempered clavier were tried first to train the model because they can be considered as a standard *de facto* for this type of application. The reason for that is probably that the prototypicality, the inner regularities of each piece and the complementarity of the different pieces are the best guarantee of the homogeneity of the training set (there is one prelude and one fugue in each and every major and minor key). A Pascal program transformed the MIDI files into an ARTIST input file by extracting only the note events. One note event is comprised of a note and its duration and velocity, so as to make it directly usable by ARTIST. That is, all messages relating to timbre changes or other controller changes for instance, irrelevant for the model, were ignored. The same process was applied to the MIDI files used as stimuli for the experiments of Chapter 7. One little modification was brought to these stimuli. The notes they contained were all of equal velocity, which does not strictly conform to the way this music is supposed to be played. The performers are usually supposed to accentuate the first note of every measure (and the note immediately following the middle point of the measure) to convey a sense of meter (the regularity of the rhythm) to the listener. To conform to this, the velocity of the first note of every time slice was increased by a fixed amount.

The short stimuli containing only a few notes, such as the short test melodies (e.g., 'Twinkle twinkle..', Chapter 3) or the musical scales used as contexts (Chapters 4 and 5) were programmed with MATLAB and directly used by ARTIST, itself programmed with MATLAB.

Procedure

The model was trained with Bach's 24 preludes from the Well-Tempered clavier, so there was one piece in every major and minor key (C major, C minor, C# major, etc..). Each piece was presented to the model 12 times (in 12 different transpositions), so that the tonic covered all the chromas of the middle octave. The $24 \times 12 = 288$ presentations were given to the model in a random order. To be fed into the network, a piece was decomposed into time slices of equal length, one half measure in this case. The presentation of one slice consisted of the following four steps:

1) combining in F1 the residual input activation left over from before, new inputs and top-down activation propagated from the categories' winner in F2. This takes into account the activation decay over time. The result is normalized.

2) propagating (computing) the activation to the abstract level F2

3) recording F2 pattern of activation if we are interested in the output at this point

4) choosing the abstract winner

5) updating winner's synaptic weights if in the learning mode

The 288 presentations resulted in the learning of $41,004$ time slices containing a total of $218,340$ notes, creating 709 categories with the vigilance value set at 0.55 . It is the network in this final state that will be used for most simulations, unless noted otherwise. After learning with a more stringent matching criterion of 0.7, the model reached a similar architecture with the creation of 787 categories.

Before any learning occurred, no abstract node is committed, so no category has been defined. Upon beginning of learning, the architecture of the network undergoes big changes. On the average, more than 16 categories are created for each piece presented over the 9 first presentations ($146/9 = 16.2$). Indeed, all these inputs are new to ARTIST and they are occurring for the first time. After many pieces have been learned by the network, learning is quite stable. That is, new pieces are readily interpretable with the mental schemata already created. Figure 2.4 shows that after about 100 pieces have been learned, less than two categories on the average needed to be added to the existing ones in order to understand a whole new piece (precisely, 349 categories were created after piece #100, for an average of $349/188 = 1.86$ new categories per piece).

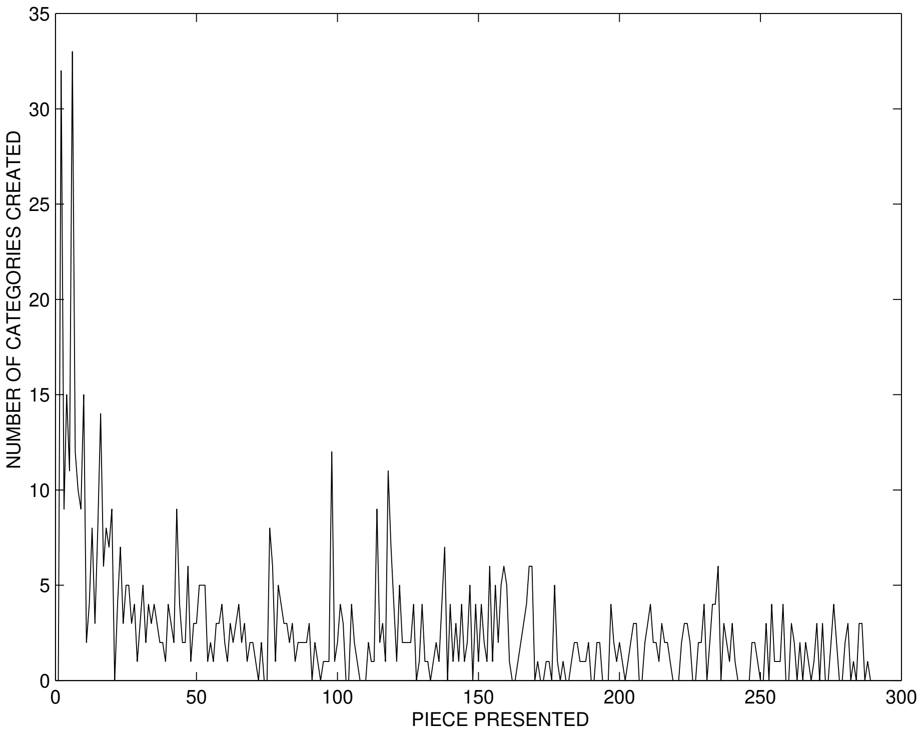


Figure 2.4: New categories created as a function of piece presented.

At this point, ARTIST is quite familiar with his musical environment and it can be tested on some musical tasks like an adult. The following chapter presents a first simple test of ARTIST, to see how its memory for very familiar tunes operate, to observe the role of top-down activation, and mostly to acquaint the reader with the basic functioning of the model.

CHAPTER 3

SIMULATION 1: RECOGNITION OF INTERLEAVED MELODIES

Dowling (1973) showed that when distractor notes are interleaved with the notes of a very familiar melody, the only way to recognize the melody and name it is to know before hand that it comes from a very limited set of possibilities. This is true as long as the distractors blend with the original notes, but not if two auditory streams are spontaneously formed, whether on the basis of frequency range, timbre, loudness or any other physical cue (Bregman, 1990) or even cognitive cue such as tonality. In other words, in the absence of salient perceptual or cognitive cues to segregate distractor notes from melody notes, the subjects need to explicitly test hypotheses regarding the identity of the melody. They need to know that it will be 'Old McDonald' or 'Twinkle twinkle' to be able to direct precisely their auditory attention and pick out the actual notes of the melody. As the number of alternative increases, the likelihood of recognizing the tune decreases. If subjects know that the melody is very familiar but are not provided with the explicit possibilities, they are very unlikely to recognize it.

This lead to the conclusion that expectancies play a crucial role for directing auditory attention: When the listener is familiar with a melody, he knows which note to expect, and when. 'Which note' refers to the pitch dimension, and 'when' refers to the time dimension. Therefore, an expectancy can be defined as the focus of auditory attention on a particular pitch-time window (Dowling 1987, 1990). Since ARTIST has a built-in top-down mechanism, it should be able to pass the interleaved melody task, and recognize a very familiar melody even when distractors are intermingled.

Procedure

The prerequisite to be tested with this task is to be very familiar with a melody. To impose this on the model, it was sufficient to run the model (already trained with Bach's preludes) in the learning mode with the learning parameters close to their maxima (vigilance and learning rate equal to .9 and 1, respectively) and present the soon-to-be-very-familiar melody. This ensured a fast and perfect exemplar learning, because the high vigilance forced the creation of a new node especially devoted to the memory encoding of 'Twinkle, Twinkle'. Had the vigilance parameter been lower, an already existing node may have resonated with the presentation of 'Twinkle, Twinkle', and the prototype for this category would probably have been substantially different from 'Twinkle, Twinkle' because of previous influences. As a consequence, the melody may not have been perfectly learned.

Learning the tune resulted in the creation of 2 new categories, named '710' and '711' because they were created after the 709 others. The nodes created at the abstract level of the network then act as a label for this particular tune. Two kinds of responses were recorded from the network: The activations of the two label nodes and the ranks of those activations compared to the activations of all the other astract nodes in F2. Both activations were summed, and so were the two ranks. If we assume a decision rule based on the activation levels of the nodes, the higher the summed activations, the more likely ARTIST is to respond that 'Twinkle' was recognized. Activation and rank vary in opposite direction, because the most activated node has the lowest rank, which is 1. So if we assume a decision rule based on the ranks of the nodes, the lower the summed ranks, the more likely ARTIST is to respond that 'Twinkle' was recognized.

The recognition involved three conditions plus a control condition to know whether the model can use its top-down knowledge the way we do. Condition one measures the model's likelihood to recognize the hidden melody by simply presenting the melody interleaved with distractors.

Condition two corresponds to the case where the human subject has some a priori knowledge of the melody that may be presented. The same stimulus as condition one is presented, but the label node is also given a high activation (equal to 1) to allow a top-down flow of activation to propagate and converge with the top-down activation. The activation of the label node is then reset to zero so that what is recorded is not due to the hypothesis tested but to a real match between the hidden melody and the hypothesized one.

A control condition needs to be added to rule out the possibility that the testing of a hypothesis is not always followed by a confirmation but only when there is a real match. We do not want the model to validate the 'Twinkle..' hypothesis when the Prelude #1 is being played. Therefore the response of the model is recorded after presentation of the Prelude #1 interleaved with the same distractor notes, and with the 'Twinkle nodes' activated during presentation. The Prelude #1 was chosen for this control condition because it is in the same key as 'Twinkle' and contains three out of the four pitch classes found in 'Twinkle'.

The last condition had the model faced with 4 hypotheses to test, in order to simulate the task where human subjects are given a list of 4 possible tunes containing the one to be recognized among the distractors. To realize 4 different hypothesis testing simultaneously, 4 different nodes were activated and simultaneously propagated their top-down activation. One of these 4 nodes was 710 or 711 (for first and second measure, respectively), whereas the three others were randomly drawn. To emulate the performance of human subjects, the model would have to show poorer recognition in this condition than in the case where only one hypothesis is being tested.

Results

The activations of the label nodes following the presentation of the original melody were both .86. Those were the highest and second highest activations in F2. So the totals for the activations and the ranks were $1.72(= .86 \times 2)$ and $3(= 1 + 2$, the lowest possible score), respectively. The summed activations and summed ranks for all the conditions are shown in Figure 3.1. They are the highest and lowest across all conditions, indicating that the original version is the most likely to be recognized.

With no hypothesis testing (no forcing of top-down activation), the presentation of the interleaved version elicits less activation than does the original version, and the ranks of those activations total 25 instead of 3.

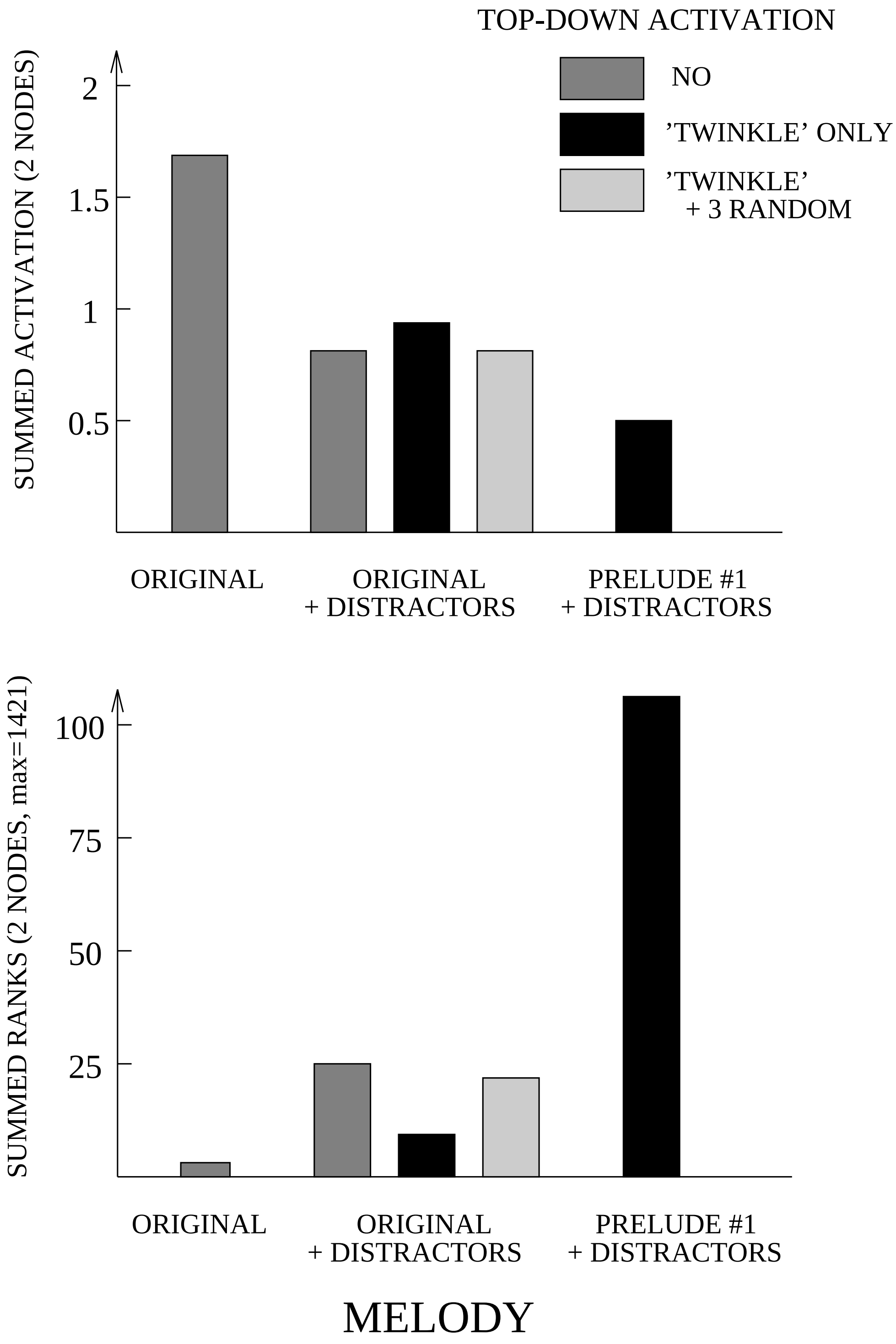


Figure 3.1: Summed activations (top) and activation ranks (bottom) of the 2 label nodes for 'Twinkle twinkle', as a function of stimulus played and hypothesis tested.

Therefore the probability for ARTIST to respond that 'Twinkle' was recognized is much lower than in the previous case, just as human subjects who were not given any hint were unable to recognize the hidden melody.

However, propagating top-down activation from 710 and 711 while the tune is being presented results in a higher sum of activations and lower sum of ranks than for the previous condition. This indicates a higher probability to recognize the tune than in the previous condition, similarly to the situation where humans are informed of the possible identity of the tune. Nevertheless, this probability is not as high as in the 'no distractor' condition.

When the first notes of the Prelude #1 are presented to the model, interleaved with the same distractors as above, and top-down activation from the 'Twinkle' nodes is forced, there is no consistent false alarm recognition of 'Twinkle', if we consider that the sum of activations is the lowest and that the sum of ranks is highest across all conditions.

For multi-hypothesis testing, the three random nodes 164, 430 and 674 are also used in the top-down process. The results are comparable to those obtained without any use of top-down knowledge, exactly like for human subjects.

Conclusion

These simulations revealed that ARTIST can memorize some particular tunes (very short for now), and that this knowledge can be used to recognize a very familiar melody among distractors. As with human subjects, the recognition performance drops as the number of possibilities concerning the hidden melodies increases, to approximately the same level as when no hint regarding the identity of the tune is provided. Those results are the same whether we assume a decision rule based on the absolute activations of the label categories or based on how many other categories have a greater activation than they do.

Figure 3.2 shows the input activations for the first measure of 'Twinkle, Twinkle...', without and with distractors, respectively. Without the distractors and after the tune has been learned, the inputs before and after top-down propagation are identical because the abstract node is a perfect replica of the input (tune is perfectly familiar). Middle panel of Figure 3.2 contains all the notes of the top panel, plus new ones (the distractors).

The flow of top-down activation is what enables ARTIST to pick the relevant notes out of the apparently meaningless string of notes in order to recognize the hidden melody. The activation from the top projects to the input nodes corresponding to the notes in the melody, and reinforces their activations compared to the distractors' activations. It does not explicitly inhibit distractors' activations, even though the latter is reduced after the input is renormalized. Figure 3.2 (bottom) shows that irrelevant notes are filtered out after top-down activation has spread, and they have little residual activation.

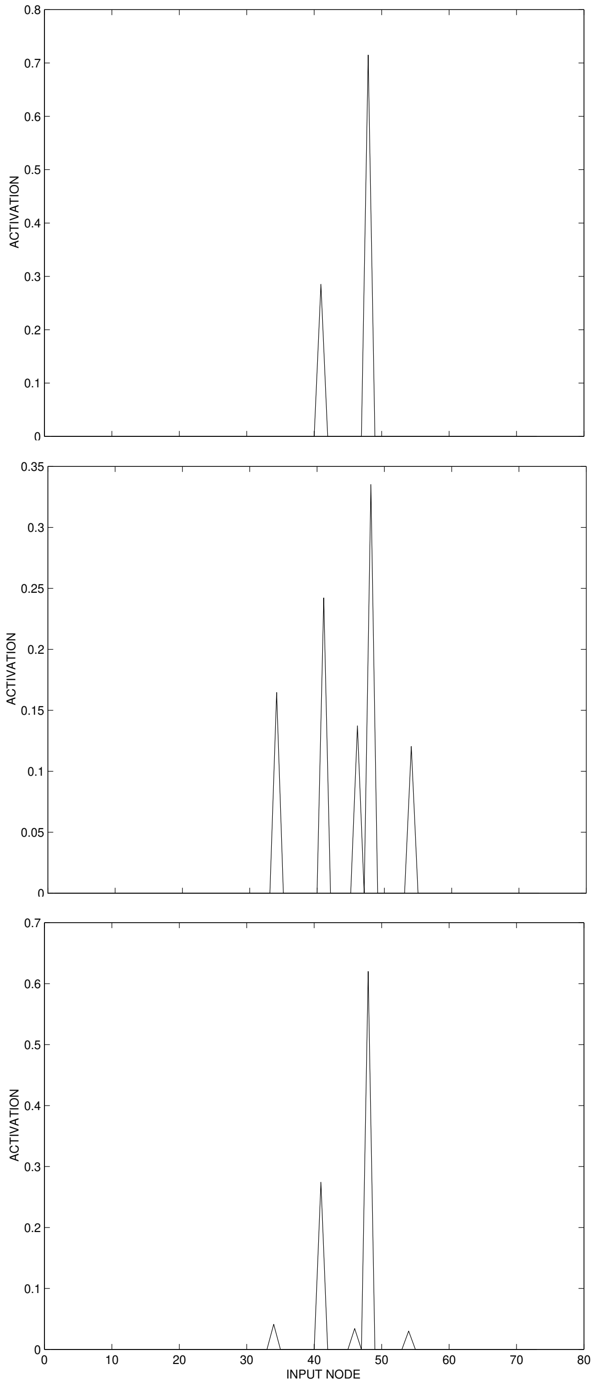


Figure 3.2: Input activations for the first measure of 'Twinkle, twinkle' (notes: C-C-G-G) without distractors before or after top-down (top; this is identical to a plot of the category (i.e., synaptic weights of the label node) because there was perfect exemplar

learning), with distractors before top-down activation (middle) and with distractors after they have been filtered out by top-down activation (bottom).

The node from which top-down activation comes has to be synchronized with the hidden melody, since it would be useless to reinforce the right notes' activations at the wrong time: the forced winners were 710 for the first measure, and 711 for the second. In summary, the description of this process as being the focus of attention on pitch-time windows (Dowling, Lung and Herrbold, 1987) is perfectly appropriate to describe ARTIST's functioning.

CHAPTER 4

SIMULATION 2: INTERNALIZATION OF THE INVARIANTS OF TONALITY

4.1 The tone profiles

4.1.1 Krumhansl's contributions

The difficulties in clearly defining what tonality is were highlighted in the introduction. Fortunately, Krumhansl and Shepard (1979) provided us with the probe-tone technique, an empirical and reliable way of measuring tonality. The results born from this technique are now considered the cornerstone of music research on tonality, because they describe the most salient invariants of tonal music. The technique lead to a description of what listeners know about pitch relationships, in the context of traditional Western music. In other words, it characterizes the relationships between musical elements in a psychologically relevant fashion.

The outstanding contribution of Krumhansl is twofold. First, it lies in her building a direct bridge between the disciplines of music theory and psychology: starting from the simplest measure of perceived consonance (the straightforward rating by subjects of how good or bad a musical sequence sounds), she was able to rediscover and account for the main musical relationships known from music theory. Moreover, these results appear to be very robust, being very similar across all the combinations of tones and chords used as musical contexts. Second, she made musicological descriptions take the step from discrete to continuous, from qualitative to quantitative: it has always been known that the tonic is more stable than the fifth, but as we will see below, this stability relationship is now quantified. That is, each note is given a corresponding numerical value to reflect its stability in a given context, ranging from 1 to 7 on a continuous scale. This measure of stability will enable us to test the model's predictions in a more precise way than the qualitative descriptions of music theory would.

According to Krumhansl (1990), "this system [of pitch relationships] arises from stylistic regularities identifiable in the music" (p.9). The fact that the (mental representation of) pitch relationships uncovered by the probe-tone technique fit quite well with those found in the stimuli is a strong argument for this explanation. Therefore, Krumhansl's results being the best account to date of the invariants of tonality and their associated schemata, they constitute the best test one could give to a subject (or a model) to check if he has internalized the regularities of Western tonal music. Before trying the probe-tone technique on ARTIST, an overview of Krumhansl's related work is provided.

4.1.2 The probe-tone technique

To characterize how the 'Mind's ear' hears a note in a musical context, and how those are related, Krumhansl and Shepard (1979) and Krumhansl and Kessler (1982) established the probe tone technique. It consists of playing a short musical sequence to the subject, a strongly prototypical one in order to establish a strongly tonal context (either major or minor), followed by a probe tone. The subject is then asked to rate the probe tone on the basis of how well it completed the preceding context; in this case, the range of the rating was 1—7, from 'very bad' to 'very good'. Using all of the 12 possible pitches of Western music as probe tones enables the construction of a tone profile, the graph showing the 'goodness' ratings as a function of the note used as a probe. To prevent the results from depending on the possible idiosyncrasies of the key of C major, all the contexts were used several times, transposed to other keys. Data from the different keys were highly correlated and therefore were averaged for subsequent analyses.

To establish the tonality, different kinds of contexts were used to make sure that the findings were not due to the particular stimulus used to establish the tonality but rather to the induced tonality itself. Some contexts were used in only one of the two studies whereas some others were common to both.

Specifically, the contexts used in either experiments were: complete or incomplete ascending scales (e.g., C-D-E-F-G-A-B(-C) for C major), incomplete descending scales (e.g., C-B-A-G-F-E-D), chords (e.g., C-E-G played simultaneously) and cadences (i.e., sequences of three chords ending on the chord establishing the tonality, e.g. F major-G major-C major for the tonality of C major). All these were played in both major and minor modes and resulted in two tone profiles, one for each mode.

4.1.3 The major and minor key profiles

Figure 4.1 shows the two tone profiles following musical contexts respectively defining the keys of C Major and C minor. The tone profiles for the other major and minor keys are identical and can be inferred from the C profiles by transposition (shifting along the X-axis). The results of the probe-tone technique lead to the reconstitution of the hierarchy predicted by music theory, namely that the most stable tones are in decreasing order the tonic (C), the fifth (G) and the major third (E), followed by the other diatonic notes (D,F,A,B), and then the chromatic notes (F#,C#,G#,D#,A#). These groups of notes define the four levels of what is known as the *tonal hierarchy*. Figure 4.1 shows the name of all the notes next to their data points so that the levels of the tonal hierarchy could be visualized more easily, by reading the notes from the highest to the lowest point. The profiles thus found are very robust. That is, there are only minimal variations between profiles obtained with different types of contexts.

By establishing a context in C minor instead of C major, a similar tone profile is obtained. The main difference between the two profiles relates to the major third's vs minor third's stabilities, since the major third is only present in the major mode and the minor third only in the minor mode. A surprise comes from the minor key profile in that the minor third surpasses the fifth in stability, becoming the second most stable tone after the tonic! Intuitively, it could be so because the minor third being the trademark of the minor mode, it is used to emphasize this mode as opposed to the more common major mode, and thus may have become more closely associated with the minor mode than the fifth, which does not disambiguate between major and minor modes.

The other main difference between profiles concerns the stabilities of the augmented fifth and of the sixth (respectively G# and A). These can also be explained by their inclusion or not in the major vs minor scales. However, this does not mean that the differences are attributable to a kind of sensory priming: they were observed even when

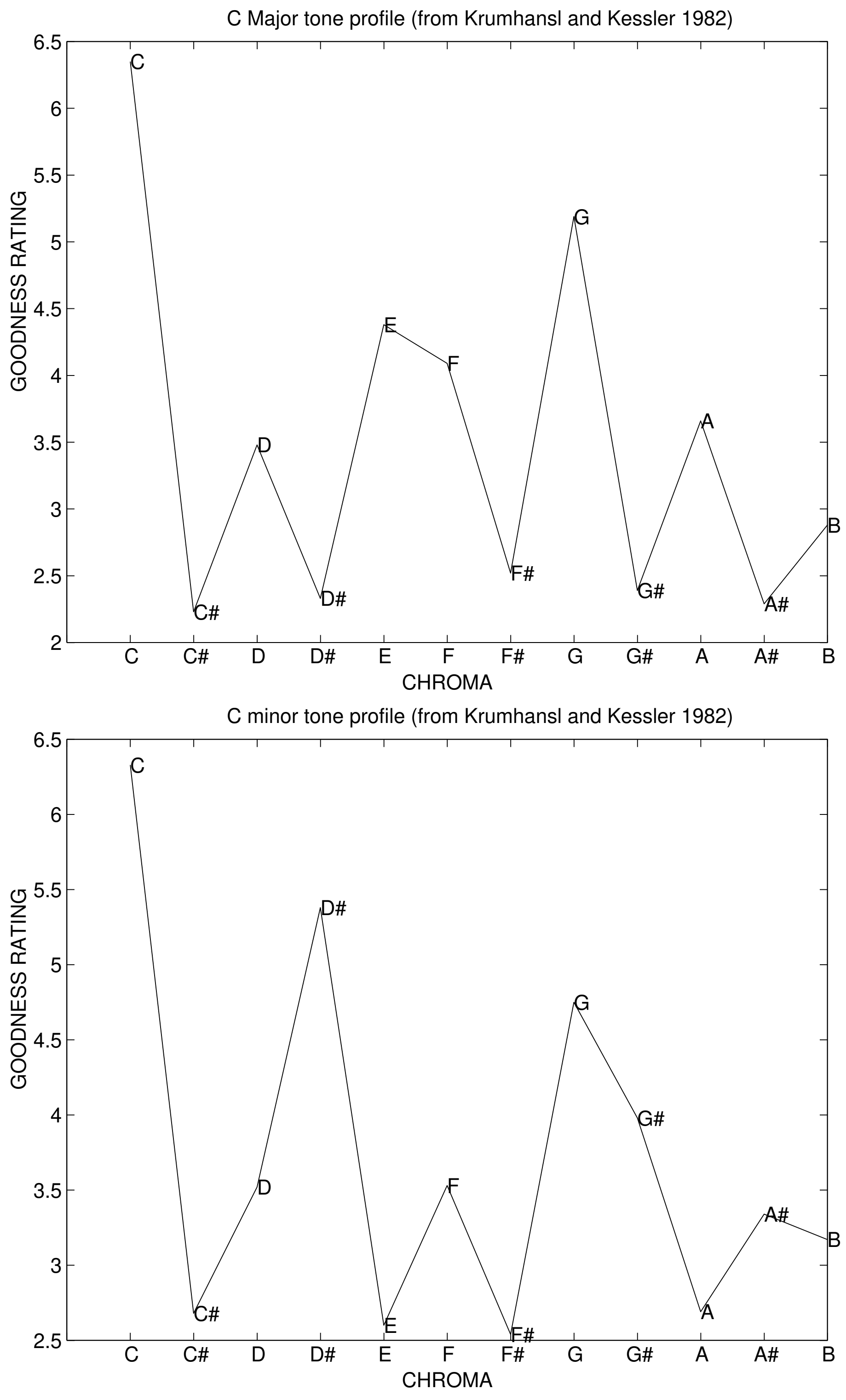


Figure 4.1: The C Major and C minor profiles: relative stabilities of pitches following major and minor key contexts.

the contexts were simple chords which do not include those notes. They have to be attributed to semantic priming. That is, even simple chords established a tonality, which in turn prepared all the scale notes for preferential processing.

A less fundamental result, nevertheless noteworthy, comes from another study attempting to characterize tonal stabilities, this time relating to the stability of a pair of notes (Krumhansl, 1979). The same methodology as before was used, with the exception that two probe tones instead of one were sounded after the context, and the subjects were asked to rate how well the second tone follows the first, given that particular context. This technique will also be used in the developmental study of the emergence of tonal schemata, reviewed in details in Chapter 6. The results obtained with two probes obviously correlated highly with the tone profiles for a single probe. The interesting finding is the perceptual asymmetry of the two probe tones, showing that the second (and final) tone has a stronger influence than the first one on the ratings. Such an asymmetry was expected, since it is well known that the temporal order of the notes in a melody is very important. However, even if these findings make intuitive sense because it is the last tone that will convey the sense of suspension or of resolution following the musical extract, they all can be attributed to the fact that subjects were explicitly asked to focus their judgment on the second tone; had the subjects been asked to rate how well the first tone precedes the second (as opposed to how well the second tone follows the first), or how the last interval fits in the context, the results might have been slightly different. In any case, this might be a phenomenon related to the asymmetry found for the order of presentation of tonal vs atonal stimuli, discussed in the introduction.

4.1.4 Distances between keys

Since the tone profile of a key seems to characterize all (or most) of its aspects of consonance, a measure of the psychological distance separating two keys can be derived from their respective tone profiles by correlation: the better the two profiles correlate, the closer the two keys should be perceived psychologically. Krumhansl (1990) computed the correlations between the tone profiles from Krumhansl and Kessler (1982) for all possible pairs of keys, including major and minor keys. This is shown in Figure 4.2. Not surprisingly, and in agreement with music theory, it was found that this way of measuring distances is almost monotonously related to the interkey distance measured around the circle of fifths: only a few data points break the monotony of the relationship, none of them coming from the correlations between two major keys.

Thus, those local maxima are found in major/minor or minor/minor keys correlations, and their similarities to the reference chord is more than can be accounted for by the circle of fifths. As Krumhansl (1990, p.39) notes, this is explained by the priviledged status of the parallel and relative major-minor relationships (e.g., C major's parallel minor is C minor, and its relative minor is A minor). The parallel and relative keys are priviledge in the sense that their chords both share two notes out of three with the chord of the key of reference (C major: C-E-G; C minor: C-Eb-G; A minor: A-C-E). In fact, the points breaking the monotony of the relationship was predictable considering the advantage of the minor third over the fifth in the minor key tone profiles, mentioned earlier (e.g., the C is very important in A minor, making it very close to C major).

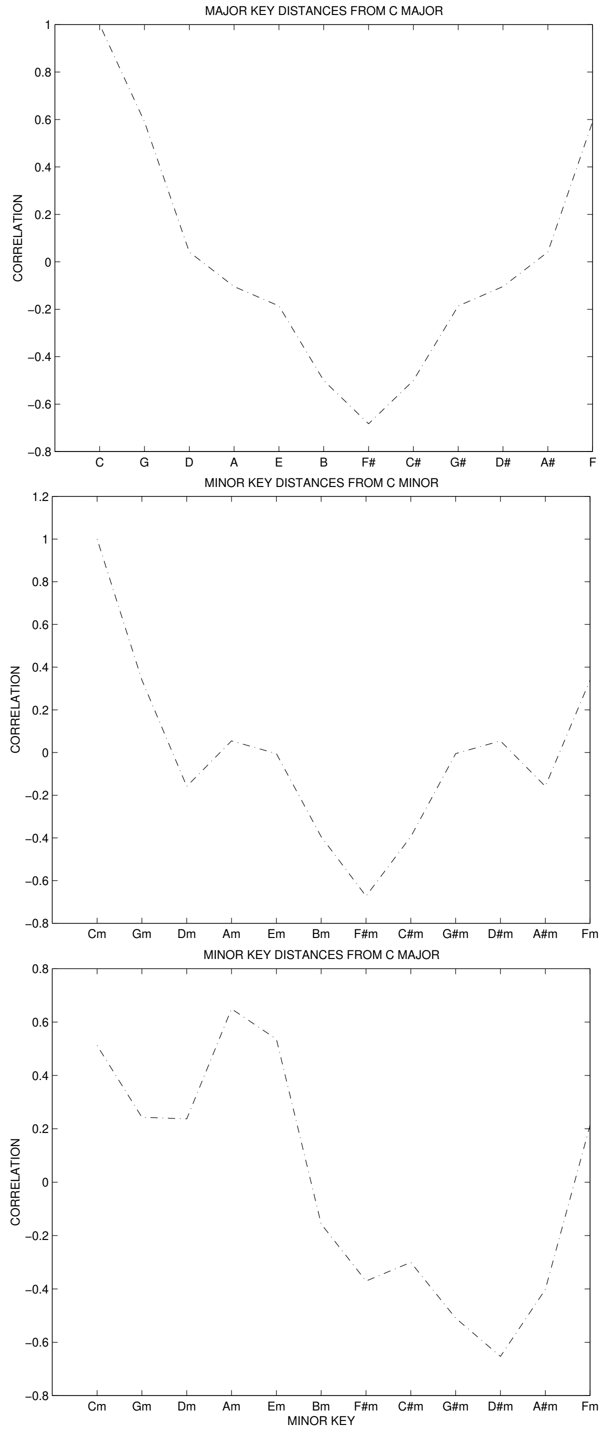


Figure 4.2: Major-major (top), minor-minor (middle) and major-minor (bottom) interkey distances, derived from the correlations between C major/C minor profile and all major/minor keys tone profiles (Krumhansl and Kessler, 1982).

Krumhansl and Kessler (1982) showed that a multidimensional scaling (MDS) analysis performed on all the correlations lead to a 4-dimensional solution that, when projected on 2-dimensional spaces, reconstitute perfectly the circle of fifths (dimensions 1 and 2) and other circles based on relative and parallel major-minor relationships (dimensions 3 and 4).

Using the same method, Krumhansl (1990) estimated interkey distances in a second way. It differs from the one just mentioned in the stimuli used for the collection of data. First, the musical sequences used to establish major and minor contexts were a little different, but the ratings have been shown earlier as being very consistent across different sequences establishing the same key. The significant difference between stimuli was the use of 'probe chords' instead of probe tones: the 12 probe tones following the context were replaced by the 24 chords built on the 12 tones as tonics, thus including all 12 major and 12 minor chords (48 chords built on the 12 tones, thus including all 12 major, minor, diminished and augmented chords).

This provided a direct measure of both major and minor *harmonic* (as opposed to tonal) hierarchies for each context mode (C major and C minor). The harmonic hierarchies pertaining to the other keys could easily be inferred after assuming transpositional invariance. As explained just before, the correlations between the harmonic hierarchies were computed for all possible pairs of keys, to derive a measure of interkey distances. A plot of these distances shows great similarity with the plot of interkey distances obtained earlier, even though more flattened, not as contrasted. Also, the

4-dimensional MDS solutions look almost identical to the MDS solution derived from tonal hierarchies, confirming the validity of the harmonic hierarchies.

4.1.5 Conclusion

Many studies have since replicated or extended Krumhansl's findings, sometimes with substantially different paradigms (e.g., measures of errors or of reaction time), confirming the cognitive reality of the tonal hierarchy (Jarvinen, 1995; Cuddy, 1993; Repp, 1996; Sloboda, 1985; Janata and Reisberg, 1988).

The probe tone technique is a tool to explore the cognitive representation activated at any time during a musical sequence. Tone profiles can serve as references in that they reflect the cognitive states after presentation of prototypes. There is evidence for the robustness of both the technique and the profiles, coming from replications of the results and from convergence with other paradigms and concepts of music theory. Therefore it seems appropriate to use those tools as a basis for the evaluation of how well a system emulates humans' representation of tonality. Specifically, the proposed model's second test will be to reproduce the tone profiles when subject to the probe tone technique. Indeed, given the diversity of musical issues relating directly or indirectly to the tone profiles, there would be little hope from a model that could not come close to exhibiting them.

4.2 Simulation 2: ARTIST and the probe tone technique

All we need to submit ARTIST to the probe-tone technique is to present different contexts followed by every pitch, and record its answer of how good the sequence sounds. The only problem is that ARTIST does not have a defined output. It was never taught to give a 'sounds good' or 'sounds bad' judgment in response to a musical sequence. All it knows is to recognize and classify musical patterns through the activation of nodes, that represent abstract categories formed through passive learning.

If the idea presented in the introduction that familiarity determines what is liked is valid, then measuring ARTIST's degree of familiarity with a musical sequence should give us its rating of how good or bad the sequence sounds. Along the same lines, Katz (1995) proposed a theory of positive affect, based on the old principle of 'unity in diversity'. It is argued that 'unity in diversity' is perceived when two usually mutually exclusive principles are simultaneously realized. This translates directly at the level of neuronal activations: it will occurr when two mutually inhibiting neurons are simultaneously active. This is not an impossible situation because of the delay involved between one neuron becoming active and the other one being inhibited. During this delay both neurons can be active to some degree and their summed activations can momentarily be boosted. This principle was applied to a fairly simple connectionist model and did account for the aesthetic effects of many musical figures. It is also mentioned that Matindale (1988) used this measure to explain preferences for familiarity and prototypicality.

Following these reasonings, the total activation present in the network at the F2 level will be taken as an index of familiarity and of aesthetic judgment.

Procedure

Using all 16 different types of contexts used in Krumhansl and Shepard (1979) and Krumhansl and Kessler (1982) with ARTIST would be computationally tractable but quite fastidious. Moreover, most of them yield very similar results, whereas some others lead to significantly different profiles and were not included in the final analysis. So only the most simple and prototypical contexts were retained to test ARTIST: for each mode, the corresponding chord and the ascending and descending scales were used.

Moreover, chords and scales are complementary in the sense that they instantiate orthogonally the two dimensional aspects of music: verticality and horizontality, repectively. The vertical dimension of music refers to harmony, the pitch relationships of notes played simultaneously. The horizontal dimension refers to the time dimension. This is literally the way music is notated on a score. In general, music makes use of both dimensions: it consists of patterns of notes, some of them played simultaneously, unfolding in time. Thus a chord can be considered a purely vertical musical stimulus, because it consists of three notes played together and involves as little as possible of the time dimension. That is, only the duration of the notes makes use of the time dimension, but not their starting times, all identical. In contrast, a scale is primarily a horizontal stimulus: it is a sequence of notes unfolding in time, without two notes ever being played simultaneously.

After ARTIST completed learning as described in Section 2.4.4, each of the three context was used 12 times (vs 4 times with human subjects in Krumhansl and Kessler 1982) to allow the tonic to take any pitch class value. The 12 tone profiles obtained for each context were then averaged. Using different pitch classes as tonics ensures that the results (ARTIST's as well as humans' results) are not due to the choice of a particular pitch of reference. There seems to be no particular reason for ARTIST's 12 tone profiles to be very different from each other, since ARTIST was trained with the same pieces transposed in all 12 keys. However the categories developed differently on different pitch levels, due to the randomness of the order of exposure to pieces and keys. Therefore the profile for one given context was obtained by averaging ratings over the 12 pitch classes as tonics. Presenting one context on one given tonic involved 12 trials, one for each probe tone. Thus the major key profile was obtained after recording ARTIST's rating on 432 trials (3 contexts $\times$ 12 tonics $\times$ 12 probes), and the minor key profile required 432 more trials.

A major improvement Krumhansl and Kessler (1982) brought after the first probe-tone study by Krumhansl and Shepard (1979) was the use of Shepard tones (1964). The first study found that as the subject's musical experience increases, he relies more on the pitch itself of the probe tone (and its relationship to the established key) to give a judgment, rather than just relying on the size of the interval like beginners do. In order to isolate the effect of pitch class on musical judgment, and to obtain more consistent results between experienced and inexperienced listeners, the second study used complex tones that have a definite pitch class but no definite height. For instance, a Shepard C definitely shares the same quality common to all Cs, but one could not tell the octave it belongs to, if the note is C1, C2, C3, etc... These tones were used to give the illusion of ever-ascending scale (Shepard, 1964). They are simply made of frequency components from all the Cs in different octaves. It would be very much like playing all the Cs of the keyboard at the same time, C0 with C1 with C2 and so on...

This is exactly how we can simulate the presentation of Shepard tones to ARTIST, by playing the pitch presented on all octaves simultaneously. An amplitude envelope similar the one used in Krumhansl and Kessler (1982) was imposed over the frequency range: amplitude is highest for the middle octaves and decreases for octaves close to both ends of the frequency spectrum. Figure 4.3 shows how the notes C-D-E-F are coded on the input layer when presented as Shepard tones.

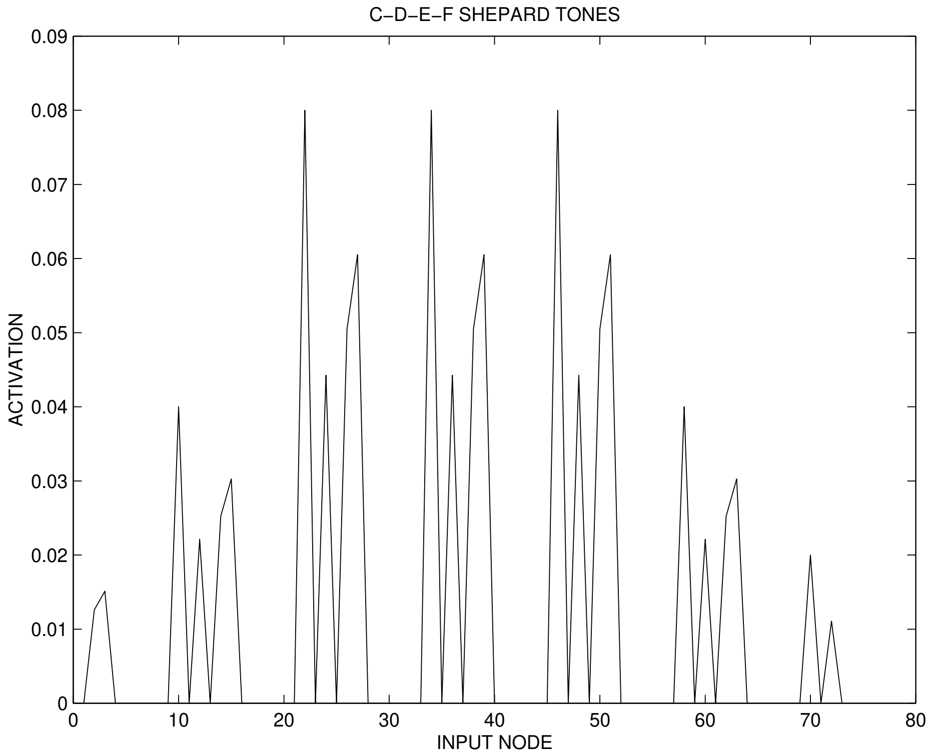


Figure 4.3: Input activations after presentation of 4 Shepard tones made of 6 harmonics each.

Results and discussion

To measure how well ARTIST's tone profile fit Krumhansl and Kessler's (1982), the Pearson Product Moment Correlation Coefficient between the two was computed, for both major and minor modes. Both were significant, respectively $-.95$ and $-.91$, $p < .01$ (2-tail). Surprisingly, the correlations were negative. ARTIST's profiles are shown by the solid lines in Figure 4.4 but were inverted for easier comparison with the reference profiles (dashed lines), hence the negative values of activations.

This means that Katz' (1995) theory of aesthetic judgment based on 'unity in diversity' does not strictly apply in this case. In fact, it did not even apply to Krumhansl and Kessler's (1982) human subjects, who gave high ratings to stimuli generating a sense of unity (when the probe tone belongs to the key established by the context) and low ratings to those generating a sense of diversity (when the probe tone does not belong to the key). The stimuli were probably too short and too simple to afford any sense of unity in diversity. However Katz' measure might have given way to a positive correlation between profiles if lateral inhibition had been implemented in ARTIST, because inhibition was one of the premises of Katz' reasoning.

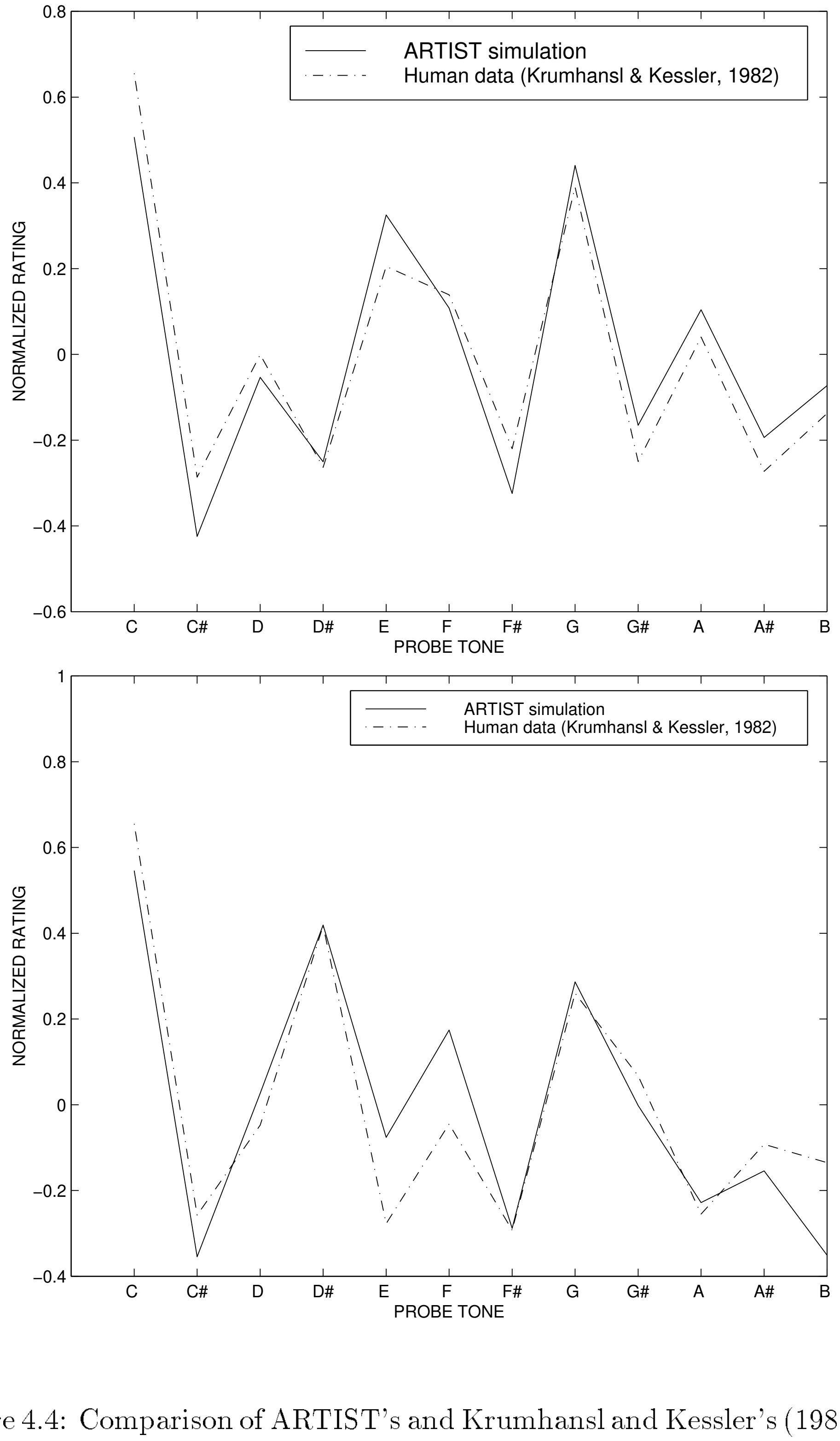


Figure 4.4: Comparison of ARTIST's and Krumhansl and Kessler's (1982) tone profiles, for major keys (top, correlation = .95) and minor keys (bottom, correlation = .91).

Thus the previous results may be better understood in terms of cognitive economy. The familiar or prototypical musical patterns are closely associated with only a few categories, and they have well-defined target nodes, the activation of which results in the sense of unity. In contrast, very unusual patterns will only be understood as a complex combination of categories, resulting in the activation of many abstract nodes and provoking a sense of diversity. For instance, the C major scale followed by the probe-tone C establishes unambiguously the key of C major and activates mostly the nodes relevant to this tonality. But if the probe-tone is F#, nodes relevant to both keys of C major *and* G major will become active because all 7 notes of both keys have been played. Thus, stimuli conforming to familiar patterns minimize the total activation of the categories, whereas those deviating from what is familiar increase the total activation.

Now ARTIST's C major and C minor profiles are established, we can deduce the profiles of any key by transposition (shifting the graph along the X-axis). Following Krumhansl's method of computing the correlation between tone profiles of different keys, we can infer the inter-key distances implied by ARTIST's schemata. The key distances from C major and C minor are graphed at the bottom of Figures 4.5 to 4.7. They are very similar to those obtained by Krumhansl (Top of Figures 4.5 to 4.7) even though some local maxima are not as well defined (e.g., around Am and D#m in Figure 4.6 and around C#m in Figure 4.7). The correlation between ARTIST's and Krumhansl's inter-key distances profiles were highly significant, between .97 and .99, $p < .01$.

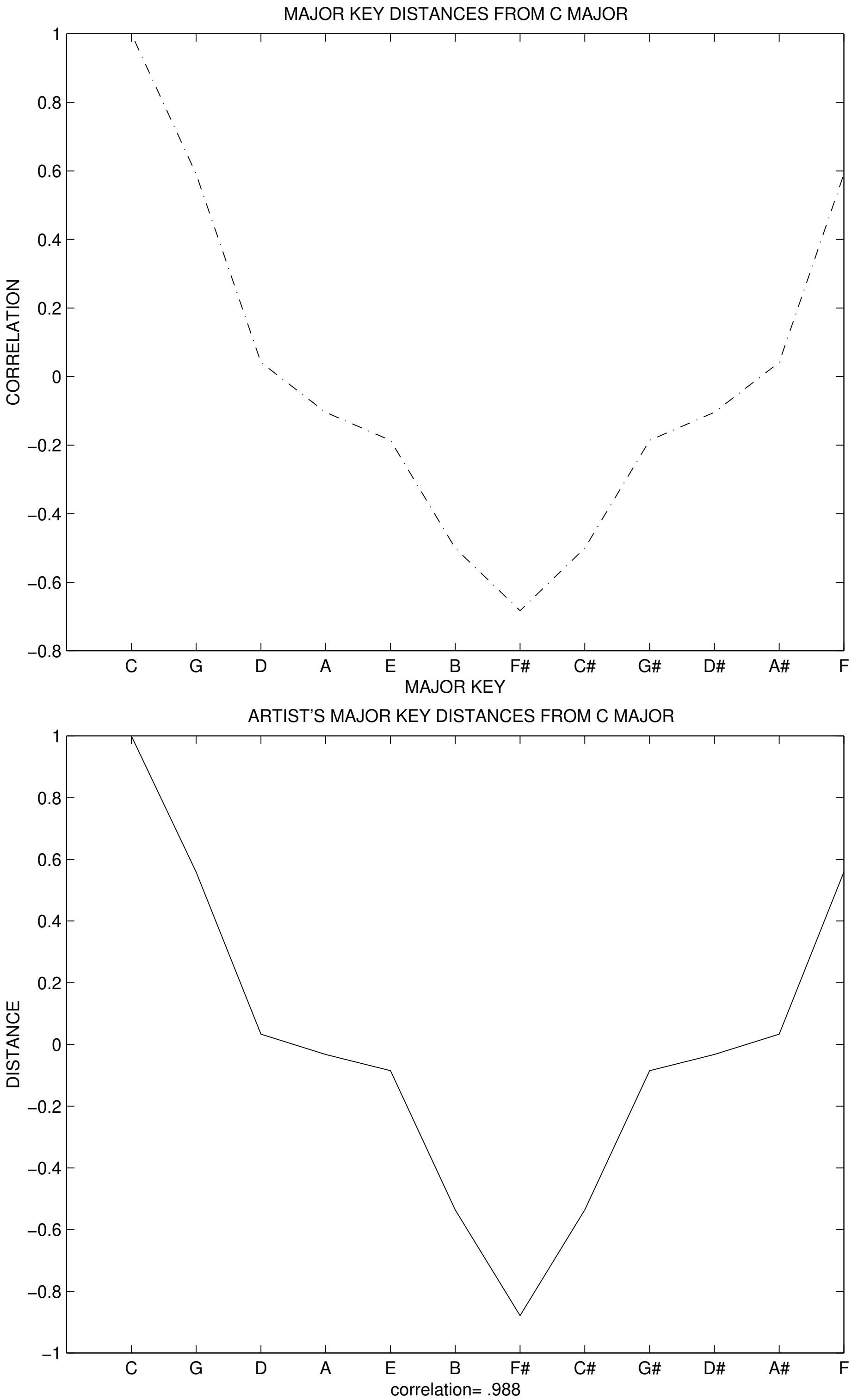


Figure 4.5: ARTIST's major-major interkey distances (bottom) and interkey distances obtained with human data (top) from Krumhansl (1990).

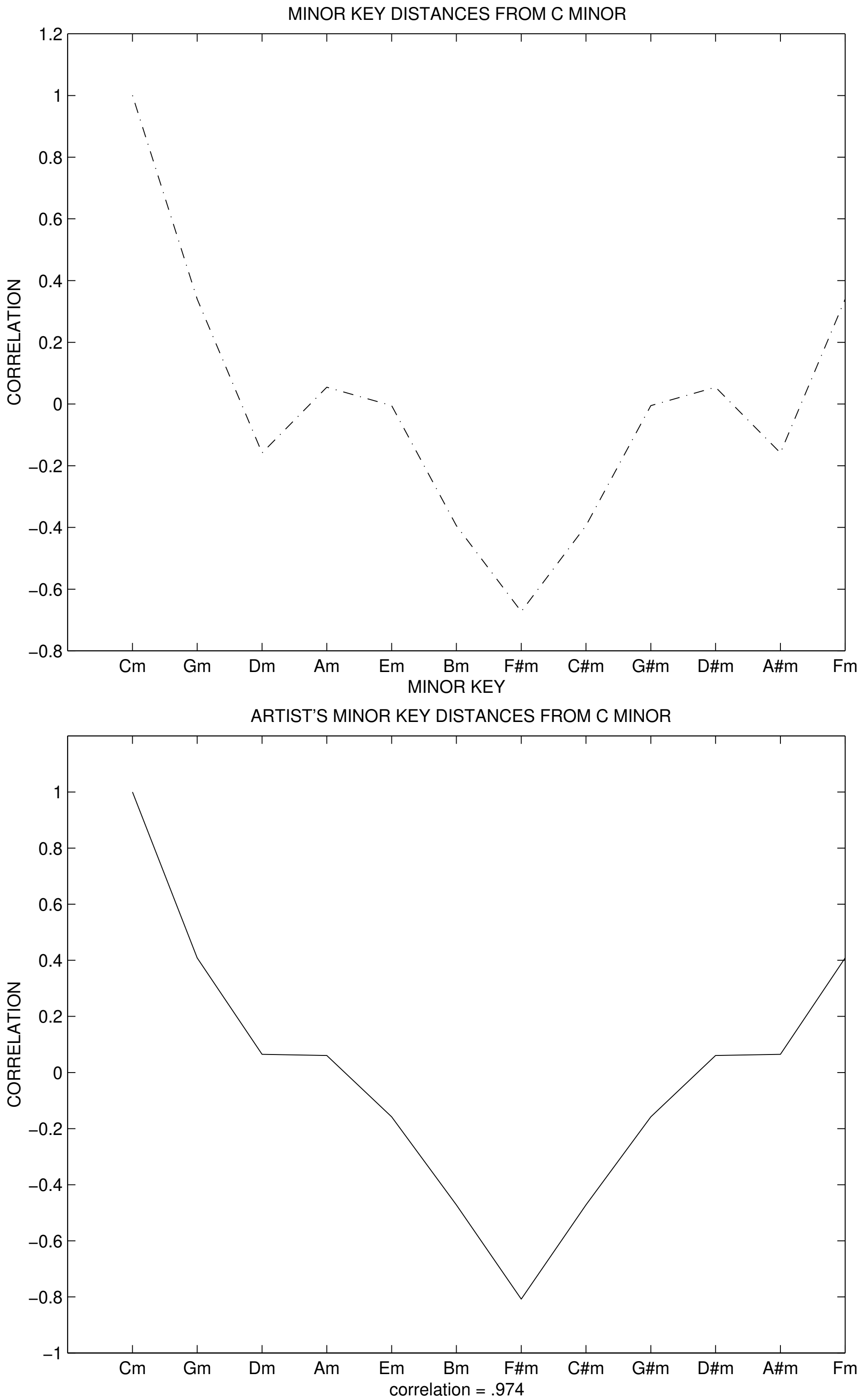


Figure 4.6: ARTIST's minor-minor interkey distances (bottom) and interkey distances obtained with human data (top) from Krumhansl (1990).

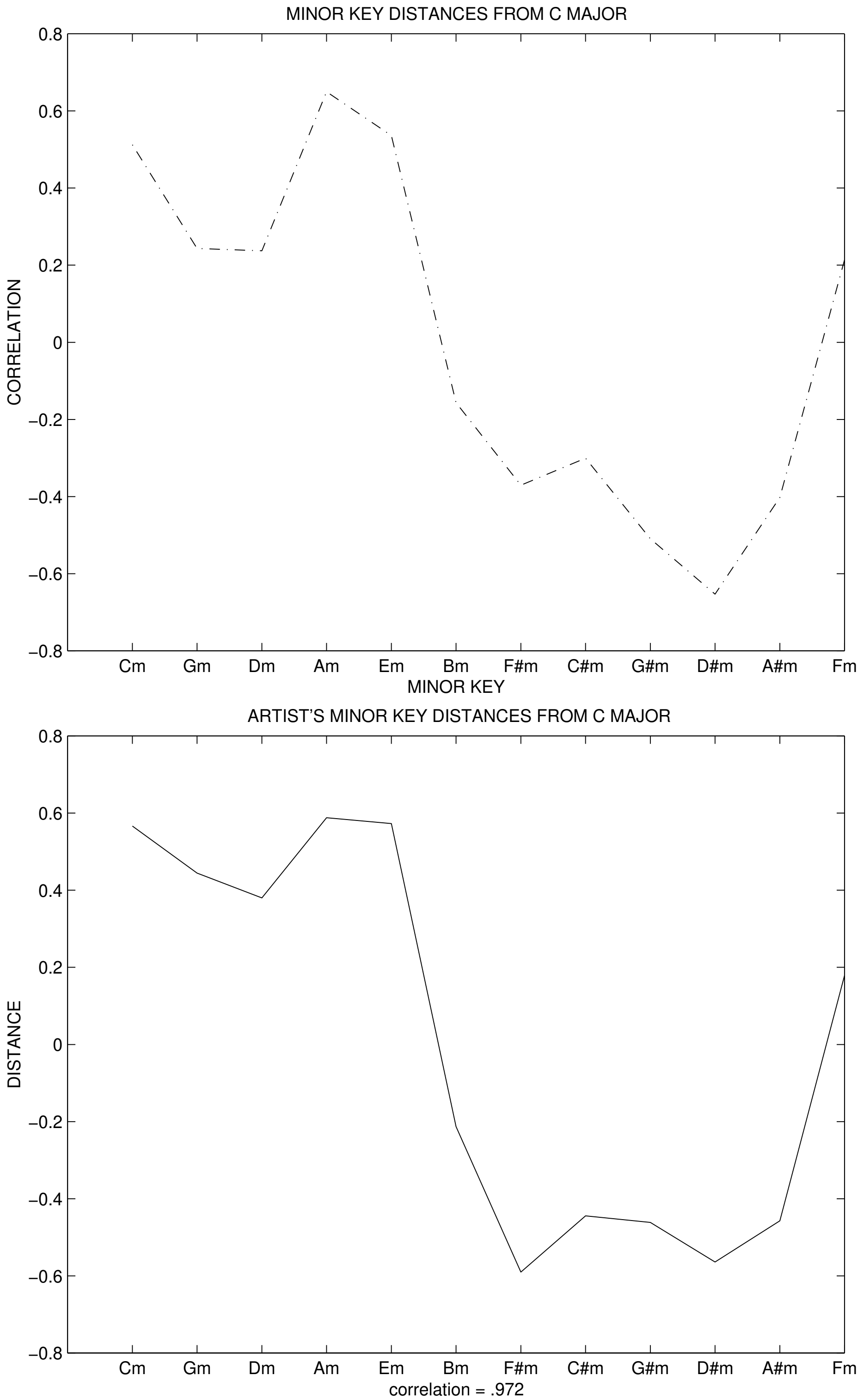


Figure 4.7: ARTIST's major-minor interkey distances (bottom) and interkey distances obtained with human data (top) from Krumhansl (1990).

Conclusion

The simulations carried out in this chapter show that the tone- and key-distances profiles can be internalized from simple exposure to music. ARTIST was able to extract the tonal invariants from its environment, even without any bias towards octave equivalence or similar acoustical relationships (e.g., 3:2 frequency ratio between fifth and tonic). The schemata formed are amazingly close to humans', especially if we consider that its musical universe contains only 24 different pieces (but each piece was 'heard' 12 times, in different keys) and that it would take less than 12 hours for a human to listen the 288 pieces!

These results seem to be very robust. Chapter 6 explores how different levels of exposure to music affect ARTIST's responses on a similar task involving two probe tones. For this, some random subsets of the complete corpus of 288 pieces were used to train ARTIST. They included 24, 48 and 144 pieces, and were not balanced regarding to the number of pieces in each key. Still, the pattern of results on the two-probe tone task were consistent. Regarding the vigilance parameter, it was mentioned in Chapter 2 that increasing it from .55 to .7 led to a similar architecture. Therefore it seems that mid-range vigilance values should not affect severely the present results, even though this was not explicitely tested. The learning rate was also manipulated in a few instances and the learning converged in the same way each time, but again, the responses to the probe-tone task were not compared. Section 5.3 explores the robustness of the model's behavior when trained with important variations in the coding of the input.

CHAPTER 5

A MARKOV MODEL IN ARTIST'S SHOES

5.1 Can low-level information in the environment explain the tone profiles?

The two previous chapters demonstrated how a connectionist model can internalize some invariants present in the music. The invariants were extracted from the environment and internalized through the process of learning. The result of learning was assessed by the probe tone technique, so we can wonder what is the kind of information present in the stimuli that enables ARTIST to perform similarly to humans on this task. Does ARTIST 'pick-up' on the same information as humans to internalize the structure of its environment?

Naturally, examining the statistical regularities of the stimuli themselves could give us some clues to start answering this question. For instance, if the learning process is simply viewed as synaptic reinforcement resulting from exposure, it is straightforward that the number of occurrences of each note during learning could be the basic information mostly responsible for the probe tones responses. The note count is a greatly simplified way of representing pieces of music, mostly because it completely ignores the order of appearance of the notes. Brown and Butler (1981) and Brown (1988) have demonstrated the importance of order information and therefore the present approach can only give an incomplete account of human data. Still this first step can shed some light on the relationships between the basic regularities of the environment and the internalized invariants.

Krumhansl followed precisely this approach in investigating the kind of information picked up by people. However a big problem arises here as it is almost impossible to have an extensive list of all the music an adult was exposed to since birth. Therefore a relatively large corpus is needed in order to have a sample sufficiently representative of the musical environment adults grew up in. Krumhansl (1990) compiled data from Youngblood's (1958) and Knopoff and Hutchinson's (1983) who computed the frequency distributions of notes in a variety of compositions by Schubert, Mendelssohn, Schumann, Mozart, Hasse, and Strauss. Thus Krumhansl and Kessler's (1982) probe-tone profiles could be compared to tonal distributions profiles containing about $20,000$ and $5,000$ notes for pieces in major and minor keys, respectively. That is, the goodness ratings were correlated with the frequencies of occurrence. The two data sets for major keys and the two data sets for minor keys closely resembled each other, and the correlations were respectively .89 and .86 (both significant at $p < .05$).

If the tone occurrences in the pieces are weighted according to their durations instead of simply being counted, the match between profiles is even stronger; the correlation was .97 with Hughes' (1977) summed tone duration profile of a Schubert piece (op.94 no.1). It is not clear whether this better match between profiles is only due to the idiosyncrasies of the analyzed pieces or is rather due to Hughes' more sensitive approach. Taking into account the durations of the notes probably improved the match significantly, showing that the tonally important notes are temporally stressed by being given longer durations, a fact well-known in music theory. Also, the fact that Hughes' analysis included the notes of the accompaniment instead of just the notes of the melody may have been of importance.

In any case, the strong resemblance between the note distributions and the probe tone profiles suggest that people internalize the distribution of notes, plus maybe other things. Furthermore, it could also suggest that it is the internalization of the note distribution itself that determines the shape of the tone profiles. This is only speculation however, since the relationship exhibited so far between profiles is limited to a strong correlation, and no causality can be inferred.

Consequently, it would be interesting to check if this holds true for ARTIST's case. How well does the note distribution match ARTIST's tone profile? The good thing in that case is that we know exactly everything ARTIST ever 'heard' during learning. This may help us in finding out the extent to which the simple regularities of the environment (tones distribution) explain ARTIST's probe-tone profile.

Figure 5.1 shows the note distribution in ARTIST's musical universe. It accounts for every single note ever presented to ARTIST, and the assumption of having a corpus representative of the whole musical experience does not apply here as it does in the case of humans. The corpus IS the environment. The note frequency plot shows that all 6 octaves making up the input range were not equally represented. There was a tendency to use mostly the middle octaves, and the preludes rarely used the highest and lowest octaves. Hence the note distribution across all octaves approaches a bell curve.

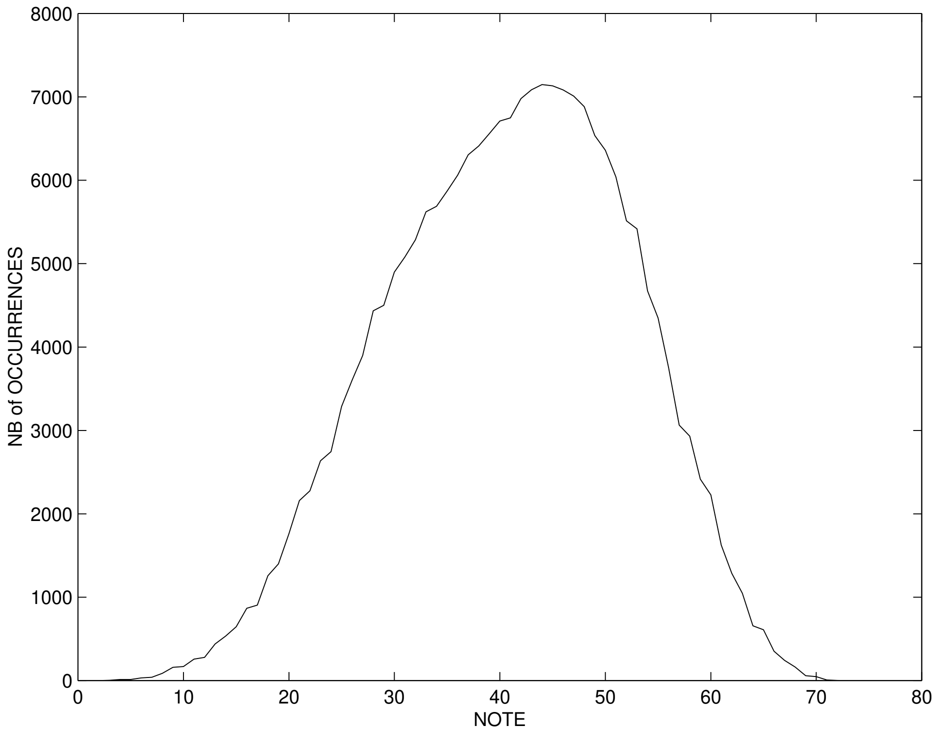


Figure 5.1: Occurrences of notes in the 288 pieces, corresponding to 24 preludes each presented in 12 keys.

The frequency distribution of the pitch classes in ARTIST's environment can be computed from the note distribution by collapsing accross the 6 octaves. The result is a perfectly flat function. Remember that ARTIST's musical experience entailed a single exposure to 288 preludes. Each of Bach's 24 preludes was presented 12 times over 12 different keys. Therefore all the keys and all the pitches were equally represented. Thus the frequencies of occurrence were strictly equal for each pitch class, namely 1/12. Each of the 12 pitch classes occurred 18,195 times during learning. Therefore the frequency distribution of the pitch classes cannot account for ARTIST's tone profiles' peaks and valleys. Obviously, something else than the simple notes distribution determines ARTIST's responses to the probe-tone task.

Given that the pitch class distribution resembles the tone profile for humans, there seems to be an important difference between humans and ARTIST. In fact there is not. The actual note distribution in pieces forming humans' musical experience is probably very similar to ARTIST's, a bell shape, because many keys must be widely represented in those pieces. In fact, the distribution of notes from humans' musical environment, so similar to the C major profile, was obtained from pieces all transposed to the key of C major. No piece played in any other key was included. If only the pieces in one particular key are taken into account, say C major, the distribution of notes spans only one portion of the musical environment (one twelveth if all keys are equally represented). Thus the note distribution explained the human tone profile only as far as the population used to compute the distribution was already carefully pre-selected. In other words, the tone profile resembles note distributions which are not representative of human musical environment. The note distribution spans one (carefully chosen) twelveth of the environment, but not the distribution over the whole environment.

Figures 5.2 and 5.3 show that for ARTIST too, the note distribution resembles the probe-tone profiles when only the pieces in the key of C (major and minor) are taken into account. Figure 5.2 shows the note distribution for the 24 preludes transposed in C. Figure 5.3 collapsed this data across the 6 octaves and thus shows the pitch distribution in the 24 transposed preludes. Its correlation with Krumhansl's C major tone profile is .71 ($p < .01$).

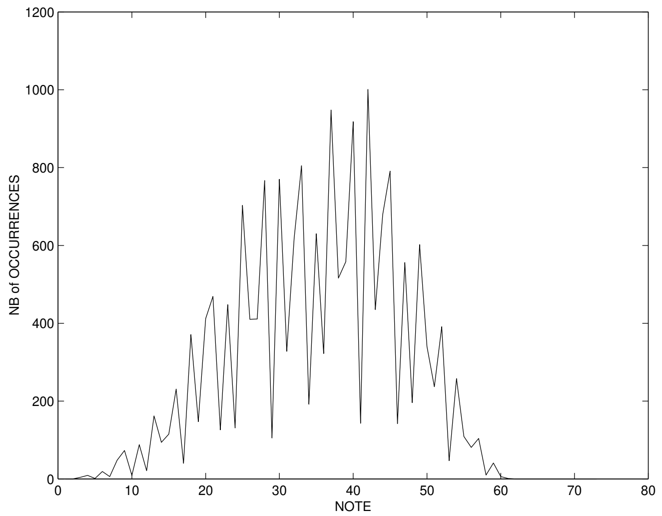


Figure 5.2: Distribution of notes in the 24 preludes all transposed to C major.

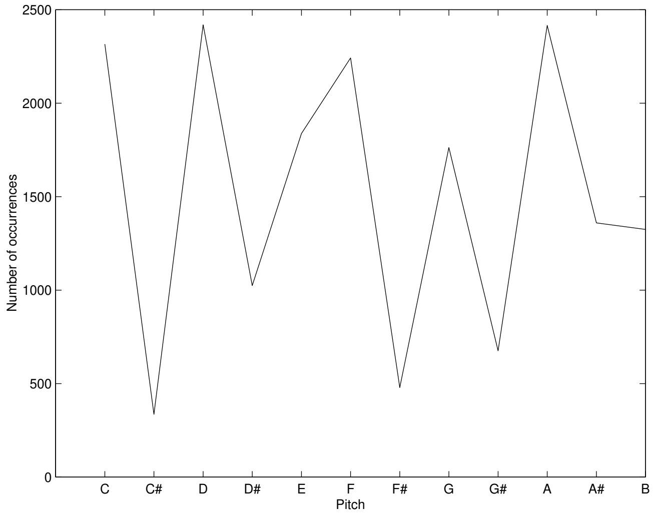


Figure 5.3: Distribution of pitches in the 24 preludes all transposed to C major.

In fact it is as if humans and ARTIST responded to probe tones according to particular sub-distributions present in the environment and not according to the whole distribution. The particular sub-distribution activated must depend on the context preceding the probe tone. For instance, a context played in C major elicits probe-tone responses resembling the note distribution in the pieces in C major. An F minor context elicits responses resembling the F minor note distribution. Another way of understanding this is that it is the *relationship* between a tonic (implied by the context) and the other notes that has to be encoded or retrieved, and not the occurrences of the particular notes themselves. This is obvious from the principle of transpositional invariance that governs the perception of music. However encoding the relationships instead of the notes themselves is not easily done, because of the circularity of the problem: encoding the relationships assumes knowledge of the tonic, which probably requires the use of the mental schemata responsible for the tone profiles. So is it possible to explain ARTIST's tone profile by only taking into account some musical context and the basic regularities of the environment?

There is a class of models, called Markov models, whose working principle is to make predictions regarding elements in a sequence depending on the preceding context; those predictions are made on the basis of the basic regularities embedded in all the sequences of a corpus. Therefore they are perfectly adapted to address the question of what kind of probe tone profiles emerge from just learning the tone occurrences in the environment. Building a Markov model using the same environment as ARTIST's, and testing it on the same probe tone task should help us understand better ARTIST's way of functioning. Specifically, this should distinguish the parts of the probe tone profiles that can be explained by simple occurrences in the environment from those that cannot. The present chapter compares ARTIST's performance on the probe tone task with that of Markov models. The latter models will not be designed with the goal of emulating humans' responses as accurately as possible. Rather, the goal is to develop them in the same conditions as used for ARTIST whenever possible, with the same data and assumptions, in order to better understand ARTIST.

5.2 Markov models and the probe-tone task

A markov model makes predictions of which element comes next in a sequence given the previous item(s). Those predictions are made by assigning a probability of occurrence to the items that could possible occur. This can easily be applied to music because music can be roughly defined as a sequence of notes. The predictions are based on the statistical regularities of the model's environment; so the model needs to be provided with a corpus of examples that constitute its environment. In our case, it is straightforward that the corpus needs to be made of the 288 preludes to match ARTIST's environment. For a given corpus, a whole family of Markov models can be built. Members of the family differ in what is called their *order*. There can be Markov models of 1st order, 2nd order, etc...

Using the same Markov models notation as used by Miller and Chomsky (1963), an nth order Markov model considers all the sequences of n elements occurring in the corpus and stores them in a contingency table. It can then be used to predict the chance of occurrence of any element following a sequence of $n-1$ elements (with the more common notation, an nth order Markov model predicts the $(n+1)$th element on the basis of the preceding n). This is simply done by looking at the final elements of the sequences with n elements that start with the $n-1$ elements specified. Replacing the word 'character' by the word 'note' in the following passage from Miller and Chomsky (1963) gives a good explanation of the notation as it is applied to the computations of note sequences: "It is convenient to define a zero-order approximation as one that uses the [notes] independently and equiprobably; a first-order approximation uses the [notes] independently; a second-order approximation uses the [notes] with the probabilities appropriate in the context of the immediately preceding [note]; etc." (p. 428) For example, a 3rd order Markov model records all the 3-note sequences present in the corpus. If we want to know what the likelihood is that the note E will follow the sequence C-D, the model computes what proportion of sequences starting with C-D are C-D-E. This gives us the conditional probability of E following C-D according to the statistical regularities of the corpus. Note that the basic assumption of markov models —that the occurrence of a note depends on the immediately preceding note— is strongly supported by the failure of Dowling's (1973) subjects to identify very familiar melodies when distractor notes are interleaved with the melodies' notes.

Since the goal of this chapter is to understand the influence of the statistical properties of the corpus on ARTIST's responses to the probe-tone technique, the Markov models need to generate a tone profile. This implies generating a rating value for each of the 12 pitches, following 3 different contexts: ascending and descending scales and chord contexts. In the case of Markov models, the probability of occurrence will be taken as the rating value. The probability of occurrence of a particular pitch class equals the probability of occurrence of any note having this pitch, regardless of its octave (it could be any of the 6 octaves covered by the corpus). For example, the probability of occurrence of the pitch class B is the probability of the notes B1, B2, B3, B4, B5 or B6 occurring.

5.2.1 0th and 1st order Markov models

The 0th and 1st order Markov models are the two particular cases where no context is taken into account to predict the occurrence of an element. The 0th order Markov model by definition gives equal probability to all the notes. In our case, the probability of any pitch occurring is 1/12 since there are 12 pitches. The probe-tone profile obtained with this model is therefore perfectly flat and does not explain any of the variance found in ARTIST's responses profile.

The 1st order Markov model assigns probabilities according to the frequencies of occurrence of the pitches in the corpus. As mentioned in Section 5.1, all the pitch classes appear equally often in the corpus because the 24 preludes were presented exactly once in every key. Therefore in this case the 1st order Markov model is identical to the 0th order model and does not make any interesting prediction: The tone profile obtained is perfectly flat. This confirms the relevance of Brown and Butler's (1981) and Brown's (1988) argument regarding the crucial importance of the order of the notes in a musical sequence.

5.2.2 2nd order Markov model

This model takes into account a context of one 'musical event' to predict the next pitch. A problem arises here as a definition of musical event is needed. To keep things simple, the best would be to define a musical event as being one note played. However in reality the musical event immediately preceding a note can be any combination of several notes played simultaneously. ARTIST also implicitely uses this definition as its input can be presented with any number of notes played simultaneously in any combination.

So the Markov model needs to take into account chords -several notes played at once- in order to make its task as similar as possible to ARTIST's.

Already we reach some limits of the Markov models as they can be applied to our particular musical problem. The contingency table of the model needs one entry for each musical event that occurs. Even restricting the music to a melody played with 3-note accompaniment chords (the standard musical situation in real life) would open more than $26,000,000$ possibilities of combining the notes for a single event! Of course, many of these combinations never occur since the notes played in harmony usually have particular relationships: they are typically one third, one fifth or one octave apart. Still, thousands of those combinations occur over the course of the 288 preludes, making the model computationally untractable. As a consequence, only two simultaneous notes will be taken into account to form one musical event. Taking into account a third simultaneous note might be computationally tractable for the present 2nd order Markov model but not for the 3rd order so musical events were restricted to only two simultaneous notes.

This in turn forces another assumption on the Markov model. When more than two notes are played simultaneously, we need to pick only two of these to define the corresponding musical event. It is straightforward to use the highest and lowest notes in this case because the upper and lower voices of a piece of music are much more perceptually salient than the middle ones (Huron, 1989; Huron and Fantini, 1989). The upper voice plays the melody and is the voice most naturally followed while listening. The lowest voice provides the bass for the accompaniment, typically sounding the tonic or the fifth. The notes played in the middle voice are often quite predictable given the lowest note, being a third, fifth, or sixth higher most of the time.

Method

The 288 preludes constituting the corpus were scanned by a MATLAB program. A list of all the different musical events as defined above was created in a lexicon. There were 1474 combinations of highest and lowest notes used in the 288 preludes. This means that any musical input present at any time in the preludes can fit in one of the 1474 categories. Of those categories, 69 consisted of a single note, the others being combinations of two notes.

Every musical input of the preludes was thus categorized and the pitch class of the highest following note was recorded in a 1474×12 matrix. One row of this matrix represents the frequency distribution of pitches following the corresponding type of musical event. Now that those contingencies are recorded, we need to submit the model to the probe-tone technique in a way consistent with the way ARTIST was tested. The biggest difference is that the 2nd order Markov model can only hold a context of one note, whereas ARTIST could work with contexts of any length, such as the 7- or 8-note contexts used in the probe tone task. For what follows, it is important to remember that ARTIST's tone profile was obtained by playing the contexts with every pitch serving in turn as the tonic, and with Shepard tones. Thus for the scale contexts all the notes in all octaves were involved once as the last note of the context, the tonic.

Scale context

In order to make the testing of the Markov model with the probe tone technique comparable to the precedure used with ARTIST, every row in the contingency matrix was considered. The 12 numbers of the row were taken as reflecting the responses to two sets of 12 probe-tone trials, one set for each note of the context category considered as the tonic. For instance, if a C major chord context like C2-E2-G2-E3 (Figure 5.4, from bottom to top) is followed by a D, the middle notes (E2 and G2, in grey) are discarded from the context, and the D counts as one instance following C2 and one instance following E3 (symbolized by arrows). Figure 5.4 summarizes the information taken into account by the contingency table.

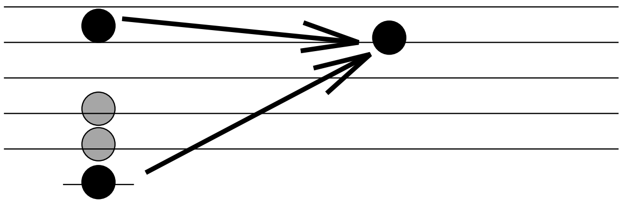


Figure 5.4: Musical information used in the Markov model. Black notes are context and predicted notes. Grey notes are discarded from the context. White note follows the context but is not used for prediction. Arrows link notes used for building the contingency table.

The Markov model tone profile is obtained by plotting the relative probabilities of occurrence of each sequence context + probe tone. Formally, this probability for the probe tone *probe* is:

Let *ascendscale* be the ordered set of notes {C,D,E,F,G,A,B,C'}. Let *scalenote* be a note element of *ascendscale* taking the values $scalenote = D$ to $scalenote = C'$ in ascending order,

$$P(C, D, E, F, G, A, B, C', probe) =$$

$$= \prod_{scalenote=D}^{C'} P(scalenote|precedingnote(scalenote)) \times P(probe|C')$$

All the terms of this product except the last one are independent of *probe* so the Markov model tone profile is determined by the values of $P(probe|C')$. To compute those probabilities from the counts stored in the contingency matrix, the 12 responses of each row are permuted in a circular fashion so that the response to the tonic always appears first. Then the sum over each column is proportional to the probability that the associated probe will follow C'. This procedure is the same as the one used to obtain humans' and ARTIST's tone profiles, when the profiles obtained with different context keys were transposed to C major before being averaged.

Chord context

A major chord is made of the tonic, the major third and the fifth. So for the chord context, only the rows corresponding to Markov contexts where the high note is the octave, the fifth or the major third of the bottom one are taken into account. The responses were all transposed to C major as explained above before the sum of ratings for one pitch was computed.

Results

Scale context

Figure 5.5 shows the Markov tone profile obtained for the scale contexts. The correlation with humans' tone profile is negative and close to 0, $r(10) = -0.20, p > .10$. The main reason for this seems to be that the model's preference for pitches close to the tonic overrides judgments based on other pitch relationships: had the peak for F been slightly lower, the 6 preferred pitches would have been the 6 closest to C, with 3 on each side.

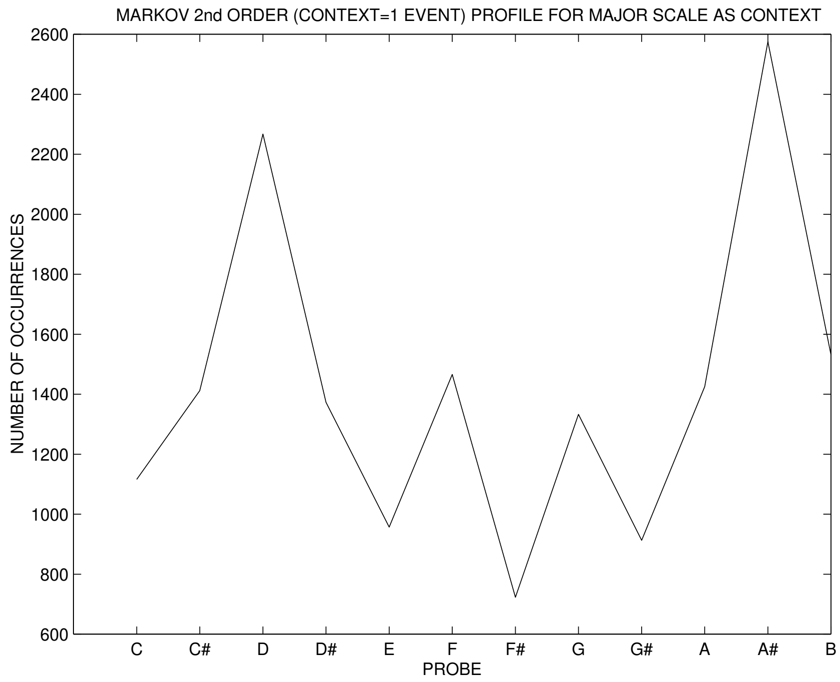


Figure 5.5: Probe-tone profile with a scale as tonal context, of which only the tonic can be used by the 2nd order Markov model.

A further important difference between profiles lies in the relative ratings of the important notes. Instead of preferring the triadic notes C, E and G to the other diatonic notes D, F, A and B like ARTIST and humans did, the Markov model preferred D and

A#. There is little doubt that the reason for this is that those are two of the notes closest to the tonic, being only two semitones away.

So why is this model mostly sensitive to pitch proximity? Reconsidering the way the profile was computed, it appears that all of it is based on the frequency of intervals between consecutive notes. It was built without any information regarding keys, and three reasons contributed to this. First, all 12 keys were equally represented with the 288 preludes of the corpus. Second, given what was just said, all 1474 contexts events were used as probe-tone trials regardless of the notes they contain. Those two remarks by themselves are not sufficient to explain the model's focus on pitch proximity, because they also apply to the way ARTIST was exposed to music and tested. The third reason combining with the others to remove key information is that the context used for prediction contains only one note. This means that each note in a sequence is in turn interpreted as being the tonic. Having no memory regarding which notes came before the pair context/predicted notes, the Markov model's musical world must be a restless flow of notes constantly modulating the key, because every new note defines a new key. This may mean that no key at all is defined, and further research may show similarities with the way humans perceive serial music.

Interpreting the Markov tone profile in terms of frequency distribution of intervals, we can easily understand why D and A# are the preferred notes, even before the tonic. It is well known that small intervals are much more frequent in music than big leaps. This was shown by analyses of pieces of music (Vos and Troost, 1989), and some theoretical justification for this can be found in Narmour's (1990,1992) influential Implication-

Realization model. The basic reason for using mostly small intervals may be to facilitate the listener's integration of music into a coherent whole, since pitch proximity is a major determinant of auditory streaming (Bregman, 1990). Given the structure of the major keys (sequence of intervals is 2,2,1,2,2,2,1) the 2 semitones intervals between adjacent notes are more frequent than 1 semitone intervals. This explains why the two preferred notes are exactly those two semitones away from the tonic, D and A#. The tonic did not receive very high rating because the unison interval (0 semitone, i.e. repetition of a note) is not very frequent.

Chord context

Figure 5.6 shows the Markov tone profile obtained for the major chord context. It is very similar to the profile obtained with the scale context and the correlation with humans' tone profiles was also quite low, $r(10) = -.30, p > .10$. The resemblance of the MARKOV profiles with the 2 different contexts is probably due to the shortness of the scale context taken into account (1 event), which makes the contexts quite similar. In any case, it tends to confirm the conclusion drawn from previous result.

Conclusion

The profiles obtained with the 2 contexts were summed. The result is shown in Figure 5.7, but nothing new appears since the two profiles were very similar and the correlation with humans' tone profile was still low, $r(10) = -.24, p > .10$.

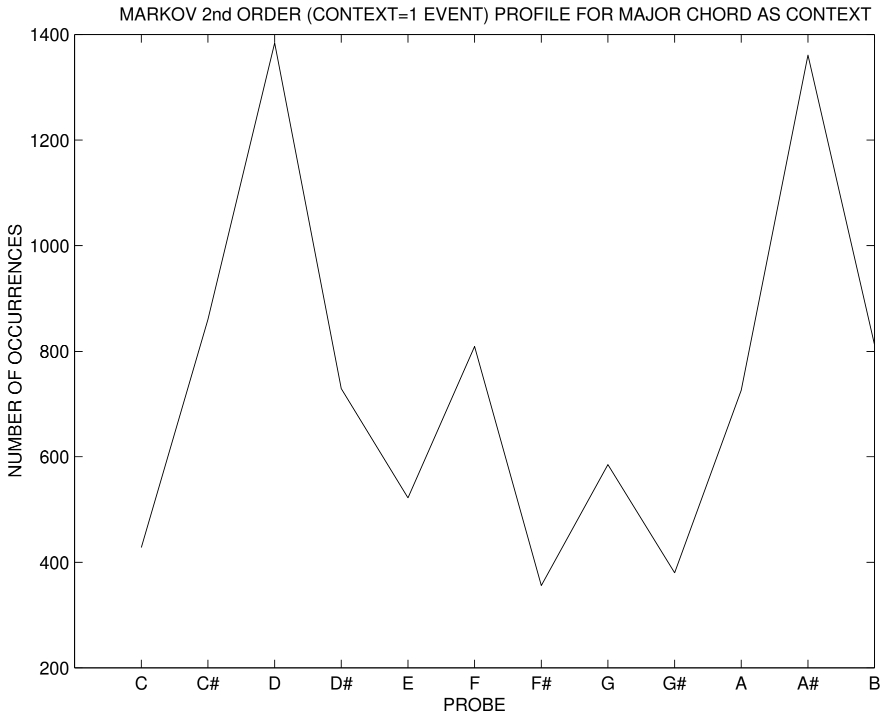


Figure 5.6: Probe-tone profile with the tonic major chord as tonal context.

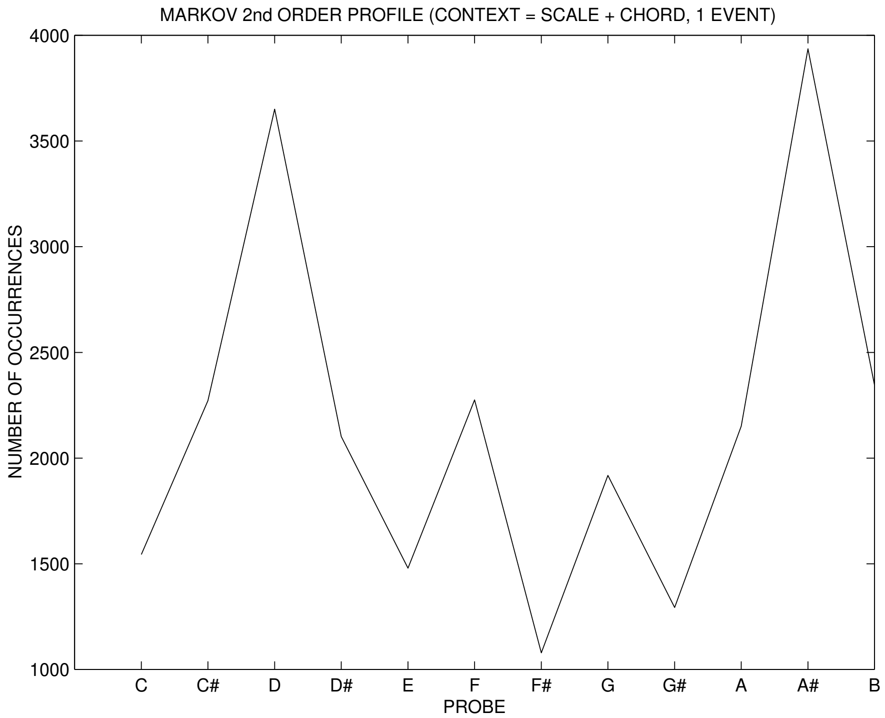


Figure 5.7: Average probe-tone profile for the 2nd order Markov model.

The model does not seem to have extracted the notion of key. It responded according to pitch proximity because the most frequent intervals are small. The high rating of A# might be explained by its presence in some minor modes. There exists three variants for each minor key, whose names are quite explicit: the harmonic minor mode is generally used when a chord is played, and the ascending and descending melodic minor modes are mostly used when melodies are played with respectively ascending and descending contours. Nevertheless those are far from being absolute rules. For instance, Bach frequently uses the harmonic minor modes in the melodic line. A# appears in the descending minor mode, but this is unlikely to explain why it obtained the highest rating because there seems to be no reason for this to prevail over all the other modes. For example, D# belongs to all minor modes, but its rating is close to average.

The most important limitations of the model concerned the way it could be tested. The model's predictions had to be based on a one-note context, so four of the contexts were reduced to the same one: the complete major and minor scales, whether they are ascending or descending, all end with the note C, which is all the Markov model can take into account. Two notes are enough to differentiate between ascending and descending scales so a 3rd order Markov model will be able to give different predictions for those two conditions. Three notes are needed to differentiate major and minor modes, so only a 4th order Markov model, very intensive computationally and not explored here, would be able to give different predictions.

In fact there was a simpler way to obtain the exact same profile. To make sure the procedures were identical to those used with ARTIST, the preludes were played in all

12 keys. So were the contexts used to generate the tone profile, and the responses to the contexts of any key were always transposed back to C major. Those two processes are reciprocal in some way, and finally cancel each other. It is as if the information available was multiplied by 12 in a systematic way, and retrieved 12 times in the same systematic way. A simulation with only the 24 preludes in C major was run. This greatly reduced the number of possible categories, from 1474 to 686. The tone profile was computed from the 686 $\times$ 12 contingency matrix by summing the occurrences of all the rows, after they were permuted to account for the tonic. When all the numbers of this profile are multiplied by 12, it exactly equals the previous one obtained with the 288 preludes. Hence the computation time required can be greatly diminished without losing any accuracy in the results. This significant improvement was used for the following simulation and it allowed to be run in a short time instead of more than a week.

5.2.3 3rd order Markov model

The same procedures as above were used for the 3rd order Markov model, the only difference being that two musical events instead of one were used as contexts. This means that the preludes were decomposed as sequences of three musical inputs. The two first inputs were each classified in one of the 686 categories according to their lowest and highest notes. The third input was assigned one of the 12 pitch classes depending on its highest note. Thus every sequence contributed to one cell in the 686 $\times$ 686 $\times$ 12 contingency matrix. Not all the data in the matrix could be used as probe tone trials, contrary to the previous case. For the 2nd order model, any note could be, through transposition, considered the last note of the scale used as context before the probe. So all entries in the matrix corresponded to a probe tone trial. In the present case, 2 notes of the context scale can be used to predict the next pitch.

Following the same argument as in Section 5.2.2 regarding the probabilities of occurrence of the context + probe tone sequences, the Markov tone profile for the ascending scale context is given by the set of conditional probabilities $P(probe|B, C)$ because the other terms of the product are independent of *probe*. Similarly, the Markov tone profile for the descending scale context is given by the set of conditional probabilities $P(probe|D, C)$.

So the two notes of a context have to hold the same relationship as B and C (the two last notes of the scale preceding the probe tone) in order for this context to be counted in the profile. In the case of the ascending scale, the second note needs to be one semitone higher than the first, as in the pair B-C ending the C major scale. In the case of the descending scale, the second note should be two semitones lower as in the pair D-C ending the C major descending scale. So only the rows of the matrix corresponding to two context notes with the relationships just explicited were used out of the 686×686 rows. This amounted to $154,065$ rows (about one third), which were then transposed so that the tonic (2nd context note) became C.

In summary, two probability distributions were retrieved from the contingency matrix. One was the probability distribution of the intervals following an interval of one semitone upward, which simulated the probe tone responses to the ascending context scale. The other was the probability distribution of the intervals following an interval of 2 semitones downward, which simulated responses to the descending context scale.

The final profile shown in Figure 5.8 was obtained by averaging both those profiles and the chord context profile, as it was done for ARTIST and humans. The chord context profile is the same as found with the 2nd order Markov model because the chord is considered as only one musical event and increasing the context length does not change anything for the prediction.

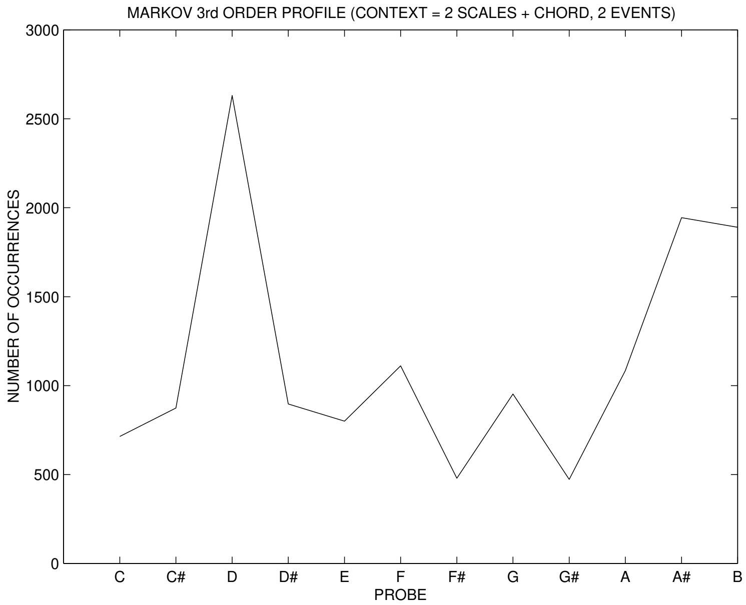
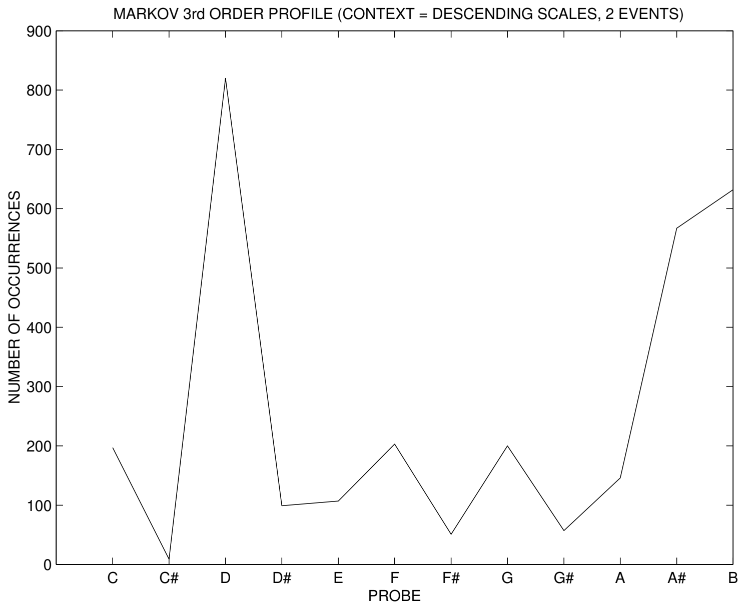


Figure 5.8: Average probe-tone profile for the 3rd order Markov model.

The profile obtained with the 3rd order Markov model is quite similar to that of the 2nd order model. Its correlation with humans' profile was still negative, $r(10) = -0.14, p > .10$. Diatonic notes are still not differentiated from the others, indicating that the notion of key was not extracted. The preferred notes are still close to the tonic, so pitch distance was still a major influence on the model's responses. The only difference of importance is that the 3rd order model gave higher ratings to B than the 2nd order model, so its responses are closer to humans'. This increase may be better understood by looking at the decomposition of the average profile into the ascending and descending scale profiles, shown in Figure 5.9.



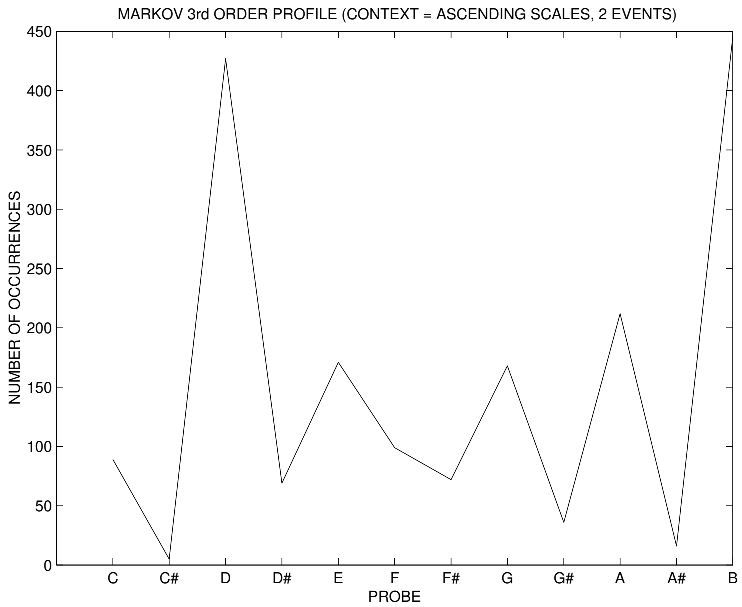


Figure 5.9: Probe-tone profiles of the 3rd order Markov model for descending (top) and ascending (bottom) scale contexts. Note the difference in A# rating.

The only big difference concerns the rating of A#: It is very close to the lowest rating with the ascending scale but third highest with the descending scale as context. This seems to validate the explanation mentioned in the previous section, based on the presence of A# only in the descending melodic minor mode but not in any other mode. The rating for B was high in both cases, because the sequences B-C-B and D-C-B commonly occur.

The rating for E was also higher in the ascending scale condition. Thus, two new local maxima (peaks) appear with the 3rd order model. This made the correlation with human data positive, $r(10) = 0.14$, vs -0.04 for descending scale context, $p > .10$ for both. Even though this correlation is not significant, the model extracted some important qualitative information: the local maxima (peaks) and minima (valleys) correspond to the same probe tones for humans and the 3rd order model. This indicates that the model has extracted the notion of key by giving relatively higher ratings to the notes belonging to the key of C. On the other hand, the relative heights of the local maxima do not match at all, which explains the almost null correlation. The differentiation of notes within the diatonic set concerned D and B, the notes closest to the tonic, instead of the triadic notes C, E and G for humans.

5.2.4 Conclusion

The application of Markov models to our particular problem of music perception suffers limitations. Only short musical contexts can be used for prediction because of their computational demand. For instance, major and minor scale contexts reduce to the same information available to the models. As a consequence, they cannot generate one major and one minor key profile. Also, only the 3rd order model have different contexts for ascending and descending scales. Moreover, the computational demand of the models compelled us to make some choices regarding the information to be fed to the model.

In spite of this, with one context the model was perfectly able to isolate the notes within the key from those outside. This implies that the frequencies of occurrence of intervals can lead to identification of the key. The effect of pitch distance was also apparent in the 3rd order model's responses, but to a smaller extent the difference between diatonic and chromatic notes had some influence.

This reminds us of the novice listeners' tendency to rely mostly on pitch proximity for their judgments (Krumhansl and Shepard, 1979). Moreover, as the next chapter explains, the distinction between notes in vs out of key is the first step children accomplish in the developmental sequence of sensitization to tonality. Together, this evidence suggests that lower order Markov models may be appropriate to simulate novices' behaviors. As novices become more sophisticated, they may integrate information over longer periods of time, and higher order Markov models are needed to account for their responses. It would be interesting to be able to implement higher order models and check if their sensitivity to pitch proximity keeps decreasing. The contribution of melodic contour would also be worth assessing, since it is known that children are especially sensitive to contour (Dowling, 1986)

Finally, the present results indicate that ARTIST may be exploiting the low-level statistical regularities of the stimuli to extract the notion of key and show preference for the diatonic notes. However, these regularities cannot account for ARTIST's insensitivity to pitch proximity. Neither can they account for ARTIST's preference of triadic notes over the other diatonic notes (and moreover in the same order of preference as humans). ARTIST may be able to exploit higher-order regularities because the number of notes presented to its input is not limited. The influence of its top-down system may have the same consequences because it acts somewhat like a working memory buffer. Most likely, it is the interplay of these elements that may be responsible for ARTIST's performance. The next section briefly explores ARTIST's behavior when its 'rhythmic attention' is perturbated and the time span between inputs varies.

5.3 ARTIST's robustness to input variations (varying window sizes)

The stimuli used were originally MIDI files, in which the unit of time is the 'tick'. So for the sake of convenience this was used as ARTIST's time unit too. The setting used was 120 ticks per quarter note, so there were 480 ticks per 4/4 measure in the stimuli. All the results reported so far concerning ARTIST were obtained using the same temporal window to scan the musical input. This window was 240 ticks wide, and ARTIST's input layer was updated every 240 ticks (or half measure). That is, each update consisted of computing the decay over 240 ticks for the old inputs (present before the presentation of the current half-measure), and of integrating the musical events that occurred in the last 240 ticks to provide new inputs. This choice was logical for both musicological and psychological reasons. In music, most of the time (and it is the case in Bach's prelude) one note is accentuated every half measure and there is evidence that this is a strong clue used to segment musical input (Handel, 1974; Drake and Palmer, 1993).

Now we can wonder whether ARTIST is robust enough to exhibit the same behavior and the same kind of performance with different time windows. This may not be the case because as it has just been explained, the time window used so far is a perfect subdivision (respectively, multiple) of basic musical temporal units, such as the measure (respectively, beat). So the particularity of the half-measure as a parsing unit of the inputs may have been crucial to ARTIST's performance. The present section addresses this question. In Section 5.3.1, ARTIST is tested with some new constant window sizes. Section 5.3.2 explores ARTIST's robustness when the window size keeps changing as the music is presented. Section 5.3.3 takes a look at the possible interaction between the window's average size and its variability.

The results presented in the following sections required ARTIST to go through the learning of the 288 preludes more than 10 different times, and through the probe-tone task more than 30 times. So ARTIST could not be as thoroughly tested as in Chapter 4: the tone profiles were computed only for the major ascending scale context. As a consequence, the following results cannot be directly compared with Chapter 4's results which also include the descending scale and the chord contexts. This probably explains the slight decrease of the correlation with human data from Chapter 4 (.95 to .88, even though the vigilance level was increased to 0.7; this was the only other difference from Chapter 4's simulations).

All simulations for Chapter 5 were conducted in the same conditions and they can be compared with each other. Even the random order of presentation of the stimuli during learning was kept the same, so the only difference between different simulations of Chapter 5 concerned the time window.

5.3.1 Fixed time windows

ARTIST was exposed to the 288 preludes under 5 different conditions of time window sizes. Even though ARTIST was already tested with 240-tick windows in Chapter 4, this window size was used again here because of the few changes mentioned above. Smaller windows (120 ticks wide) were tried, to put ARTIST in a situation more similar to that of the Markov models, which could only handle two-note contexts at best. This doubled the number of time slices presented to ARTIST, which then had to learn more than $80,000$ slices. This dramatically increased the number of categories created, reaching $1,000$ before learning was even half completed, and the simulation soon became computationally untractable.

Three window sizes larger than the original 240 ticks window were used. They were 2, 4 and 8 times larger, being respectively 480, 960 and 1920 ticks wide. For each simulation, the correlation with humans' tone profiles are shown in Table 5.1, along with the number of categories created in each case.

Table 5.1: ARTIST's performance (correlation with human data as a function of fixed window size).

FIXED WINDOW SIZE	240 (1/2 bar)	480 (1 bar)	960 (2 bars)	1920 (4 bars)
Number of categories	787	423	201	99
Shepard tones	.88	.89	.80	.89
Normal tones	.80	.85	.74	.77

In spite of a slight decrease in performance with normal tones and increasing window sizes, ARTIST's performance remains consistently high. The size of the time window does not have a lot of influence on performance, and ARTIST's behavior is quite robust regarding the window parameter. For each of these simulations, the window size was kept constant. Furthermore, the windows include an exact number of measures. Consequently, the regularity and predictability of the rate of presentation of new input, along with the fact that this matches the musical temporal unit, may be at the origin of ARTIST's high performance. So we can wonder if the performance would degrade a lot if ARTIST was exposed to a musical world with completely unpredictable flow of inputs, where two consecutive inputs could well be separated by 50 ticks as well as by $1,000$ ticks, for instance. The next section addresses this question.

5.3.2 Variable time windows

Three simulations used variable-size windows, with different ranges of variation. A minimum size of 40 ticks was imposed on all the windows to prevent empty time slices (those containing no input) from occurring too often. The window size was varied by adding a random number of ticks to the minimum size, up to 960, 1920 or 9600 ticks.

So each new input presented to ARTIST could integrate the notes of up to a little more than 2, 4 or 20 bars, respectively. The latter value was chosen very high to test what happens in case of extreme inconsistency from one input to the next.

As seen from Table 5.2, the performance was lowest for Shepard tones with smaller windows and for normal tones with the largest window. However we do not know whether the decrease in correlation for the 2 Shepard tones conditions is reliable or whether it is within natural fluctuations. It is possible that variations in window size may be a little disturbance for the model. Still, the result for the greatest windows, which sizes varied a lot (from less than 1 bar to more than 20), was surprisingly high, around .8 on average. This suggests that it is not crucial that the frequency of update of the inputs match musical temporal units.

Table 5.2: ARTIST's performance (correlation with human data) as a function of window fluctuation.

WINDOW SIZE (equivalent in bars)	40 + 0-960 (2+ bars)	40 + 0-1920 (4+ bars)	40 + 9600 (20+ bars)
Shepard tones	.64	.69	.85
Normal tones	.83	.89	.77

5.3.3 Variation ratios

In what precedes, the average window size for a particular condition is about half of its maximum size (if we neglect the addition of the 40 ticks). Therefore, the ratio between average window size and variability was roughly constant. It is possible that this ratio be of great influence on the performance of the model. It could be the size variation *relative* to the mean size that affects the inputs' consistency 'perceived' by ARTIST. So ARTIST was trained in four time-window conditions, corresponding to the crossing of two average sizes (800 and 1600 ticks) by two maximum width variation (500 and 1480 ticks). The ratios maximum variation/average size were different in all conditions. They appear in Table 5.3 along with the models' performances, with and without Shepard tones.

Table 5.3: Performance (correlation with human data) of 4 models taught with different maximum width variation/average size window ratios.

AVERAGE SIZE (ticks)		WINDOW VARIATION (ticks) ± 250	WINDOW VARIATION (ticks) ± 740
800	Ratio	0.625	1.850
	Shepard	.87	.71
	Normal	.79	.83
1600	Ratio	0.312	0.925
	Shepard	.86	.83
	Normal	.83	.80

Once again, performance was quite consistent across all conditions, even more so than with the previous simulations (Sections 5.3.1 and 5.3.2). The performance with Shepard tones was slightly better than with normal tones except in one condition. Like before, the performance differences for different types of tones or for different windows are not large enough or systematic to hint at any particular influence of the window size or window variation.

Taken together, these results suggest that fixed window sizes optimize ARTIST's performance with Shepard tones: the three correlations around .88 in the fixed window conditions were the highest of all the simulations with Shepard tones. However, the perturbations caused by varying windows, if any, seems minimal.

5.4 Conclusion

Markov models match human data poorly if we judge by the low correlation between responses on the probe tone task. The correlation was not significantly different from 0 even for the 3rd order Markov model and virtually none of the variance in human data was accounted for. However, they gave interesting results: the 2nd order Markov model bases its responses mostly on pitch proximity, and the 3rd order model shows distinction of the notes within a key vs those outside the key. Given what is known of the perception of music by novice listeners (see next chapter), this suggests that Markov models may be good to simulate novices' behavior.

In contrast, ARTIST accounts for 90% of the variance in human data, with strictly the same information available as that to the Markov models. ARTIST's better performance can probably be explained by the fact that its functioning allows for integration of information over longer periods of time. It is possible that ARTIST would perform equivalently to Markov models if the notes were fed three by three to its input as they are to the 3rd order Markov model. Unfortunately, the proliferation of categories in this situation prevented the test of this hypothesis by making the simulation untractable.

Further simulations showed that changing the number of notes (through changing the time window) fed to ARTIST's input at each step does not substantially affect the performance. This does not mean that performance will remain unaffected if the time window gets so small as to contain only 2 or 3 notes. But the convergence of top-down activation with input activation would probably increase the number of notes active at the input layer, thus re-establishing a context of 5 notes or more. Therefore it seems unlikely that putting ARTIST closer to Markov model's situation by reducing the time window size would reduce the .95 correlation to 0. In any case, ARTIST's exceptional robustness regarding the temporal aspect of the coding of the input is another feature it shares with the humans it emulates. It is possible that feeding inputs to ARTIST according to musical time creates optimal conditions for ARTIST to match human responses but in no case is this crucial to a good account of human data.

If Markov models are good candidates to mimic humans' development of the perception of music, we do not know yet how well ARTIST performs on this point. The following chapter explores ARTIST's developmental steps to compare them to humans'.

CHAPTER 6

SIMULATION 3: ACQUISITION AND DEVELOPMENT OF THE TONAL HIERARCHY

ARTIST can acquire a great deal of musical knowledge just by 'listening' and learning the pattern it is exposed to. Understanding how the process of acculturation progresses with age and exposure to music is important to know whether the learning process of the model is plausible. Specifically, it would be useful to observe stages in ARTIST's musical development, and see if the sensitivities to different musical features emerge at different times. Since the tonal hierarchies are the main indicators chosen to evaluate the model's cognitive resemblance to humans, the following chapter reviews the gradual acquisition of the tonal hierarchies by children.

6.1 Tonal hierarchies and musical correlates

Krumhansl examined how the tone profiles correlate with other measures applied to musical material, such as tonal consonance or tonal distributions. Tonal consonance refers to the smoothness or roughness (high or low consonance) perceived when two notes are sounded together. Following von Helmoltz's intuitions (1885/1954) it is widely accepted that consonance is primarily a consequence of the acoustical properties of the signal and the way this signal is processed by the peripheral auditory system (Greenwood, 1991; Plomp and Levelt, 1965). Even though the details of von Helmoltz's theory do not appear to be totally accurate, consonance and dissonance are still believed to be primarily consequences of innate human characteristics, the strongest argument for this being the universality of the principles of consonance across virtually all cultures in the world, short of a couple of exceptions only. As expected from humans' perceptual system tremendous flexibility and adaptation to the environment, our capability to learn and change over time can come to influence our perception of consonance, through the learning of associations for instance. Learning as a secondary source of influence on the perception of consonance was also recognized by Helmoltz, and a modern version of his two components model of musical consonance (the innate and learned components are respectively called *sensory consonance* and *harmony*) is outlined in Terhardt (1984; see also Huron 1994 for a discussion of this issue).

The relative consonances of the tonic/probe-tone pairs probably play a role in shaping the tone profiles, because the contexts used for the probe-tone technique clearly imply a tonal center. Thus, the context + probe-tone sequence should receive a rating similar to the tonic + probe-tone sequence rating, itself probably similar to the rating of the tonic/probe-tone pair sounded simultaneously. To have an idea of how well the tone profiles reflect tonal consonance, Krumhansl (1990) compared her tone profiles with six different measurements of consonance, coming from theoretical as well as experimental studies. She found a moderately strong match with them, and concludes that the tone profiles reflect more than the simple acoustic properties of the notes. Which other factors may influence the shape of the tone profiles?

The previous chapter suggests that the distribution of pitch classes in the environment may be important to explain the tone profiles because it is internalized to some extent by listeners. As mentioned in Section 5.1, Krumhansl (1990) found a strong match between the tone profiles and the distribution of pitch classes in a variety of compositions. That the distribution of pitch classes accounts for the tone profiles better than the measures of consonance do indicate the importance of the learning mechanism responsible for the gradual internalization of the distribution of notes in the environment.

6.2 The order of appearance of the levels of the hierarchy

Studying if and how the tonal hierarchies emerge at an early age could give us important insights regarding their origins and the respective importances of innate biases and of learning in shaping them. The perception of music has been studied in infants and children, and all studies do not agree about the precise age at which a given perceptual ability appears, which is understandable given potentially big differences in the paradigms used, in the education infants received, and in their musical environments. In spite of these differences, all studies seem to point in the same direction: humans' perception of music gets refined by exposure to music and learning, and goes from being sensitive to gross structural and perceptual features to being tuned to more and more subtle features.

Note that it is not necessarily desirable to have a perceptual system that would be perfectly tuned to the smallest features, picking up on all the differences between two stimuli. However, picking up only on meaningful differences presupposes that the system knows which differences are meaningful, and this depends on the task the system is asked to perform. Focusing on insignificant differences can impede the generalization process: two similar stimuli that we would want to be processed in the same way could end up being processed in two different ways because our perceptual system noticed the subtle differences and activated different mental schemata for the respective stimuli. For instance, listeners possessing absolute pitch do not perform as well as average listeners in recognizing intervals or melodies under transposition (Miyazaki, 1993a, 1993b). Conversely, a perceptual system better able to generalize than another one is not necessarily the best since this can mean it is insensitive to some differences in stimuli: infants younger than a year old sometimes outperform adults in detecting mistunings (Lynch et al., 1990) or within-key changes to a melody (Trainor and Trehub, 1992).

The two probe tone task

Krumhansl and Keil (1982) tested children of different ages and adults with the two-probe tones technique (mentioned in Chapter 4) to find out if the distinction our perceptual system draws between the groups of tones revealed by the tone profiles (diatonic set, major triad, tonic, represented in Figure 6.1) emerge at different stages. The four-tone sequence chosen to establish the context contained only notes from the major triad, C E C G. The context was followed by two notes (probes) and the children were asked to judge how good or bad was the resulting 6-note sequence, considered the beginning of a melody. Children indicated their judgments by pointing to one of seven dots. The extreme and middle dots were labeled with smiling, frowning and neutral faces. The pair of probe tones contained either two chromatic notes or two diatonic notes, with all possible pairs for the latter case. Each pair of probes fell into one of five categories: triad-triad, nontriad-triad, triad-nontriad, nontriad-nontriad and nondiatonic-nondiatonic (e.g., C-E, F-G, G-F, D-A and F#-G#, respectively). These categories are listed in decreasing order of importance in the hierarchy, therefore this order should correspond to progressively lower average ratings. The non-trivial case concerns the relative order of the nontriad-triad and triad-nontriad categories (conditions 2 and 3). However, we know that the former should correspond to stimuli rated higher because the unstable note is played first and is assimilated to a more stable note over time (refer to discussion of asymmetry in Section 1.4).

The results led to a surprisingly clear picture of the developmental changes in sensitivity to tonal functions, summarized in Figure 6.2. Distinctions between diatonic notes and chromatic notes appeared first (grades 1 and 2, condition 5 is different from others). Then triadic notes get progressively more differentiated from other diatonic notes: both notes need to be in the triad for a better rating of the pair by 3rd-4th graders, whereas one triadic note is sufficient to distinguish the pair from non-triadic pairs by 5th-6th graders (cond. 1 vs 2,3,4 in 3rd-4th grade; cond. 2,3 vs 4 in 5th-6th grade). Finally, only adults showed reliable different ratings depending on the order of presentation of one triadic and one non-triadic note (cond. 2 vs 3, the asymmetry discussed earlier). The differences in ratings between probe pairs categories monotonously increased with age, reflected by a steeper function for older children and adults. This summarizes the general tendency of increased differentiation between probe pairs with age.

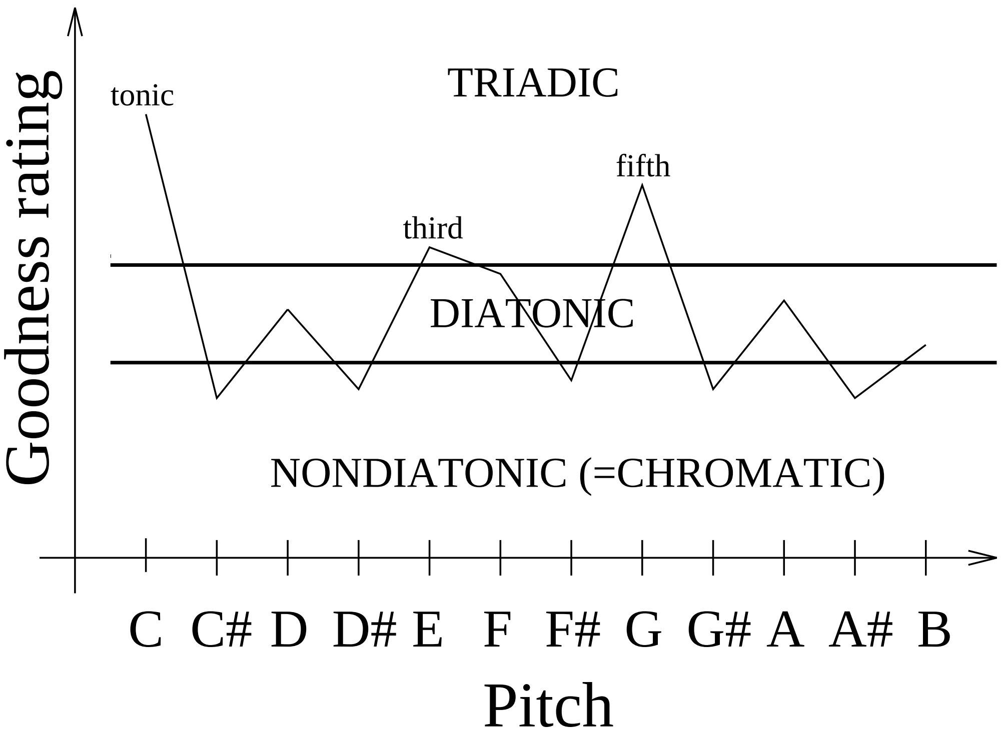


Figure 6.1: The 3 hierarchical levels in the C major profile.

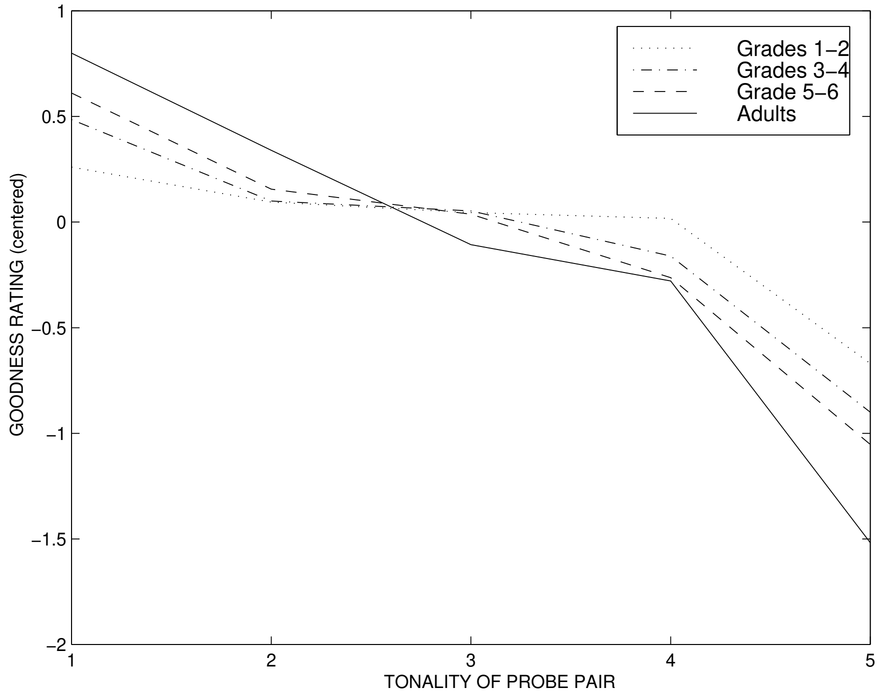


Figure 6.2: Average centered ratings on the 2 probe-tone task as a function of age and type of probe pair (Krumhansl and Kessler, 1982).

This order of acquisition of the hierarchy seems natural from the point of view of information theory, because it must be optimal in some sense (maximizing the variance) to extract big clusters of elements from a population first and then work on progressively smaller clusters, instead of extracting successively small clusters. However, as pointed out by the authors, two facts are surprising from a psychophysical point of view, both suggesting that the children, especially the youngest ones, did not rely on the low-level acoustic properties of the stimuli in their judgments (or at least not enough to be detected).

First, the tonic notes (C and C', an octave apart) did not receive higher ratings than the other triadic notes until the adult stage, in spite of being the tonal center which is uniquely defined by the set of diatonic notes, and in spite of constituting half of the context stimulus (2 notes out of 4). Second, the notes of the major triad, that constitute the whole musical context and therefore should elicit sensory priming, did not receive higher ratings before 3rd-4th grade. The authors acknowledge that this does not mean that children do not perceive such simple physical properties, but only that other factors had more influence on the responses.

The single probe tone task

Using only slightly different experimental conditions, Speer and Meeks (1985) and Cuddy and Badertscher (1987) showed that children could exhibit preferences for the tonic vs other triadic notes and for the triadic vs other diatonic notes as early as 1st and 2nd grade. The latter study included a condition using the same context stimulus as Krumhansl and Keil's, so it is probably the change from two probe tones to only one probe that enabled the children to exhibit differentiated sensitivity to tonic and triadic tones in both these studies. Using a complete major diatonic scale instead of a 4-note melody to establish the context may have helped too.

In fact, Cuddy and Badertscher's study concludes that the major triad context is more efficient than the major scale in instantiating the tonal context. Looking at the profiles generated by those two contexts, it is clear that the former gives rise to a profile much closer to that obtained by Krumhansl and Kesslser's (1982) than the latter. This means that children can handle a 4-note context more easily than an 8-note context. But this does not mean that the triadic/diatonic distinction precedes the diatonic/chromatic distinction. The same graphs reveal that besides very high ratings for the tonic, children gave very similar ratings to all diatonic notes, clearly differentiating them from the chromatic notes but not differentiating the triadic notes from the other diatonic notes.

Thus, the results of those different studies may not be as incompatible as suggested at first look: as suggested by all studies, the diatonic scale structure is internalized by the time children get to first grade. This also agrees with experimental results using a totally different method, such as Dowling (1990) and Andrews and Dowling (1991) who showed that tonality effects emerged around 7 years of age in a complex task such as the recognition of a familiar melody with interleaved distractors. Also, as the more recent studies suggest, the importance of the major triad and of the tonic arises around this time but is not so well established, and is therefore exhibited only with a simpler task (only one probe tone) or with context stimuli more sensitive to the triad and tonic differentiations (major scales). Lamont and Cross (1994) addressed the apparent inconsistencies of the studies mentioned above with two experiments. One involved the probe-tone technique with new types of contexts (chord sequence and major scale notes in random order), and the other used a totally different method, trying to actively involve the children in musical 'games' so they have the chance to exhibit their highest degree of sophistication. The results of both experiments give partial support to the previous studies, especially Krumhansl and Keil's, and good support to the present conclusion overall, even if the complexity of the results obviates any simple developmental sequence that accounts for all the data: the 4-way analysis of variance [Age $\times$ Sex(School) $\times$ Context $\times$ Probe tone] found three main effects, five one-way interactions, four two-way interactions and the three-way interaction!

6.3 An innate bias?

The numerous studies by Trainor and Trehub (1992, 1993, 1994; see also Trehub et al. 1986) can also help understanding the difficulties of replicating the sequence of apparition of sensitivities to tonal functions, even though they did not specifically address this issue. First, it was shown that children are already sensitive to diatonicism by ages 4 to 6, consistent with the fact that by age 7 the diatonic scale is well established. Supporting an explanation of this phenomenon based on progressive acculturation, 10-month olds never exhibited such sensitivity. However, infants that young are affected by the distance between the keys of transposed melodies when detecting a semitone change in one note. Is it possible to exhibit a key-distance effect without having internalized the structure of the diatonic scale? It could be, if we consider that infants more readily detected the alteration of an interval between consecutive melodies when those melodies were a fifth apart than when they were a major third apart (respectively one and four steps apart around the circle of fifths). However, as the authors point out, we do not know whether this effect is systematically related to the key distance, because only two distance conditions were contrasted, near key (transposition to the fifth) vs far key (transposition to the major third). The effect may simply be due to the particularity of the fifth.

This study, along with previous research (Trehub, Thorpe and Trainor, 1990; Trehub and Unyk, 1991), confirms the idea that the perfect fifth may hold a special status in the way it is processed by the perceptual system, constituting a prototype for auditory patterns because of the simplicity of the frequency ratio (3:2). The authors conclude that the enhanced processing of the perfect fifth results from an innate bias of our auditory perceptual system, after rejecting the possibility that it could be the result of the learning occurring during the first ten months of life. The latter explanation was rejected based on the fact that unlike adults, 6-month olds detect the mistuning of notes equally well in melodies based on the Western major scale or in melodies based on the Javanese Pelog scale (Lynch et al., 1990).

However, all that is shown in Lynch et al.'s study is that the categorical perception of pitch is acquired rather than innate, and is not completely acquired by six months of age, because detecting mistunings is a task that reveals the extent to which pitch perception is categorical. This experiment rules out the acculturation explanation in Trainor and Trehub's (1993) study only under the assumption that acculturation does not happen faster for triad processing than for categorical perception of pitch. Replicating Trainor and Trehub's results with newborn babies would rule out the acculturation explanation of the key distance effect more clearly. Even though it is not known whether the priviledged processing of the perfect fifth is present at birth or learned during the first 10 months, in either case it seems to be the second feature of the tonal hierarchy humans are sensitive to, second only to the unison/octave relationship, already perceived at the age of three months (Demany and Armand, 1984).

In summary, for children younger than 7, the naturally priviledged position of the perfect fifth coexists with an increasing sensitivity to diatonicism, and probably translates into a preference for the major triad over other diatonic notes in certain circumstances. However the latter may not always be detected since its acquisition is slower and requires more time than the preference for diatonic notes over chromatic ones. Table 6.1 summarizes the main results of the developmental studies. An 'X' in the column labeled 'Tonic' means that the ratings received by the tonic were significantly different from those received by the other triadic notes. An 'X' in the column labeled 'Triad' means that the ratings received by the triadic notes were significantly different from those received by the other diatonic notes. An 'X' in the column labeled 'Diatonic' means that the ratings received by the diatonic were significantly different from those received by the chromatic notes.

TABLE 6.1: Development of tonal sensitivity exhibited by the 2 probe-tone technique

— means no reliable effect; X means a differentiation between the notes in the set and those in the embedding (immediately larger) set; XX means X plus a differentiation between one and zero (out of two) note in the triad; XXX means XX plus a temporal order effect of the triadic note (first vs second probe); ? means no data available.

Article	Age group	Context	Tonic	Triad	Diatonic
Krumhansl and Keil 1982	1st + 2nd Grd	CECG (2 probes)	—	—	X
	3rd + 4th Grd		—	X	X
	5th + 6th Grd		X	XX	X
	Adult		X	XXX	X
Cuddy and Badertscher 1987	1st-6th Grd (3 groups; no difference)	CECG (1 probe)	X	X	X
		Ascending scale	X	X	X
		Diminished triad	—	—	—
	Adult (3 exp. levels; no difference)	Triad	X	X	X
		Ascending scale	X	X	X
		Diminished triad	X	—	X
Speer and Meeks 1985	2nd & 5th Grd (no difference)	Ascending scale	X	X	X(5th Grd)
		Descending scale	X	X	X
Trainor and Trehub 1994	5-yr-old	melodies(10 notes, I-V-I modulation)	?	?	X
	7-yr-old		?	X	X
	Adult		X	X	X

6.4 ARTIST's early years

It was seen in Chapter 4 that exposure to 288 pieces was sufficient for ARTIST to reach the adult-like stage of musical sophistication. We can wonder whether ARTIST's learning is so fast that it will directly reach 'adulthood' upon leaving its originally blank state. Alternatively, substantially less exposure to music could bring ARTIST to a partial learning of tonality. If this were to be the case, would that partially developed sense of tonality be similar to children's?

6.4.1 Procedure

ARTIST was tested with the two-probe tone technique at three different points during learning plus at the final stage, reflecting different levels of exposure to music. We know that the rate of creation of the categories during learning is very fast in the beginning and slows down pretty quickly. Therefore the different levels of exposure chose to test ARTIST were not equally spaced in terms of amount of exposure (number of pieces learned). ARTIST was tested mostly during the very early stages of development since the major changes in architecture (addition of categories) happen early (see Figure 2.4). ARTIST was tested with the two-probe tone technique after exposure to 24, 48, 144 and 288 pieces. The 24 preludes were presented an equal number of times in a random order. That is, the 4 stages respectively corresponded to 1, 2, 6 and 12 exposures to each prelude. Each exposure of a prelude was in a random key, independently from all the other exposures.

The context used was the 4 notes sequence C-E-C-G, identical to Krumhansl's. Following the context, all 144 pairs of tones were used to probe ARTIST in order to cover all possible cases and have rating averages as reliable as possible for each of the 5 tonality conditions. Probe pairs therefore included all 70 diatonic-nondiatonic (unordered) pairs, absent in Krumhansl's test, as well as all 25 nondiatonic-nondiatonic combinations, of which only 7 were present in Krumhansl's study. These differences only resulted in different number of trials for tonality condition 5 (95 vs 7). As it was done in simulation 2, each sequence (144 in this case) was used 12 times, to allow the tonic to take all 12 pitch values. This insures maximum reliability of ARTIST's responses by involving all of its knowledge rather than just the knowledge pertaining to one particular pitch as the tonic. This is even more important than in simulation 2 because at intermediate stages of learning, the exposure of ARTIST may not be balanced regarding to the keys of the stimuli (tonic pitches). Thus, the total activation in F2 was recorded for the 1728 trials at each of the 4 developmental stages. The opposite was taken as ARTIST's rating of the stimuli, since we know from simulation 2 that preferred stimuli minimize F2's total activation.

6.4.2 Results and discussion

Main effect of probe pair type

The results are plotted as 4 different functions (ratings as a function of probe pair type), one for each developmental stage, in Figure 6.3. The resemblance with Krumhansl and Keil's results (1982; see Figure 6.2) appears immediately. For all developmental stages, ratings decreased as the pair of probes occupied lower positions in the hierarchy: there was a slow decay between tonality conditions 1 to 4, and a sharp drop for tonality condition 5. ARTIST accounts for the main effect of stimulus category with the same pattern of average rating per category as humans.

Interaction between musical exposure and probe pair type

Part of the interaction between age and type of probe pair is also accounted for: the function gets steeper as amount of learning increases. With more exposure to music, ARTIST's ratings increased for good sounding stimuli (tonality condition 1) but decreased for those sounding less good (tonality condition 5) in a systematic way. This shows that with more experience at listening to music, ARTIST is better able to differentiate its answers depending on the type of stimulus.

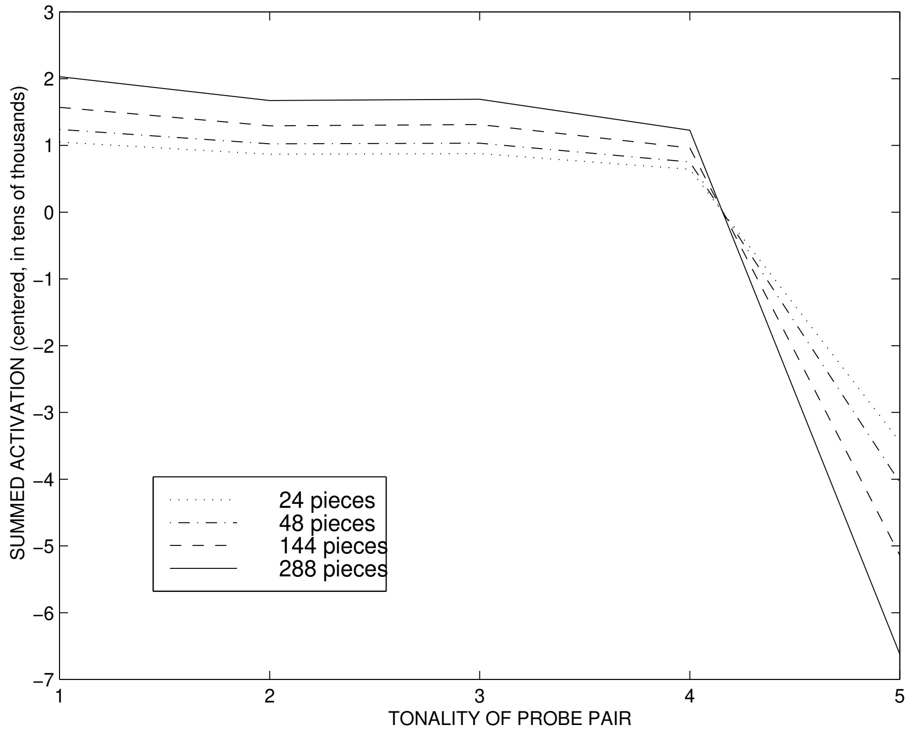


Figure 6.3: ARTIST's goodness ratings of probe tone pairs for the 5 tonality conditions of pairs' position in the tonal hierarchy by developmental (learning) stage.

All the above observations concerning ARTIST's development are strictly the same as Krumhansl and Keil's (1982) for human data. This implies that in both cases, the four functions of developmental stages cross somewhere between tonality conditions 1 and 5 to reverse their order (e.g., adults at the top in tonality condition 1 and at the bottom in tonality condition 5). However there is one difference between ARTIST's and humans' data: the exact place where the functions cross. For ARTIST, they cross between tonality conditions 4 and 5, meaning that the four functions are in the same order for tonality conditions 1 through 4. In fact, those functions look parallel on this plot.

In contrast, human listeners' functions are not parallel but cross between tonality conditions 2 and 3, as shown in Figure 6.5 (close up from Figure 6.2). The relative order of the functions is already completely inverted in tonality condition 4 (adults at the bottom) compared to tonality condition 1 (adults at the top). This means that with more exposure to music, the increased differentiation between ratings was already apparent between tonality conditions 1 and 4. This appeared not to be the case with ARTIST. However, the lines between tonality conditions 1 to 4 may not be completely parallel. It could be that the fluctuations of the functions are so small compared to the drop of the ratings in tonality condition 5 that they were made invisible by the small scale. To test this hypothesis, Figure 6.3 was redrawn as Figure 6.4 after tonality condition 5 was dropped and the 4 functions were centered. This revealed that the interaction between musical exposure and probe pair category for ARTIST was in fact almost identical to that for humans.

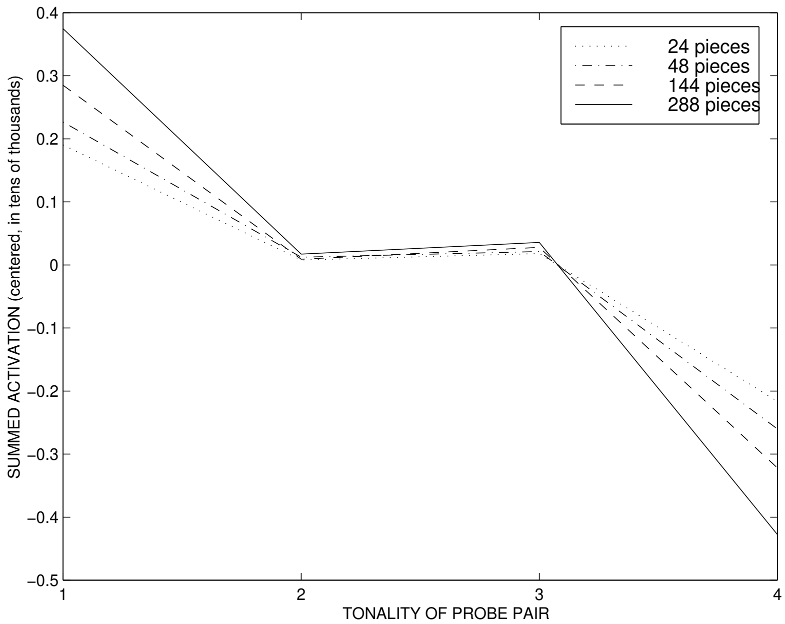


Figure 6.4: ARTIST's goodness ratings of probe tone pairs for only 4 tonality conditions of pairs' position in the tonal hierarchy and amount of learning (close-up of Figure 6.3 with functions re-centered).

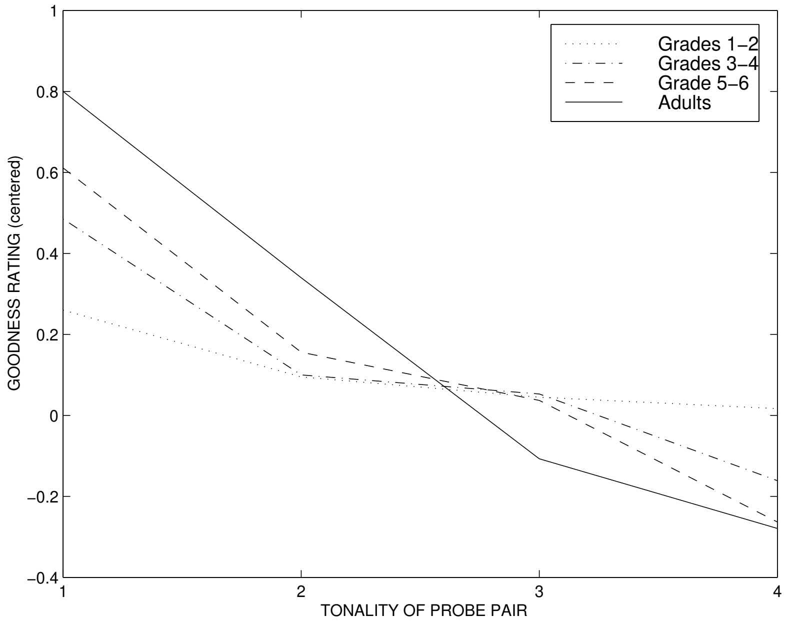


Figure 6.5: Human data from Krumhansl and Keil (1982) restricted to tonality conditions 1 to 4 (close-up of Figure 6.2).

In fact, ARTIST's increased differentiation between types of stimuli is present even between tonality conditions 1 and 4, where the function gets steeper with musical exposure. In summary, ARTIST accounts for a very large part of the interaction.

The limits of ARTIST

There are still two minor phenomena present in human data that ARTIST cannot account for. They are the lack of main effect of musical exposure for tonality conditions 1 to 4, and the asymmetry effect (tonality condition 2 vs 3). Each is discussed in turn.

Unlike human data, ARTIST's data shows that for all tonality conditions 1 through 4, ratings are highest with the most exposure to music. But many parameters affect the relative vertical positions of the four functions. For instance, the way the functions were centered before they were plotted determines the vertical gap between them. Originally, before centering occurred, the function with smallest exposure to music provided the highest ratings in all conditions, consistently with human data where younger children gave higher ratings overall. The human data were centered and normalized for each subject in order to compensate for this. It is apparent from Figure 6.4 that had the normalization been based only on the rating values for tonality conditions 1 to 4 (ignoring condition 5), all the data would have matched close to perfectly, with only the asymmetry unexplained. So anything that affects the way humans' or ARTIST's functions are centered can potentially explain the difference between data. This could be the different number of trials between both tonality conditions 5, a non-linear use of the 7-point scale of goodness by humans, or the fact that the number of nodes in F2 varies greatly from one developmental stage to the other (many of them are created by learning in between those stages, thus affecting the total output). It is possible that a different vigilance or learning parameter would yield a perfect match between data by influencing the number of nodes created during learning. In any case, ARTIST and human data only differed by an additive constant, which makes the difference between data sets insignificant and meaningless.

On the other hand, the difference between human data and ARTIST's that seems meaningful concerns the ratings of the tonality conditions 2 vs 3. Adults showed a preference for the probe order nontriad-triad rather than the opposite. It is well-known that the stability of the last note of a musical sequence is of prime importance in determining how good the sequence sounds. Almost all melodies end on the tonic for this reason. In fact, we can be rather surprised that this effect did not appear sooner than at the adult stage. ARTIST did not exhibit this effect, may be because it had not reached its 'adult' stage yet, and it is possible that even more exposure to music would eventually trigger the effect. It is also possible that ARTIST will not differentiate between the two orders of the probe tones because they occur close to each other in time. Thus decay of activation may not have effects different enough on the tones to distinguish them. Increasing the activation decay or playing the stimuli at a slower tempo may solve this problem.

There might be also a more likely explanation. Looking at closely tonality conditions 2 and 3 of Figure 6.4, we can see a slight asymmetry that goes in the opposite direction than the asymmetry exhibited by humans. It is very slight but consistent, it appears at all four developmental stages. The slight asymmetry in ARTIST's rating is in fact systematically related to the slight difference in activation decays for both tones. After all, ARTIST may reliably find the order triad-nontriad more stable than the opposite. Considering the input activations after presentation of the whole musical sequences of tonality conditions 2 and 3, only four nodes are activated, and three of them constitute the major triad C-E-G. This pattern is probably overlearned at the highest degree by ARTIST's F2 nodes and may be a special case. C-E-G constitute the C major chord, which was probably learned in many combinations with repetitions, inversions, etc... Thus the addition of a fourth note in this pattern could drop significantly the matching activations of many nodes, the more recent the note, the bigger the drop. Taking into account that ARTIST's rating is inversely proportional to the total activation, a drop in activation would yield a higher rating of stability.

In summary, the recency of a fourth note intruding on the overlearned major chord pattern of notes may be responsible for a decline in activation, that is, an increase of the ratings. More research is needed at this point to validate this hypothesis.

Conclusion

Except for the order effect of the probe tones, ARTIST can account for all the human developmental data of Krumhansl and Keil (1982). In the beginning, ARTIST is only sensitive to the presence of nondiatonic notes in the probe pair. As exposure to music increases, triadic notes are progressively extracted from the other diatonic notes for preferred processing. Ratings also get more contrasted from one category of probe pair to the other, showing a more reliable differentiation similar to humans'.

CHAPTER 7

SIMULATION 4: THE MUSICAL MODES

Chapters 4 and 6 showed that ARTIST internalized the invariants of the music it was exposed to, and that using this knowledge led to behaviors similar to humans' on both the one and the two probe-tone tasks. The behavior in question entails giving ratings of how good or bad musical sequences sound. The musical sequences used were prototypical contexts (scales or chords) followed by one or two probe tones, but were not anything like real musical excerpts. However the tone profiles appear to be also relevant to more general musical situations. For instance, Cuddy (1993) showed that the tone profiles obtained by using real melodies as contexts were very similar to those obtained by Krumhansl and Kessler (1982). Thus it is likely that ARTIST's preference for some probe tones as a function of a particular context captures the essence of tonality in music, and generalizes to judgments involving more real musical situations, such as preferences for some melodies over some others. This claim should nevertheless be tested, and this is one of the two goals of the present chapter. The other goal is to test whether ARTIST is sensitive to different forms of music within tonal music. Thus far, ARTIST's responses have been shown to match humans', but only in the context of major and minor modes. In fact, several other modes exist in tonal music, even though their use has become rare nowadays. This explains that they have been the focus of very little research, and that no data was readily available from the litterature with which to compare ARTIST's responses. Therefore, an experiment manipulating the mode of melodies had to be conducted with human subjects.

The quality of a model is often assessed through its ability to make accurate predictions, and the problem at hand gives us an opportunity to do just that. That is, given several groups of melodies with different characteristics, ARTIST can be used to predict those harmonious to human ears from those less harmonious. Then an experiment involving human subjects can be designed to test the accuracy and relevance of ARTIST's prediction. ARTIST's and humans' responses to the different modes are presented respectively in Sections 7.1 and 7.2. Since we are free to choose the stimuli of the experiment (as long as they are the same as the simulation's stimuli), this approach also enables us to gather further interesting information along the way, as the next paragraphs explain.

The first advantage of this approach is that ARTIST can be tested on some rather subtle points. All the melodies used in the present simulation are constrained to be tonal, which is a prerequisite to being associated with a particular mode (the concept of mode is explained in more details at the beginning of the next section). The characteristic chosen to differentiate one group of melodies from another is the mode. One reason for this is that the difference between two melodies with distinct modes is rather subtle (because they are both tonal), compared to the difference between a tonal melody and an atonal one, for instance. ARTIST most probably responds differently to tonal vs atonal melodies because atonal melodies lack the kind of regularities found in its environment, made exclusively of tonal melodies in the major and minor modes. In contrast, tonal melodies of all modes exhibit the same kind of regularities. Specifically, each mode is associated with a particular pattern of intervals. Given its close approximation of Krumhansl's tone profiles, ARTIST seems to have extracted the notion of key, and therefore should be sensitive to the pattern of intervals. Specifically, it should exhibit a preference for the major and minor modes which are the only familiar ones. That is, this preference would probably appear with the probe tone task, using the scales of the different modes as contexts. But it is not obvious that such a preference will be exhibited with less uniform musical sequences such as real melodies. Thus it is a good challenge for ARTIST to exhibit a sensitivity to the mode instantiated by real melodies.

Second, the human data can be interesting in itself, as it will reveal an aspect of 20th century listeners' tastes. The results are not generalizable to listeners of other times because the use of modes other than major and minor was more widespread in early music than it is now.

Finally, the beginning of an answer can be brought to the following question: Can the aesthetic differences between the 'unused' modes (other than major and minor), if any, come from exposure to only major and minor modes? Can these differences be explained by their degree of similarity to the familiar modes? Because ARTIST was only exposed to the major and minor modes during learning, a good match between ARTIST's and humans' data would go in the direction of an affirmative answer to this question. A poor match would suggest that ARTIST was missing some crucial information or that it is inherently limited and cannot solve this kind of problem.

7.1 Predictions from music theory

This simulation is designed to investigate whether the model can predict the 'goodness' ratings given by humans to melodies of different modes.

The concept of mode is in some sense complementary to that of key. All major keys share the same pattern of semitone intervals ascending from the tonic (2,2,1,2,2,2,1). It follows that any key transposed to C becomes identical to the key of C. Nevertheless, different keys have different tonics by definition and therefore the sets of the pitch classes they contain are different.

Conversely, in their usual notation, the seven modes contain the same set of 7 pitch classes (the diatonic notes), but have different tonics. It follows that each mode is associated with a unique pattern of intervals. Consequently, when transposed to a common tonic like C, the modes contain different pitch classes.

Therefore there are 7 different modal scales, shown below with the patterns of intervals (tones or semitones) that uniquely define them and with an example of the scale after transposition to the tonic C:

C D E F G A B C' (C major scale = Major mode, intervals 2,2,1,2,2,2,1)

D E F G A B C D' (2,1,2,2,2,1,2) Dorian transposed to C: C D Eb F G A Bb C'

E F G A B C D E' (1,2,2,2,1,2,2) Phrygian

F G A B C D E F' (2,2,2,1,2,2,1) Lydian

G A B C D E F G' (2,2,1,2,2,1,2) Mixolydian

A B C D E F G A' (2,1,2,2,1,2,2) Aeolian (Natural minor)

B C D E F G A B' (1,2,2,1,2,2,2) Locrian

The traditional explanation of the arrangement of the keys around the circle of fifths is that the more pitches shared by two keys, the closer the keys around the circle. The similarity between the sets of pitches of two keys is believed to be at the origin of the perceptual similarity of the keys. If the same principle holds for the modes, counting the number of pitches common to the different modes should give a good index of the perceptual similarity between modes.

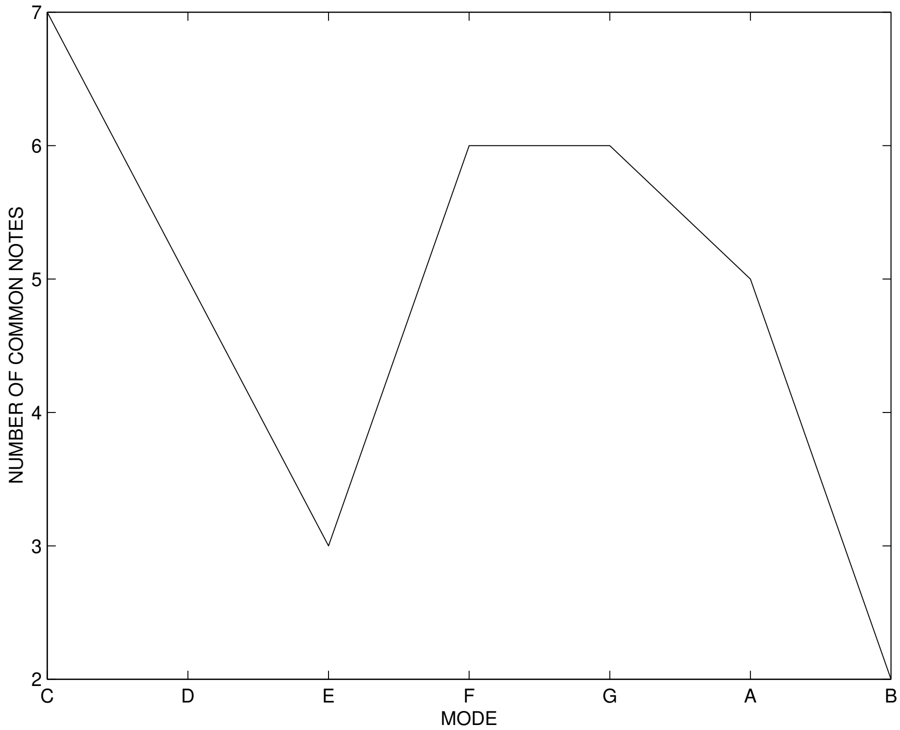


Figure 7.1: Relative goodness of modes based on the number of notes shared with the major mode (when transposed to the same tonic).

In turn, if the similarity between modes generalizes to the similarity of their respective goodness, the simple count of overlapping pitches between modes can be taken as an index of modes' goodness. It is likely that the major mode, being the most familiar, will be the preferred one. It follows that the other modes' pleasantness may be proportional to the number of pitch classes shared with the major mode. If this is the case, the human data will look like the graph in Figure 7.1, based on the number of common pitches between each mode and the major one. For instance, we can count from the example above that the dorian mode shares five pitch classes with the major mode (we do not count the tonic twice even though it appears at the beginning and at the end of the scale).

Another reasonable hypothesis is that people perceive the similarity between keys or modes based on the similarity of their sequences of intervals. For example, the dorian mode has three intervals that are identical and in the same place as in the major mode (positions 1, 4 and 5). As seen in Figure 7.2, the prediction based on shared intervals is very similar to that based on shared notes, the only notable differences being the ratings for the E and B modes (Phrygian and Locrian, respectively).

Note that these predictions are consistent with the key distances as measured around the circle of fifths. This is because the distances between modes are monotonous ly related to the distances between keys. However, this relationship is not strictly monotonous and in some cases, two keys neighbor around the circle of fifths have the same 'modal distance' from the major mode.

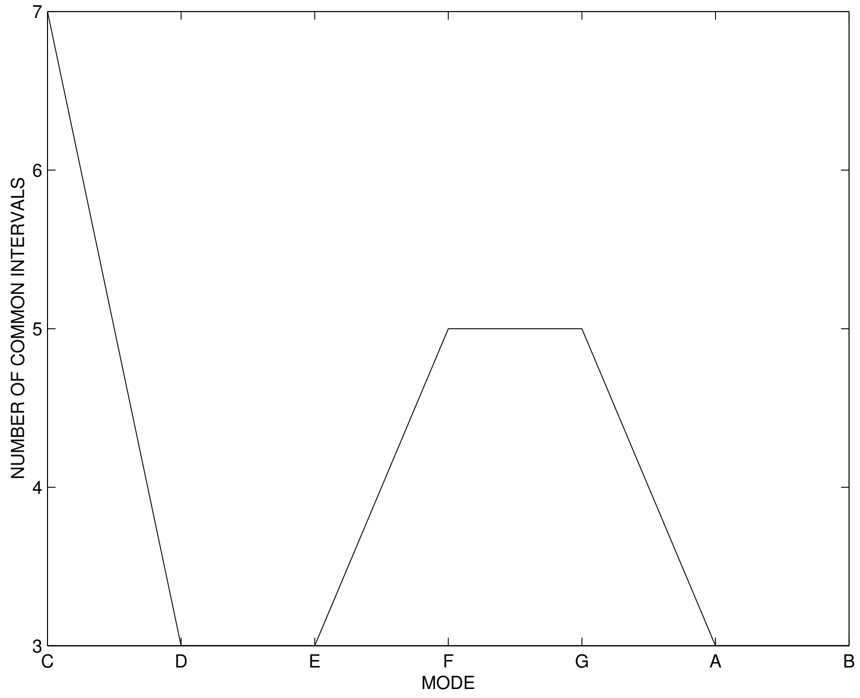


Figure 7.2: Relative goodness of modes based on the number of intervals shared with the major mode.

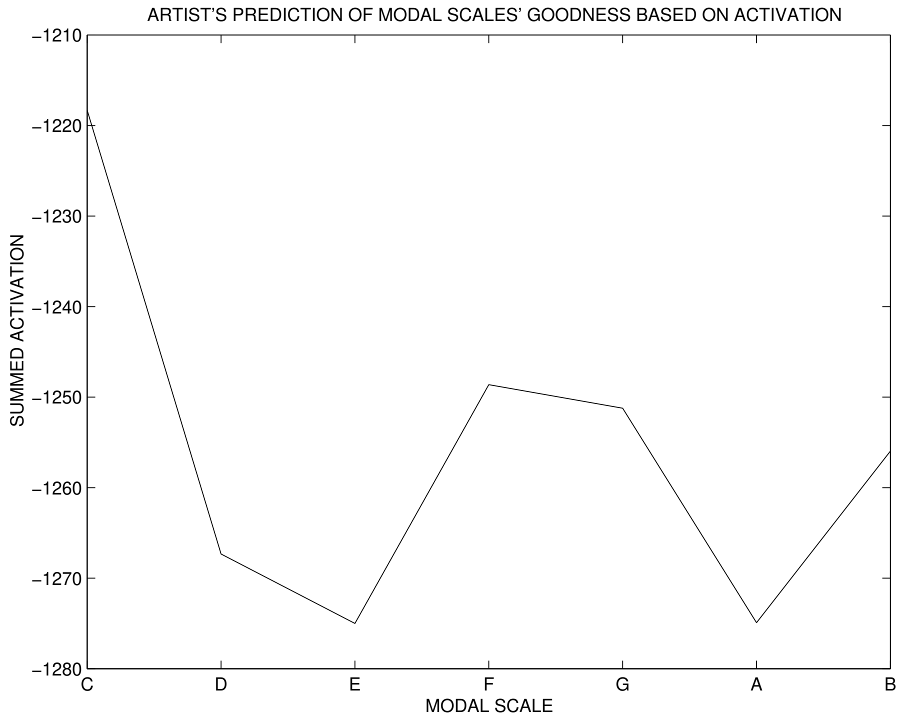


Figure 7.3: Relative goodness of modes based on the activation evoked by the modal scales.

Out of curiosity, the modal scales were given to ARTIST as inputs. The total activations were recorded for each, and are plotted in Figure 7.3. ARTIST's prediction fits the theoretical predictions quite well, except that the Locrian mode (B) rating is a little too high. ARTIST's prediction seems to depend on both the number of notes and intervals shared with the major mode, because the rating for the Dorian mode (D) is in between the two theoretical predictions for this mode.

ARTIST was also subjected to the probe tone task 6 times, following the same procedure as used in Chapter 4. The only difference was that the 6 contexts were the modal scales other than the major mode scale, since the tone profile for the latter is already available. Six new tone profiles were thus obtained, for the Dorian, Phrygian, Lydian, Mixolydian, Aeolian, and Locrian modes, shown in Figure 7.4. The intermodal distances were computed in the same way inter-key distances were computed in Chapter 4: The correlations between the modal profiles and the Major mode profile were taken as an index of distance from the major mode. The results are plotted in Figure 7.5. The match with the theoretical predictions of Figure 7.2 is even better than that of the prediction based on activations in Figure 7.3, now the prediction is lowest for the Locrian mode. The resemblance between predictions from music theory and the two types of prediction given by ARTIST suggest that it processes the different musical modes in a very plausible way.

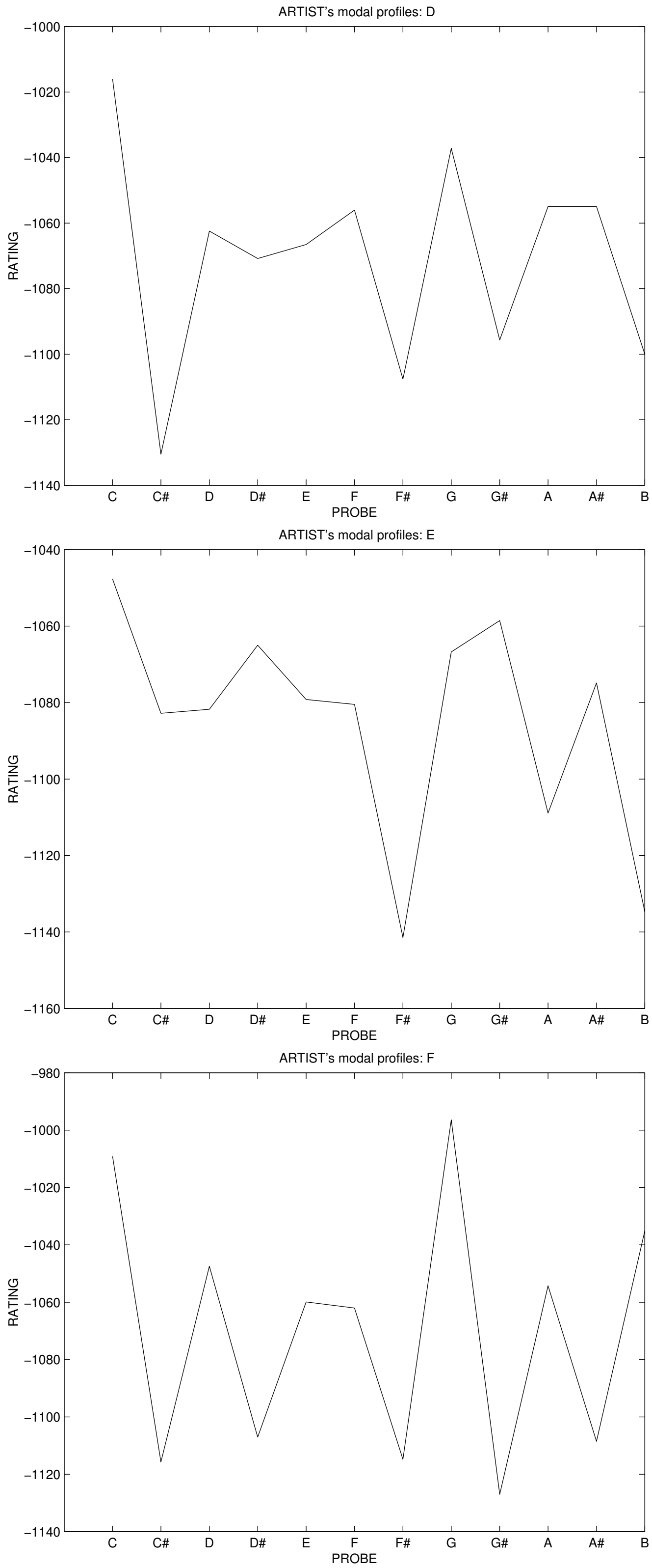
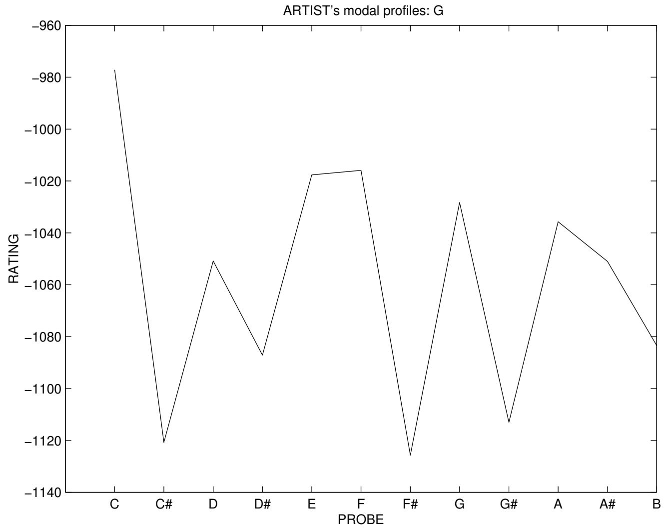
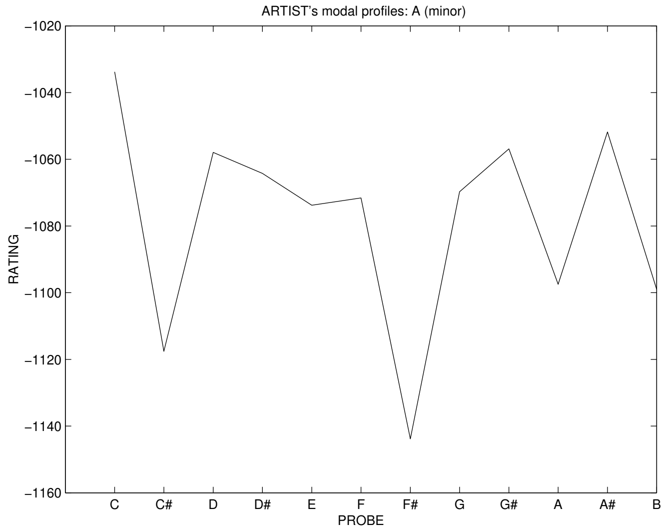
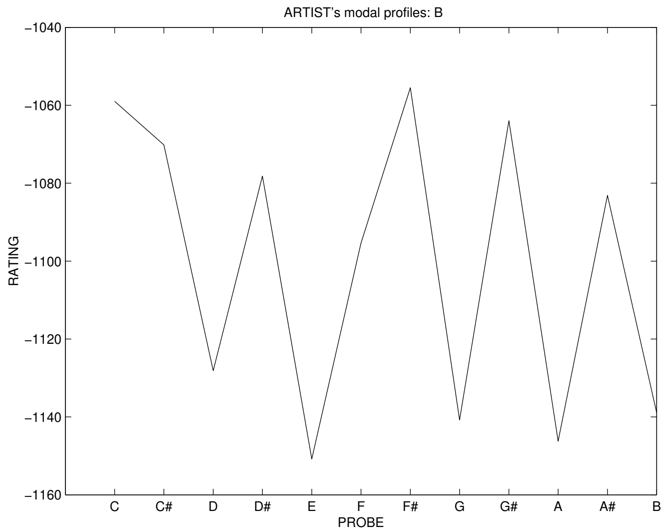


Figure 7.4: The modal profiles obtained by ARTIST with modal scales as contexts.

Figure 7.4 (Continued)







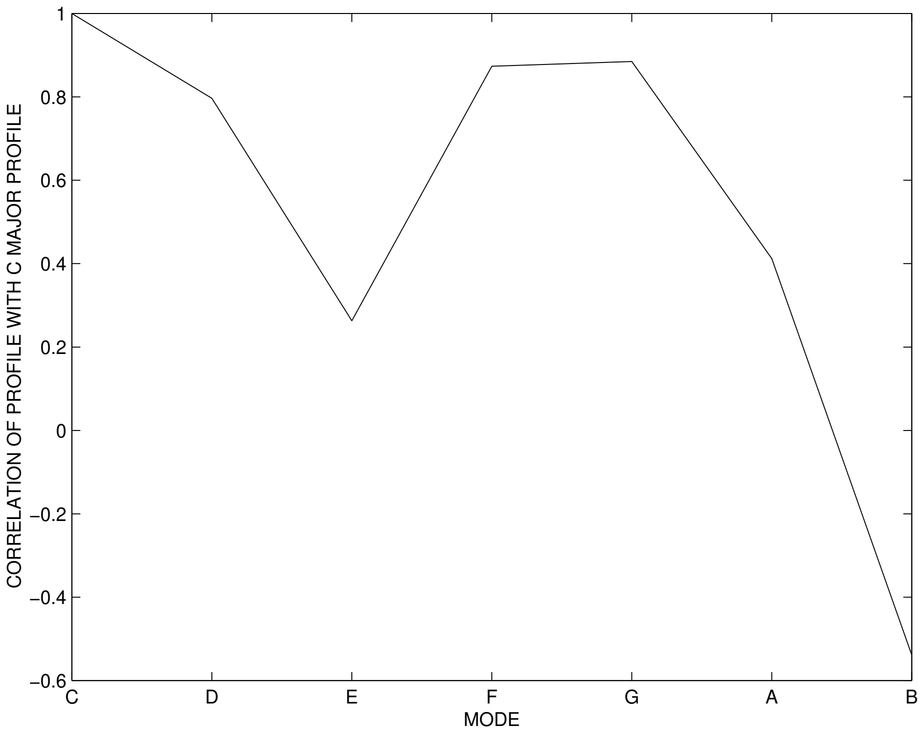


Figure 7.5: Relative goodness of modes predicted by ARTIST based on the correlation between given mode and major mode profile.

Throughout this section, the Locrian mode received the lowest scores most of the time. It contains the most dissonant intervals (tritone and minor second) and does not contain the most consonant one, the fifth. As a consequence, it is virtually unused and will not be considered in what follows. The focus of this study is thus restricted to the 6 other modes, making the experiment with humans more manageable.

7.2 ARTIST's predictions

Stimuli

For each of the 6 modes, three melodies were picked from Ortmann's compilation of traditional folklore tunes "Music for sight singing" (1959) or from "The dictionary of musical themes". Those eighteen melodies were coded in a MIDI file with the software

Cakewalk, so this file can be used for playing the melodies to human subjects or for presenting them to ARTIST.

All the melodies were modified in order to avoid other musical features that could be confounded with mode, and so that mode is the only distinguishing feature between groups of melodies. That is, each melody was translated to a 4/4 meter, was given rhythmic variations if too isochronous, and was given a tempo so that the average note density was close to 2 notes per second. Also, notes were added or changed so that each melody had at least one occurrence of every pitch class in the mode; this insured that the mode was unambiguously defined. The melodies were 14 to 24 notes in length and ended on the tonic.

The melodies were shifted to each of the other five modes, along the diatonic scale. This of course affected the patterns of intervals (in semitones) of the melodies, while rhythm and pitch contour were preserved. The identities of the melodies as defined by the usual principle of transpositional invariance were changed.

The resulting in 108 melodies were classified into 6 groups, one for each mode. Each of the 18 melodies in a group has 5 counterparts, one in each other group, that has same rhythm and pitch contour but a different interval pattern. Thus the groups differed only on the intervals present in their melodies and were identical regarding everything else.

Method/Results

As in the previous simulations, the model's rating of each stimulus was the sum of 12 ratings. Every melody was presented 12 times with the tonic taking all the pitch class values. This way, the rating is independent of the absolute pitch level of presentation of the melodies. For each presentation, the sum of the neural activations at the top layer of neurons (F2) was recorded.

This 'goodness' judgment task is the same as the judgment involved in the probe tone task, so the same measure as used to retrieve the major profile was used here. That is, the activations at the abstract level were summed, and the resulting profile should significantly correlate negatively with the human data. The profile given by ARTIST taken upside down is the prediction for the modal scales goodness.

When the neural activation is summed and recorded after the end of the presentation of a melody, the result depends critically on the content of the last measure of the melody, which provided the input processed most recently. However, we do not know the extent to which the previous measures affect the result. ARTIST's behavior could be chaotic (i.e., very sensitive to initial conditions) and the carrying over of top-down activation may lead to a result that depends on all the previous measures. It is also possible that neural activations previous to the last input are flushed through the bottom-up and top-down cycle of activation and do not contribute significantly to the final result.

Therefore, to make sure ARTIST's rating reflects the rating of a whole melody and not only of the way it ends, the neural activation needs to be computed after each activation cycle. This way, the rating of every measure in the melody is taken into account. So this measure is very sensitive to the length of the melody, and it cannot be used to compare two melodies if they have different numbers of measures. Averaging neural activation over number of measures would be a solution to this, but for the present simulation, we are comparing groups of 18 melodies rather than individual melodies. The total number of measures for the 18 melodies is the same across the six groups of melodies, so there is no need to correct for melody length.

In summary, ARTIST's goodness rating for one mode was computed by summing the neural activations for all units in F2 and after each measure, for all 18 melodies played in this mode, and on 12 presentations with different pitch heights. The ratings for all six modes are shown in Figure 7.5. This was obtained with simple tones as opposed to Shepard tones, so this constitutes ARTIST's definitive prediction because the stimuli will not be played with Shepard tones to the human subjects.

However, it could be interesting to look at ARTIST's responses with Shepard tones, because this could reflect the way very experienced listeners process music, i.e., in terms of pitch classes. It was mentioned in Chapter 4 that the tone profiles obtained by Krumhansl and Shepard (1979) for novices depend almost exclusively on pitch distances, whereas experienced listeners' profiles depended on the tonal relationships between pitch classes. This suggests that experienced listeners are less sensitive than novices to the contour of a melody and to the octave on which notes are played. For them, the pitch classes themselves are the most important features determining pleasantness. It is this idea that prompted Krumhansl and Kessler (1982) to use Shepard tones in the following study, that fit very well the results of the experienced listeners in the previous study where Shepard tones were not used.

In the most extreme ideal case, experienced listeners could be processing all notes according to their pitch class only and totally ignoring the pitch height. In other words, they would hear all the notes as Shepard tones. The reality is certainly not so extreme, but following this idea allows us to predict how the relative ratings for modes change as listeners get more experienced. ARTIST's responses with Shepard tones are shown in Figure 7.6, predicting that the modes of F, G and A will sound relatively better to experienced listeners' than to novices. The F and G modes should receive high ratings according to the theoretical predictions (Figures 7.1 to 7.3, and 7.5), so this is in agreement with the hypothesis that experienced listeners, by processing the stimuli more in terms of pitch classes, would produce results more consistent with the rules of tonality than unexperienced listeners would. Further research could tell us if simulating musical expertise by more exposure to music (like it is done in the previous chapter) would yield the same prediction.

A last caveat regarding the latter prediction should also be considered. It is possible that the Shepard tone-like processing of all notes by experienced listeners be an artifact due to the particularity of the stimuli in Krumhansl and Kessler's (1982) study. For instance, the monotonously ascending or descending contour of scales is not affected by the use of Shepard tones because scales are sequences of small steps of one or two semitones. But the contour of melodies is affected, and what was true in the context of a scale may not generalize to the processing of real melodies.

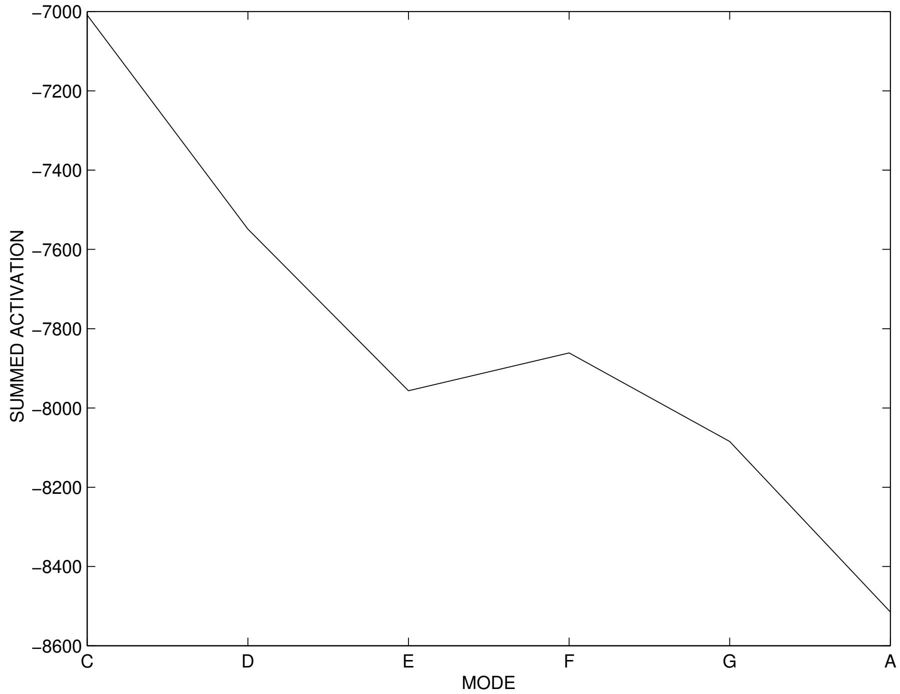


Figure 7.6: ARTIST's prediction of relative goodness of modes.

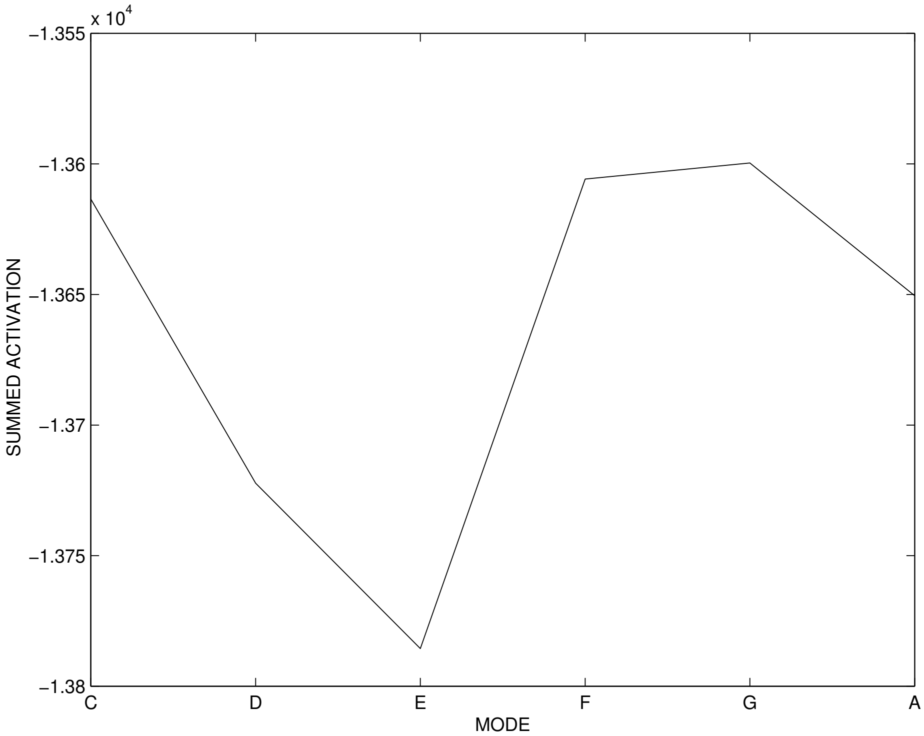


Figure 7.7: ARTIST's prediction of experienced listeners' relative goodness ratings of modes with Shepard tones.

7.3 Human experiment

Subjects

Twenty students from the University of Texas at Dallas received course credit in exchange of their participation. Half of them received some musical training for 5 years or less (mean=1.6, std=1.5), and were categorized as ‘inexperienced’. The other subjects were categorized as ‘experienced’ (mean=11.4, std=4.9).

Stimuli

The starting pitch of each melody was randomized to avoid any possible effect of pitch height. The order of presentation was randomized within the constraints that no two consecutive trials should be in the same mode (except for major, see below) or the same original melody. Otherwise, priming of the mode or of the melody could occur and there may be some undesirable effect of context.

Five out of every six melodies are in an unfamiliar mode (not major nor minor), so there might be an effect of habituation: After many exposure to strange sounding melodies, the following ones may not sound so different any more and all melodies may end up having similar ratings. So we need to remind the listener from time to time what 'normal' melodies sound like. This should insure that the perceptual contrast between modes remain. To this end, four supplementary major melodies were chosen from French children's songs and recorded. One supplementary melody was inserted every five trials, and the corresponding ratings were ignored in the data analysis. Furthermore, to counterbalance the kind of context effect just mentioned, the number of trials immediately following a major mode trial should be the same for all the modes.

However, this was not possible because the number of major mode trials was not an even multiple of 6. In the end, the number of trials following major melodies was 6 for the A, C, D and E modes, and 5 for the F and G modes.

Method

Subjects were asked to rate how good each of the 108 modal melodies sound, on a 7-point scale (from 1='very bad' to 7='very good'). They were specifically instructed to rate the entirety of the melody to prevent recency from clouding the judgments, which may naturally focus on the ending of the melody. Instructions also emphasized that judgment should be made on the pitch dimension, and that subjects should try to ignore other features such as contour or rhythm as much as possible. The six instances of a melody have same contour and rhythm, and if those features prevail over mode for the ratings, no effect of mode will be found.

Each melody was originally composed in a given mode, and then shifted to all the other modes, so two factors were associated with each trial: composition mode and played mode. From ARTIST's predictions, we expect a main effect of played mode and an interaction between played mode and expertise level. The summed neural activations in ARTIST for the two expertise conditions are not directly comparable so no prediction is available regarding the possible main effect of expertise. However there should not be any effect if subjects in both groups center their judgments around the middle of the rating scale.

Results

The 2 expertise levels $\times$ 6 composition modes $\times$ 6 played modes design was analysed by a 3-way ANOVA with melodies nested in composition mode and crossed with played modes, and subjects nested in expertise level. The ANOVA found the main effects of composition mode and of played mode to be significant, $F(5, 71) = 7.76$ and $6.19, MSe = 6.46, p < .00001$ and $p < .0001$, respectively. There was also a significant interaction between composition mode and expertise level, $F(5, 71) = 3.41, MSe = 1.16, p < .01$. The interaction between composition mode and played mode was not significant, $F(25, 71) = 0.46, MSe = 6.46, p = .98$.

On the average, the melodies played in Aeolian (A) mode received the highest ratings, followed in decreasing order by the Major (C), Dorian (D), Mixolydian (G), Phrygian (E), and Lydian (F) modes, as shown in Figure 7.8. This was also the order of preference exhibited by the group of musicians. Non-musicians ranked the modes in almost the same order, but preferred the Phrygian to the Mixolydian mode. However the difference in ratings between these two modes was slight and the interaction between played mode and musical expertise was not significant, $F(5, 71) = 1.31, MSe = 1.16, p = .27$.

The melodies originally composed in the major mode (C) received the highest average ratings, followed by those composed in Dorian (D), Aeolian (A), Phrygian (E), Mixolydian (G), and Lydian (F) modes, as shown in Figure 7.9. This factor interacted with the listeners' musical expertise: Compared to non-musicians, musicians gave higher ratings to the 3 modes rated highest (C, D and A) and lower ratings to the 3 other ones.

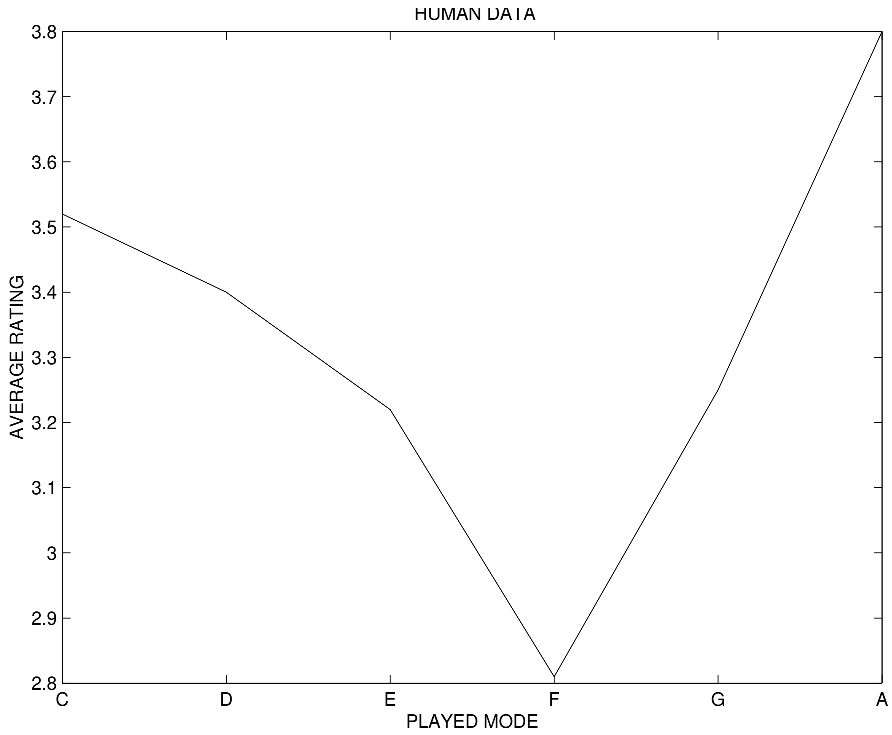


Figure 7.8: Average goodness of melodies composed in all modes as a function of their mode of play, from human data.

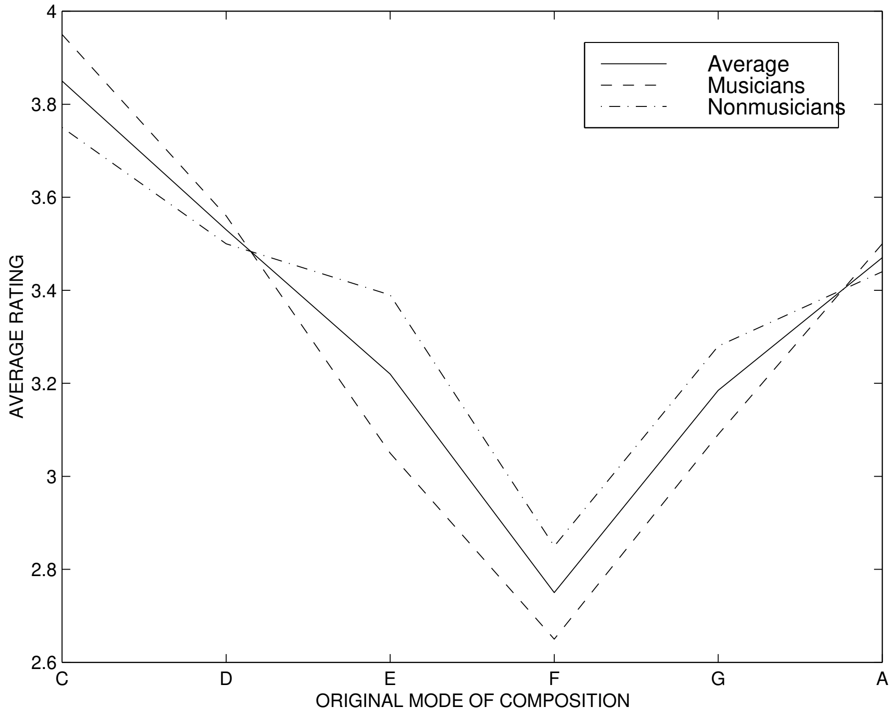


Figure 7.9: Average goodness of melodies played in all modes as a function of their original mode of composition, from human data.

A 2-way ANOVA with mode and musical expertise as crossed factors was also conducted on the ratings that relate only to the melodies played in their original mode of composition. This was done in order to check whether all the melodies originally sounded equally good in their own mode of composition, because it is possible that the difference observed for the modes in the previous analyses are due to an asymmetry of the effect of modal transposition. For instance, the original melodies composed in the modes of C and F may sound equally good, and the C melodies could retain this quality after transposition to the F mode but not the F melodies after transposition to the C mode. The ANOVA revealed that only the main effect of mode was significant, $F(5, 90) = 14.16, p < .000001$. As Figure 7.10 shows, the average ratings obtained by the different modes are very similar to those obtained with the analyses on the complete set of stimuli.

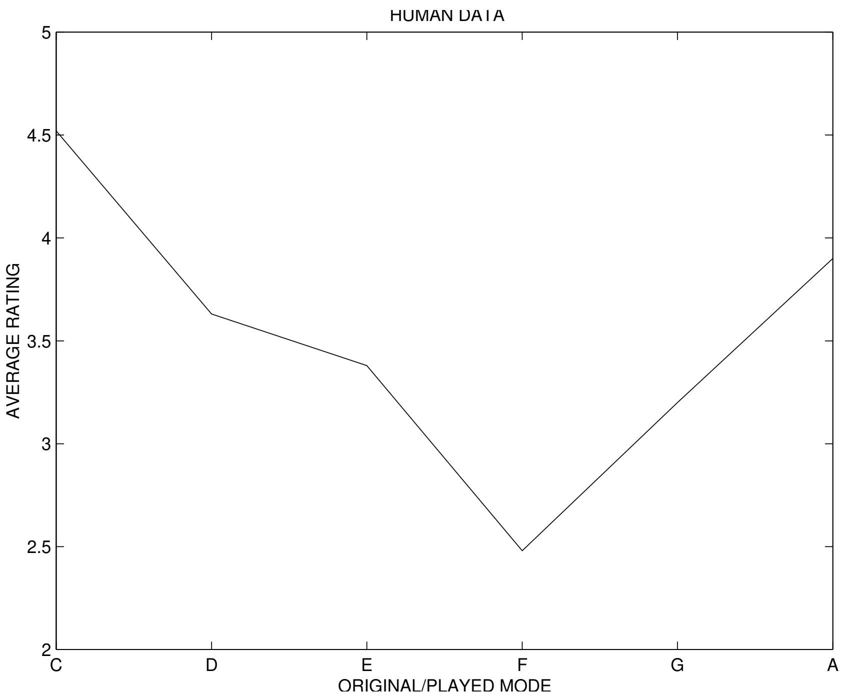


Figure 7.10: Average goodness of melodies played only in their original mode of composition, from human data.

The Pearson correlation between ARTIST's predictions (Figure 7.6) and humans' ratings (Figure 7.8) was low and did not reach significance, $r(5) = -.11,\ p > .10$. The mismatch between predicted and actual results for the A mode is mostly responsible for the insufficiency of the correlation: The melodies played in the A mode obtained the highest ratings from humans whereas ARTIST predicted the lowest ratings, in agreement with music theory. Besides this important discrepancy, examined in the general discussion, ARTIST's predictions were globally similar to humans' results. High ratings went to the C and D modes and lower ratings to the E, F and G modes. Still, the rank order of the modes predicted by ARTIST is not accurate, and the Spearman correlation coefficient computed on the rank of the modes was not significant either, $r(5) = -.09,\ p > .10$. ARTIST predicted the F mode to be slightly more pleasant than the E and G modes, whereas it was less pleasant to human ears than the other modes, and this contributed with the bad A mode prediction to make the Spearman correlation so low.

To test whether ARTIST's predictions other than for the A mode are realistic, they were compared to human data through a contrast analysis that was performed only on the 5 other modes. The set of coefficients used to represent the predictions in the contrast was the set of integers $(+5, +1, -2, -1, -3)$ and correlated almost perfectly with the original predictions, Pearson $r(3) = .999$. The quasi-F for the contrast between the predictions and human data for the 5 modes only was significant, $F(4, 59) = 5.85, MSe = 6.64, p < .05$.

Discussion

The results of the 2-way ANOVA performed on the original melodies only indicate that the effects of played mode and mode of composition uncovered by the 3-way ANOVA are not mere artifacts due to the transposition of the melodies to modes different from the original ones. It could have been that the mode of a melody and all its other features interact in such a strong way that just changing the mode would destroy the whole melodic Gestalt. But the results of the 2-way ANOVA rule out this possibility as an explanation for the effects of mode, because the effect is very similar without any change of mode.

That no main effect of musical expertise was found means that both groups of listeners centered their responses around the same value: The grand mean of the ratings equaled 3.33 (3.00 for experienced listeners vs 3.67 for unexperienced listeners) and was close to the middle of the rating scale 3.5, suggesting that both groups of subjects were able to use the rating scale properly.

The average ratings for the 6 played mode conditions only reflect the differences in tonality between the modes, since those 6 groups of melodies included the same original melodies, and differed only in terms of the pitch intervals occurring in the melodies. Therefore the main effect of played mode has to be explained by the differences in intervals between the modes. The preference for the melodies played in A, C and D modes was unexpected, considering that the theoretical predictions of Section 7.1 ranked the A and D modes towards the bottom of the goodness scale (see Figures 7.1 and 7.2). The high ratings for the major (C) mode were not surprising, even though this mode was expected to rank highest instead of numer two. The A and D modes probably received high ratings because they are identical to the natural and harmonic minor modes, respectively. This shows that the psychological distances between modes cannot be predicted from their similarity to the major mode only, and that similarities to the minor modes also need to be taken into account. This explanation is also supported if we consider that the F and G modes were not preferred to the E mode, contrary to what was expected from their respective similarities to the major mode. The E mode has only 1, 2 and 3 pitch classes different from those of the three minor modes, vs 2, 2 and 3 for the G mode and 2, 3 and 4 for the F mode. Therefore it appears that the distances to the minor modes also play a crucial role in determining the relative pleasantness of the modes.

Contrary to the played mode factor, the composition mode factor implies differences between groups of melodies based on all the features of melodies other than tonality (i.e., other than interval relationships). As mentioned before, those features can be the length, the tempo, the rhythm, the note density, the contour of the melodies, etc... Even though subjects were asked to base their judgments on tonal relationships alone, it must be virtually impossible to disregard completely the other aspects of the melodies. Consequently, there was a significant effect of composition mode. The order of preference for the composition modes was different from that for the played modes, but the general relationship was the same for both factors: A, C and D were rated above average, whereas E, F and G were rated below. This suggests that the melodies composed with the most pleasant set of intervals end up having the most pleasant contour, rhythm, etc...

This may be a natural consequence of the compositional processes used by composers; it may also be due to the fact that the melodies that had to be modified to make their mode unambiguous were mostly in the E, F and G modes, and that these modification were not done in a way as elegant as that of a professional composer. In any case, there seems to be a relationship between the purely tonal aspect of melodies and all the other features.

Musically experienced subjects gave higher ratings to the 3 best rated modes (C, A and D, rated higher than average) and lower ratings to the 3 lowest rated modes (E, F and G, rated lower than average) than unexperienced subjects did. Musicians' ratings were more differentiated as a function of composition mode, indicating that they may be more sensitive to the non-tonal aspects of melodies than nonmusicians. The reason might be that musicians have a more integrated perception of music than nonmusicians, that they have a greater tendancy to process musical stimuli wholistically. With a greater musical experience, they may have internalized the relationships between tonality and the other features to a greater extent than nonmusicians have.

Considering the relationship between tonal and nontonal features, it seems counter-intuitive that the interaction between played mode and composition mode did not even come close to being significant. However, it is perfectly possible for a melody to inherit from its original mode some goodness of form applying independantly to both tonal and nontonal features. These two orthogonal sets of features do not have to interact with each other, and the attractiveness of a melody can simply depend on the sum of their respective main effects.

7.4 General discussion

The contrast analysis shows that if we put aside the fact that ARTIST is totally confuzed regarding the Aeolian (A) mode, it gives a decent prediction of human ratings for the other modes: C was preferred to D, itself rated higher than the three other modes E, F and G. The Aeolian mode is the natural minor mode, one of the three minor modes that exist. Listeners are very familiar with this mode, and it is logical that melodies played in this mode obtained very high ratings. So why did ARTIST give it such a low rating?

The most probable explanation relates to the composition of the corpus used to trained ARTIST. In his preludes, Bach uses to a great extent the other forms of the minor mode, namely the harmonic and ascending melodic minor modes. As a consequence, ARTIST's learning of the minor mode was split between its three variations, and the natural minor probably did not have a chance to become as familiar as the major mode for ARTIST. This led to a substantially lower predicted rating for the minor mode. Another way of seeing this is that the mode of Aeolian melodies is quite ambiguous because of the great similarity between the three minor modes, and that ARTIST prefers stimuli that clearly fall in a familiar category rather than ambiguous stimuli that could fit in several categories. It is possible that the mode of an Aeolian melody is determined by the presence of one note occurring only once in the melody. As a consequence, the Aeolian melodies triggered high activations in the categories responsive to any minor mode. The number of highly activated categories increases as stimuli become more ambiguous and fit in more categories. Since we take the opposite of the total activation as an index of ARTIST's pleasantness rating, the unusually large number of categories activated by the ambiguity of the Aeolian melodies translates into a low rating for the A mode.

Note that this is the same principle that gives to the unfitting probe tones their low ratings, as seen in Chapter 4. For instance, a C major scale followed by the probe tone F# is ambiguous regarding the tonality, because it contains all the notes of the keys of C major and G major. It follows that more categories are activated in this case than in the cases where the probe tone fits with the context and does not conflict with the C major tonality. Consequently, we need to take the opposite of the total activation to get high ratings for 'good' stimuli and low ratings for ambiguous stimuli.

The interaction between musical expertise and mode played was not significant, so trying to simulate musicians' behavior by using Shepard tones may be irrelevant here. The hypothesis that musicians emphasize more the pitch class dimension than nonmusicians when processing musical stimuli was not verified. This could be because of the impossibility to disregard the rhythm, contour, and the other nontonal factors in a melodic context, as opposed to in a prototypical context of a scale and chord. This possibility is corroborated by the effect of composition mode, which is very reliable in spite of the instructions to the subjects to try to judge only the pitch dimension. Repeating the experiment using Shepard tones could increase the importance of the pitch class dimension and may validate ARTIST's predictions.

7.5 Conclusion

ARTIST is able to pick up on rather subtle differences between several forms of tonal music. The ratings given by humans to melodies played in different modes generally followed ARTIST's predictions, in spite of a large discrepancy concerning the Aeolian mode. It seems that the similarities between the three minor modes confuse ARTIST's judgements, but it is probable that ARTIST would differentiate between minor modes better with more learning or with a higher vigilance.

In spite of this shortcoming, ARTIST's predictions represent a great improvement from theoretical predictions (from music theory or from ARTIST with the mode profiles). Those theoretical predictions had very little in common with the actual results: only the high ratings for major mode (by hypothesis) and the low ratings for the E mode were accurate predictions. The ratings given to the modes when instantiated by melodies differ from the theoretical modal distances because melodies are more than a mere exhaustive list of pitch classes defining a key or a scale. Melodies repeat some notes more than others, give the notes different durations to create rhythms, alternate ascending and descending intervals to define a pitch contour, and do not have the pitch classes appearing in systematic orders. All this contributes to differentiating the melodies' ratings from the ratings obtained with scales or chords, which are in agreement with theoretical ratings. Like humans', ARTIST's ratings for melodies departed from the theoretical ratings, and the model was to a certain extent able to account for the different results obtained with real melodies.

If the results of this chapter focusing on the modes generalize to the results of Chapter 4 regarding tonality (vs atonality), then the latter results may need to be interpreted with more caution than previously thought. In Chapter 4, knowledge of the major and minor key profiles led to the estimations of all the major-major, minor-minor and major-minor perceptual key distances (Figures 4.8 to 4.10, respectively). The robustness of these results, along with their consistency with the key distances given by music theory, tend to make us think that those key distances apply to all musical situations. Even though the key distances measured by Krumhansl and Kessler (1982) were inferred from data obtained with prototypical sequences (scales, chords and chord progressions), it seemed likely, and was implicitely assumed, that the distances would be the same when the keys are instantiated by melodies instead of scales or chords. For example, we took for granted that the respective impressions of fluidity gathered from a melody modulating from C major to G major, or from the 2-chord sequence C major-G major were the same. But as we just saw with the modes, what is established from music theory and from experimental data with prototypical sequences may not always generalize to real-world music, where melodies are extensively used. This may be another example of the psychological importance of the temporal order of occurrence of the notes in a melody, a point often emphasized by Brown (1988) and Butler (1989).

CHAPTER 8

GENERAL CONCLUSION

The main contribution of this thesis is that it shows that ARTIST, a simple ANN built with minimal constraints, can extract a vast amount of musical knowledge similar to that of humans just by being exposed to a musical environment. Exposure to the music only assumes a coding of the music in terms of discrete pitches and an exponential timely decay of the input activations. The fact that the stimuli are directly derived from MIDI files emphasizes how close they are to the actual musical signal, and it is also a great advantage for further developments and applications of the model.

The model learns in a strictly unsupervised fashion, its architecture is self-organizing and its implementation is biologically plausible. These qualities make ARTIST a realistic model of human learning. It is able to mimic human behavior for very different types of tasks, which required differentiating between musical stimuli of varying degrees of tonality (Chapters 4 and 6) or between tonal melodies of different styles or modes (Chapter 7), or recognizing familiar melodies hidden among distractor notes (Chapter 3).

The most impressive performance of the model is the replication of Krumhansl and Kessler's (1982) measure of degree of tonality from human data on the probe-tone task, which established the tonal profiles of keys and permitted to infer the perceptual key distances: In one case, ARTIST's responses correlate 99% with human data. Furthermore, ARTIST is to my knowledge the first model to replicate those results solely on the basis of learning, without being explicitely given any knowledge relating to music theory. ARTIST was also given the 2-probe tone task, on which the pattern of responses show that it models the process of perceptual learning quite realistically. With more exposure to music, ARTIST goes through the same developmental stages as children as they grow up: The notion of in-key vs out-of-key is learned first, and only then are the triadic notes preferred to the other notes of the key (the diatonic notes). However, ARTIST never quite reached the adult-like stage where there is a preference for 2-note probes in the order least stable-most stable vs the most stable-least stable order. It is possible that ARTIST would start to exhibit order effects with even more exposure to music. After all, there was no hint of order effect in the human data before the adult stage. But increasing the rate of the temporal activation decay would certainly make ARTIST more sensitive to the order of two consecutive notes, since this would increase the difference in the activation levels corresponding to consecutive notes.

ARTIST also processes simple melodies in a way similar to humans. ARTIST's preference for melodies played in some particular modes resembles human preferences to some extent, showing that it is able to distinguish several categories of musical stimuli within tonal music. Further, it is able to distinguish different melodies within the same mode. It recognizes familiar melodies interleaved with distractor notes by spreading top-down activation when given the hypothetical identity of the melody, but not an unfamiliar melody even if it is in the same mode as the familiar melody.

All those results are promising for the future of ARTIST, and further development of the model could lead to useful musical abilities. Since ARTIST has internalized the rules of tonality, it should be possible to develop an algorithm to make it compose music. Manipulating the corpus used for learning could bias the composition towards different musical styles, or even towards a mix of musical styles never tried before. For instance, it could give us an idea of what a collaboration between Mozart and Hendrix would sound like. Another interesting application could be the development of an improvisation partner, which improvisation would be influenced in real-time by the other player(s).

Nevertheless, ARTIST has one major shortcoming in that it does not account for transpositional invariance, a basic property of humans' musical processing. It is very easy for humans to recognize melodies when they are played on different pitch levels, because the relationships between their constituent notes are the same and recognized immediately. Many models account for this property of our perceptual system to process inputs according to their relationships instead of according to the exact stimulation they elicit. However, those models all hypothesize the transpositional invariance and build the model according to this property (Deutsch and Feroe, 1981; Scarborough, Miller and Jones, 1989; Bharucha, 1991). According to my knowledge, no model to date acquired transpositional invariance through learning. Whether ARTIST would be able to do so by the addition of more layers of neurons to create more abstract categories remains to be seen.

To finally conclude, the general properties desirable for a good model outlined by Cross (1985, p.45—48) are summarized, as well as the way ARTIST conforms to those guidelines.

1. "Perceptible dimensions of musical experience should be involved in the creation of structures [, because] any perceptible feature may play a role". The formation of categories in ARTIST depends on pitch and metric (involving both duration and loudness) information.

2 and 3. "The listener brings to a piece of music a history of musical experience, which is itself a product of the musical history of his or her culture.[...] The extent to which [listeners are able to perceive music in terms of a fully coordinated structure] is likely to depend on the music itself, their previous experience of it, and their experience with music of the same style or idiom". ARTIST interprets any piece of music in relation to all the music it has been exposed to, and the activation of many very different categories following unfamilar kinds of musical inputs suggests that ARTIST does not 'perceive' such inputs as a fully coordinated structure.

4. "The output of the model should be related to some aspect or aspects of judgement or behaviour relating to music." ARTIST's output relates to ratings of pleasantness of musical sequences or to the recognition of familiar melodies.

5. "Structure often can be understood, at least partly, by reference to an extra-musical or historical context. For example, [...] music is associated with [...] dancing". As up to now, ARTIST's universe is purely limited to music so no reference to extra-musical context is possible.

6. “It is necessary to model both horizontal and vertical structure.” ARTIST’s input accepts any number of notes to be played simultaneously (vertical dimension) and/or sequentially (horizontal dimension).

7. “It is possible to identify a few global factors, whose operation could account for much of the patterning of musical sound [(e.g., Gestalt principles)].” ARTIST does not conform to this principle, as was illustrated by the discussion above regarding transpositional invariance, but expanding the model may allow ARTIST to identify those ‘global factors’ by itself.

8. “Groups [may be formed] and may combine to form higher-order groups”. ARTIST’s categories are a way of grouping the stimuli according to the similarity of their features. Adding more abstract layers on top of the one already existing would provide ARTIST with the structures necessary to form higher-order groups.

In summary, ARTIST’s relevance as a model is not only psychological, but also theoretical as can be seen from its conformance to those principles.

APPENDIX A

BASIC WESTERN MUSICOLOGICAL CONCEPTS

The reader is encouraged to follow the definitions on a keyboard if possible, and play along as the intervals and scales are introduced. Not only will this give the explanations their perceptual counterpart and make them easier to follow, it will also prime the reader for the explanations of the probe-tone technique and of the tone profiles. Figures A and B show an example for each definition given thereafter.

Pitch classes, octave equivalence and semitones

There is a long history of different tuning systems used in Western music. The tuning system refers to the conventions used to constrain the relationships between note frequencies. Some tuning systems are based on the simplicity of frequency ratios, whereas others emphasize more the equal spacing of the frequencies (on a log-scale). The well-tempered tuning system is of the latter kind and accounts for almost all of western music since Bach. While some of the following definitions are correct regardless of the tuning system, some apply only to the well-tempered system.

Western music uses notes that can be categorized into twelve pitch classes (also called pitches): C,C# or Db, D , D# or Eb, E , F , F# or Gb , G , G# or Ab , A , A# or Bb, and B, where C# denotes the category 'C sharp' and Db denotes the category 'D flat'. Sometimes pitch class is also called chroma and another name for the set of twelve pitches is the chromatic scale, arranged around the chroma circle. The notes with a sharp or flat are said to be altered. All the notes of a given pitch belong to the same class in the sense that they have very similar perceptual attributes, even though their fundamental frequencies and perceived heights are different. For instance, the notes C1, C2, C3, ... belong to the same pitch class C. The fundamental frequencies of two successive Cs always have the same relationship, a frequency ratio equal to 1:2. Thus, all the notes of a particular pitch class have their frequencies linearly spaced on a logarithmic scale, and they are said to share the property of octave equivalence. The span of an octave contains 12 notes (one for each pitch class) equally spaced on a log-frequency scale. The interval between any two consecutive notes is therefore constant on the log-frequency scale and is called a semitone. The semitone is the smallest interval between two notes used in Western music, and is the unit most commonly used to measure intervals between two notes. Because the range delimited by two frequencies in a 1:2 ratio is divided in twelve, the frequency ratio of two consecutive notes is $2^{1/12}$ or approximately 1.06. For instance, one semitone separates C from C# (which frequency is 1.06 times C's), whereas there are 5 semitones between C and F. The names commonly used for the intervals smaller than an octave are given in the figure. The interval of a fifth is special in that it is the most consonant of the intervals involving two different pitches (the unison and the octave involve two notes having the same pitch), presumably because the frequency ratio of the two notes is one of the most simple fraction, being very close to 3:2 ($2^{7/12}$ exactly. Unison is 1:1 and octave is 2:1).

In summary, the pattern of 12 pitches repeats itself from octave to octave, as the note frequencies double. The perception of pitch has often been described as 2-dimensional, one dimension being the height and the other being the chroma.

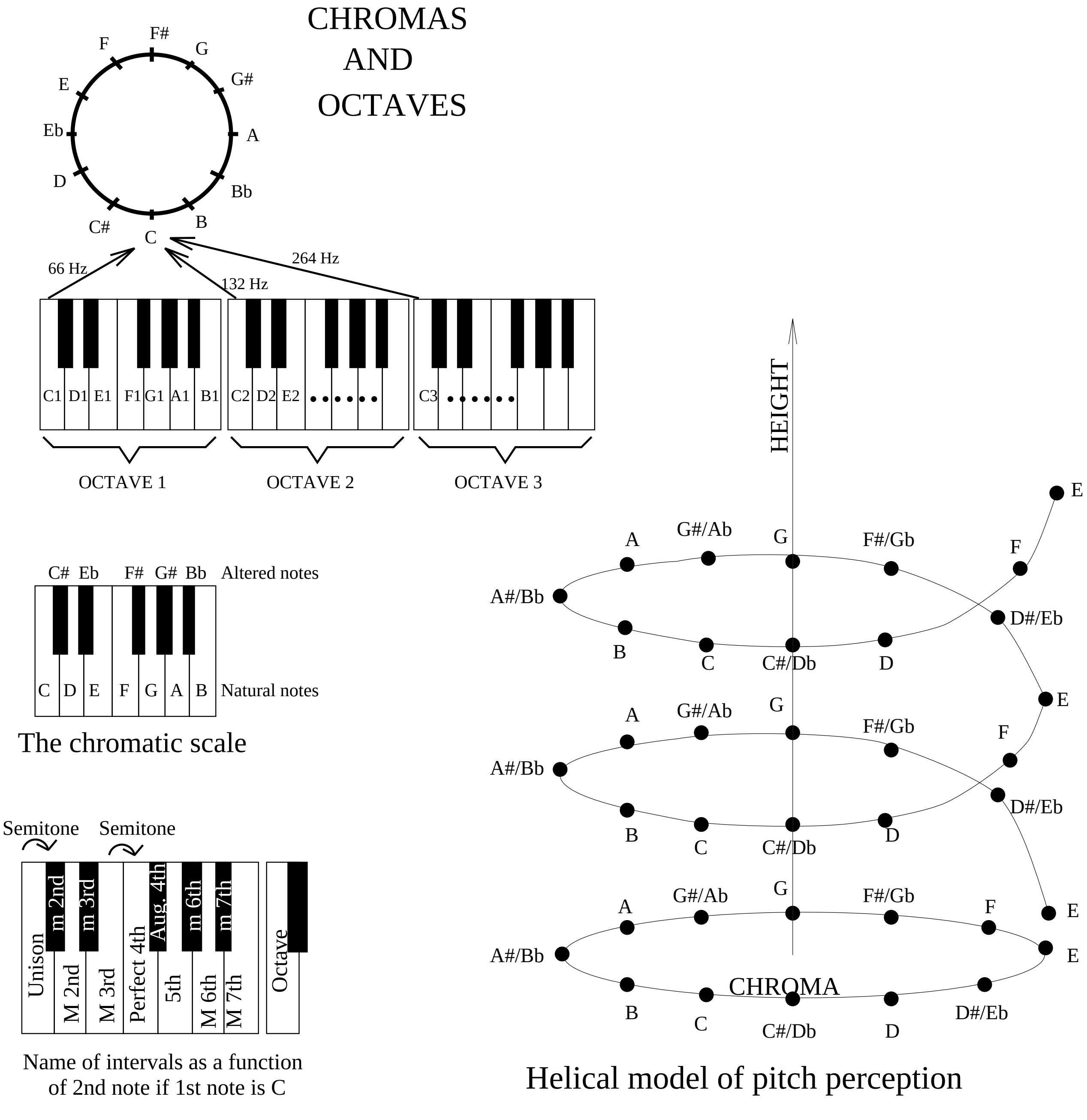


Figure A: Chromatic scale, Chromas, octave equivalence, intervals and the helical model of pitch perception.

A helical model of tones perception captures well the correspondence between those two dimensions and the two salient psychological features of note perception, and has been proposed with slight variations successively by Ruckmick (1929), Bachem (1950), Revesz (1954), Shepard (1964, 1982) and Pilker (1966), all cited by Krumhansl (1990). As can be seen on the figure, the vertical axis corresponds to note height and the horizontal plane to the circularity of chroma.

Major and minor scales

The major and minor scales (also called keys, or tonalities) are defined by taking a subset of 7 pitches out of the 12 constituting the chromatic scale. For the 7 pitches to define a major scale, their consecutive intervals must be (2,2,1,2,2,2,1) in number of semitones. Huron (1994) showed that this maximizes the number of consonant intervals available. Still, we do not know whether this scale is the most used because of its consonance or vice-versa, since the perception of consonance was measured with subjects that had been mostly exposed to that scale. For example, choosing to start at C (as the origin of the scale, it is called the tonic or root), the notes D (C + 2 semitones), E (D + 2 semitones), F (E + 1 semitone), G (F + 2 semitones), A (G + 2 semitones), B (A + 2 semitones) and C (B + 1 semitone) will belong to the scale. Thus, the scale of C major is made of the notes (C,D,E,F,G,A,B, also called the diatonic set) and is the only major scale with only non-altered notes, because if we choose any other tonic than C (we start the scale at any other place), following the semitones sequence (2,2,1,2,2,2,1) will lead us to include a # note in the scale. The two major scales most similar to C major are G major (G,A,B,C,D,E,F#) and F major (F,G,A,Bb,C,D,E), in the sense that they share 6 pitches out of 7 with C major. B major is one of the least similar to C major, sharing only 2 pitches out of 7.

Regarding the minor keys, there exist three different minor modes. The most common nowadays is the harmonic minor, corresponding to the sequence of semitones (2,1,2,2,1,3,1). Choosing C as the tonic, we find that (C,D,Eb,F,G,Ab,B) constitute the C minor scale. A-minor is defined by the pitches (A,B,C,D,E,F,G#) and shares all pitches but one with C major, just like F and G major.

Key distances

The almost total overlap between C major and other scales such as A minor, F or G major has important psychological implications. The psychological distance between two keys depends on the number of pitches they share, and so the perceived distance between those just mentioned is quite small. As a consequence, a change of key (also called modulation) from one to the other will sound very smooth, logical, like a natural evolution of the passage. Indeed, they constitute very common modulation, perhaps the most frequently found across all styles of Western music. Another implication of the resemblance between two keys is that some musical passage can be ambiguous regarding its tonality, evoking both keys at the same time. Sometimes knowing the key in which a passage is written is a matter of minute detail, and of fight between musicologists; roughly, it is the occurrence and timing (on a strong or weak beat) of the tonic that will instantiate one key more strongly than the other, but many other parameters can also play a role.

Repeating the process of counting the number of overlapping pitches for all 12 possible major keys, we can summarize the distances between any two keys by organizing them around a circle. It is called the circle of fifths because two adjacent keys are related by an interval of a fifth. It is important to note that the circle of fifths reflects both the psychological distance between keys and their number of common pitches. The two are confounded, making it difficult or sometimes impossible to distinguish music theoretical concepts from psychological ones, each kind having its counterpart in the other domain.

Chords

Several notes played simultaneously form a chord. We will only mention the two most basic types of chords here. They are the major and minor chords, made of three notes. A set of three notes (not necessarily played together) is called a triad, and two examples of triads are given.

As previously mentioned, the interval of a fifth is especially consonant, so the fifth will be present in the chords along with the root (or tonic, the note giving its name to the chord). That is, G (a fifth above C) is present in the chords built around C (C major and C minor). The last note sounded in the chords is a third above the root. If the interval defined by these two notes is a major third (like between C and E), the chord is major. If it is a minor third (like between C and Eb), the chord is minor. Thus, C major contains the notes C,E,G whereas C minor contains C,Eb,G. In summary, major and minor chords differ in the place of one note (the third), which two possible places are one semitone apart from each other (E and Eb are right next to each other on a keyboard).

Two other types of triads, quite rare relative to the major and minor triads, are sometimes used in experiments, usually to contrast their effects with the latter triads. They are the augmented and diminished triads. The notes of the augmented triad define the successive intervals of a major third and a fourth. Those of the diminished triad are separated by minor thirds. This gives for C (C,E,G#) and (C,Eb,F#) respectively.

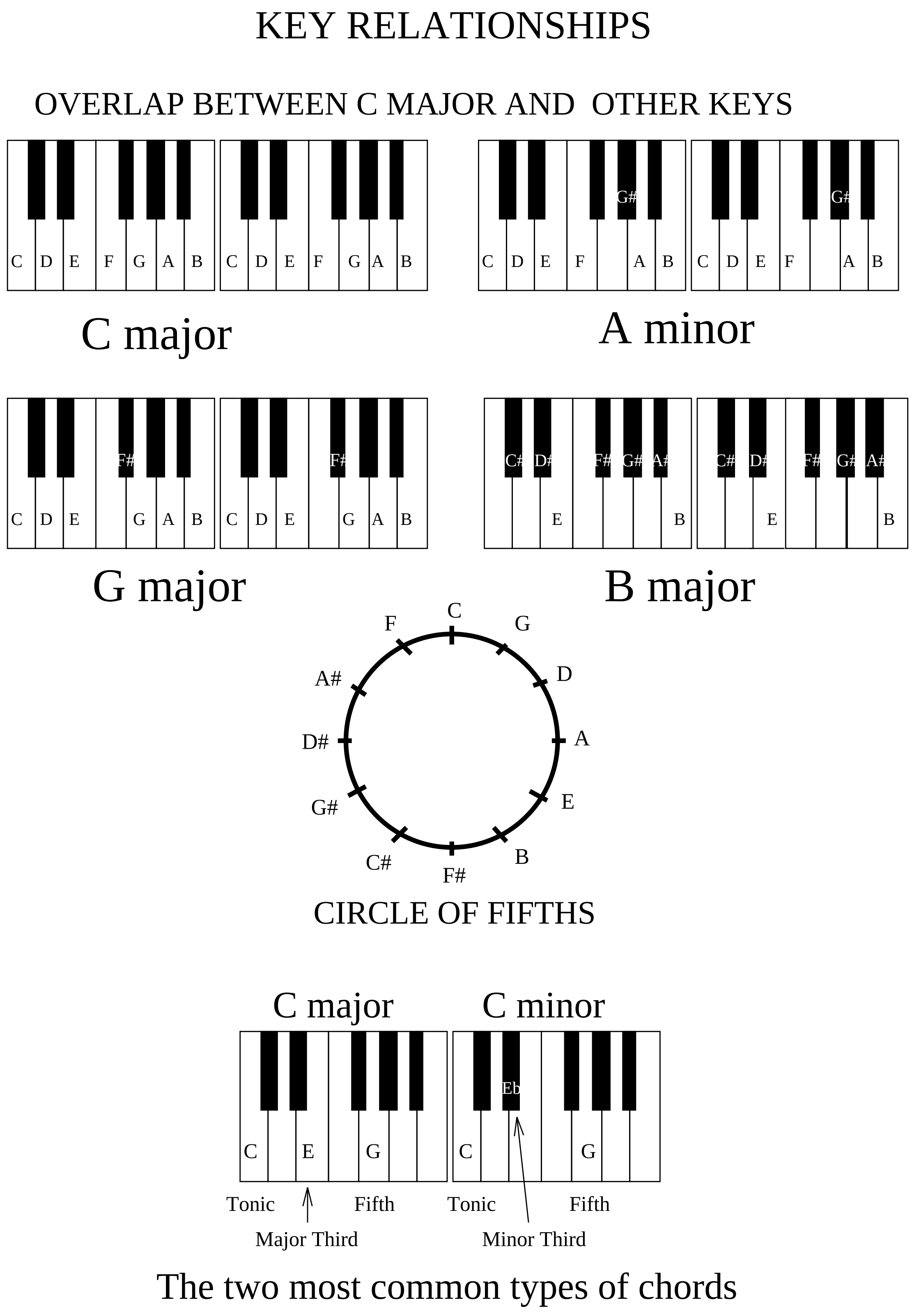


Figure B: Scales, circle of fifths and chords.

REFERENCES

Andrews, M. W., & Dowling, W.J. (1991). The development of perception of interleaved melodies and control of auditory attention. *Music Perception*, 8(4), 349-368.

Baddeley, A. (1990). Human Memory. Boston: Allyn and Bacon.

Bartlett, F.C. (1932). *Remembering.* Cambridge: Cambridge University Press.

Bartlett, J.C., & Dowling, W.J. (1980). Recognition of transposed melodies: a key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance*, 6(3), 501-515.

Bartlett, J.C., & Dowling W.J. (1988). Scale structure and similarity of melodies. *Music Perception*, 5, 285-314.

Berz, W.L. (1995). Working memory in music: A theoretical model. *Music Perception*, 12, 353-364.

Bharucha, J.J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception*, 5(1), 1-30.

Bharucha, J.J. (1996). Melodic anchoring. *Music Perception*, 13(3), 383-400.

Bharucha, J.J. (1991). Pitch, harmony, and neural nets: A psychological perspective. In P.M.Todd & D.G.Loy (Eds)., *Music and connectionism.* Cambridge: The MIT Press.

Bharucha, J.J., & Todd, P.M. (1989). Modeling the perception of tonal structure with neural nets. *Computer Music Journal*, 13(4).

Bigand, E. (1993). Contributions of music to research on human auditory cognition. In S.McAdams & E.Bigand (Eds.), Thinking in sounds: the cognitive psychology of human audition. Oxford: Clarendon Press.

Bregman, A.S. (1990). Auditory scene analysis. Cambridge, MA: MIT Press.

Brown, H. (1988). The interplay of set content and temporal context in a functional theory of tonality perception. *Music Perception*, 5, 219-250.

Brown, H. & Butler, D. (1981). Diatonic trichords as minimal tonal cue-cells. *In theory only*, 5 (6 & 7), 37-55.

Burns, E.M. (1974). Octave adjustment by non-Western musicians. *Journal of the Acoustical Society of America*, 56, S25.

Butler, D. (1989). Describing the perception of tonality in music: A critique of the tonal hierarchy theory and a proposal for a Theory of Intervallic Rivalry. *Music Perception*, 6(3), 219-241.

Carpenter, G.A. (1989). Neural networks models for pattern recognition and associative memory. Neural networks, 2, 243-257.

Carpenter, G.A., & Grossberg (1987). ART2: Self-organization of stable category recognition codes for analog input patterns. Applied optics, 26, 4919-4930.

Carpenter, G.A., Grossberg S., & Rosen, D.B. (1991). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, Vol.4, 759-771.

Carpinteiro, O.A.S. (1996). A connectionist approach in music perception. Tech. rep. CSRP 426, School of cognitive and computing sciences — University of Sussex, Falmer, UK.

Cauller, L. (1995). Layer I of primary sensory neocortex: where top-down converges upon bottom-up. *Behavioural brain research*, 71, 163-170.

Cross, I., Howell, P., & West, R. (1985). Structural relationships in the perception of musical pitch. In P.Howell, I.Cross & R.West (Eds.), *Musical structure and cognition.* London: Academic Press, pp.121-142.

Crowder, R.G. (1993). Auditory memory. In S.Mc Adams & E.Bigand (Eds.), Thinking in sounds: the cognitive psychology of human audition. Oxford: Clarendon Press.

Cuddy, L.L. (1993). Melody comprehension and tonal structure. In T.J.Tighe & W.J.Dowling (Eds.), Psychology and music: the understanding of melody and rhythm. Hillsdale, NJ: Erlbraum.

Cuddy, L.L., & Badertscher, B. (1987). Recovery of the tonal hierarchy: Some comparisons across age and levels of musical experience. *Perception and Psychophysics*, 41, 609-620.

Dainow, E. (1977). Physical effects and motor responses to music. *Journal of Research in Music Education*, 25, 211-221.

Deliège, I. (1987). Grouping conditions in listening to music: an approach to Lerdahl and Jackendoff's grouping preference rules. *Music Perception*, 4, 325-360.

Deliège, I., Melen, M., Stammers, D., & Cross, I. (1996). Musical schemata in real-time listening to a piece of music. *Music Perception*, 14(2), 117-160.

Demany, L., & Armand, F. (1984). The perceptual reality of tone chroma in early infancy. *The journal of the Acoustical Society of America*, 76(1), 57-66.

Desaing, P., & Honing, H. (1989). The quantization of musical time: a connectionist approach. *Computer Music Journal*, 13(3).

Deutsch, D. & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. Psychological Review, 88, 503-522.

Dibben, N. (1994). The cognitive reality of hierarchic structure in tonal and atonal music. *Music Perception*, 12(1), 1-25.

Dodson, C.S., Johnson, M.K., & Schooler, J.W. (1997). The verbal overshadowing effect: why descriptions impair face recognition. *Memory & Cognition*, 25 (2), 129-139.

Dowling, W.J. (1973). The perception of interleaved melodies. *Cognitive psychology*, 5, 322-337.

Dowling, W.J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, 85, 341-354.

Dowling, W.J. (1990). Expectancy and attention in melody perception. *Psychomusicology*, 9(2), 148-160.

Dowling, W.J. (1991). Tonal strength and melody recognition after long and short delays. *Perception and Psychophysics*, 50, 305-313.

Dowling, W.J., & Bartlett, J.C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology*, 1, 30-49.

Dowling W.J., & Fujitani, D.S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America*, 49, 524-531.

Dowling, W.J., & Harwood, D.L. (1986). *Music cognition.* Orlando, FL: Academic Press.

Dowling, W.J., & Hollombe, A.W. (1977). The perception of melodies distorted by splitting into several octaves: Effects of increasing proximity and melodic contour. *Perception and Psychophysics*, 21, 60-64.

Dowling, W.J., Kwak, S., & Andrews, M.W. (1995). The time course of recognition of novel melodies. *Perception and Psychophysics*, 57 (2), 136-149.

Dowling W.J., Lung, K.M.-T., & Herrbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies. *Perception and Psychophysics*, 41, 642-656.

Drake, C., & Palmer, C. (1993). Accent structures in music performance. *Music Perception*, 10(3), 343-378.

Fodor, J.A. (1983). The modularity of mind. Cambridge: The MIT Press.

Francès, R. (1958). The perception of music. (Translated by W.J.Dowling)

Frankland, B., & Cohen, A.J. (1990). Expectancy profiles generated by major scales: group differences in ratings and reaction time. *Psychomusicology*, 9(2), 173-192.

Gabrielsson, A., & Lindstrom, S. (1994). Strong experiences of music - A manifold miracle. Proceedings of the 3rd ICMPC in Liege, July 23rd-27th.

Gibson, J.J. (1966). The senses considered as perceptual systems. Boston: Houghton Mifflin Company.

Gibson, J.J. (1979). The ecological approach to visual perception. Boston: Houghton Mifflin Company.

Gjerdingen, R.O. (1990). Categorization of musical patterns by self-organizing neuronlike networks. *Music Perception*, 7,339-370.

Greenwood, D. (1991) Critical bandwidth and consosnance in relation to cochlear frequency-position coordinates. *Hearing research*, 54(2), 164-208.

Griffith, N.J.L. (1993). Modelling the acquisition and representation of musical tonality as a function of pitch-use through self-organising artificial neural networks. Unpublished doctoral dissertation, University of Exeter

Grossberg, S. (1982). *Studies of mind and brain: Neural principles of learning, perception, development, cognition and motor control.* Boston:D.Reidel/Kluwer.

Handel, S. Perceiving melodic and rhythmic auditory patterns. *Journal of Experimental Psychology*, 103, 922-933.

Helmoltz, H.L.F. (1954) *On the sensations of tones...* (A.J.Ellis, Ed. & trans.). New York: Dover. (Revised edition originally published, 1885).

Huron, D. (1989). Voice denumerability in polyphonic music of homogeneous timbres. *Music Perception*, 6(4), 361-382.

Huron, D. (1994). Interval-class content in equally tempered pitch-class sets: Common scales exhibit optimum tonal consonance. *Music Perception*, 11(3), 289-305.

Huron, D., & Fantini, D.A. (1989). The avoidance of inner-voice entries: Perceptual evidence and musical practice. *Music Perception*, 7(1), 43-48.

Janata, P., & Reisberg, D. (1988). Response-time measures as a means of exploring tonal hierarchies. *Music Perception*, 6(2), 161-172.

Jarvinen, T. (1995). Tonal hierarchies in Jazz improvisation. *Music Perception*, 12(4), 415-437.

Jones, M.R. (1981). Music as a stimulus for psychological motion: Part I. Some determinants of expectancies. Psychomusicology, 1(2), 34-51.

Jones, M.R. (1986). Attentional rhythmicity in human perception. In J.R.Evans & M.Clynes, eds., *Rhythm in psychological, linguistic and musical processes.* Springfield, IL: Charles C. Thomas.

Jones, M.R. (1993). Dynamics of musical patterns: how do melody and rhythm fit together? In T.J.Tighe & W.J.Dowling (Eds.), Psychology and music: the understanding of melody and rhythm. Hillsdale, NJ: Erlbraum.

Jones, M.R., Boltz, M., & Kidd, G. (1982). Controlled attending as a function of melodic and temporal context. Perception and Psychophysics, 32, 211-218.

Jones, M.R., & Yee, W. (1993). Attending to auditory events: the role of temporal organization. In S.Mc Adams & E.Bigand (Eds.), Thinking in sounds: the cognitive psychology of human audition. Oxford: Clarendon Press.

Katz, B.F. (1995). Harmonic resolution, neural resonance, and positive affect. *Music Perception*, 1995, 13(1), 79-108.

Kessler, E.J., Hansen, C., & Shepard, R.N. (1984). Tonal schemata in the perception of music in Bali and in the West. *Music Perception*, 2(2), 131-165.

Knopoff, L., & Hutchinson, W. (1983). Entropy as a measure of style: The influence of sample length. Journal of Music Theory, 27, 75-97.

Krumhansl, C.L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, 11, 346-374.

Krumhansl, C.L. (1983). Perceptual structures for tonal music. *Music Perception*, 1(1), 28-62.

Krumhansl, C.L. (1990). The cognitive foundations of musical pitch. Oxford psychology series, No. 17.

Krumhansl, C.L., & Keil, F.C. (1982). Acquisition of the hierarchy of tonal functions in music. *Memory and cognition*, 10, 243-51.

Krumhansl, C.L., & Kessler, E.J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 89, 334-368.

Krumhansl, C., & Shepard, R. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 579-594.

Laden, B. (1995). Modeling cognition of tonal music. *Psychomusicology*, 14, 154-172.

Laden, B., & Keefe, D.H. (1989). The representation of pitch in a neural net model of chord classification. *Computer Music Journal*, 13(4).

Lamont, A., & Cross, I. (1994). Children's cognitive representations of musical pitch. *Music Perception*, 12(1), 27-55.

Leman, M. (1991). The ontogenesis of tonal semantics: Results of a computer study. In P.M.Todd & D.G.Loy (Eds)., *Music and connectionism.* Cambridge: The MIT Press.

Leman, M. (1995). A model of retroactive tone-center perception. *Music Perception*, 12(4), 439-471.

Lerdahl, F., & Jackendoff, R. (1983) *A generative theory of tonal music.* Cambridge, Mass.: M.I.T. Press.

Lerdahl, F., & Jackendoff, R. (1983). An overview of hierarchical structure in Music. *Music Perception*, 1(2), 229-252.

Lewis, J.P. (1991). Creation by refinement and the problem of algorithmic music composition. In P.M.Todd & D.G.Loy (Eds)., *Music and connectionism.* Cambridge: The MIT Press.

Luriia, A.R. (1968). The mind of a mnemonist. Translated from Russian by Lynn Solotaroff. New York, Basic Books.

Lynch, M.P., Eilers, R.E., Oleer, D.K., & Urbano, R.C. (1990). Innateness, experience, and music perception. *Psychological Science*, 1, 272-276.

Mandler, G. (1982). The structure of value: Accounting for taste. In M.S. Clark & S.T. Fiske, eds., *Affect and cognition: The Seventeenth Annual Carnegie Symposium on Cognition.* Hillsdale, NJ: Lawrence Erlbraum Associates.

Martin, R.C., Wogalter, M.S., & Forlano, J.G. (1988). Reading comprehension in the presence of unattended speech and music. Journal of Memory and Language, 27, 382-398.

McCulloch, W.S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of mathematical biophysics*, 5, 115-133.

Meyer, L.B. *Emotion and meaning in music.* Chicago: University of Chicago Press, 1956.

Miller, (1956) The magic number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.

Miller, G.A., & Chomsky, N. (1963). Finitary models of language users. In R.D.Luce, R.R.Bush & E.Galanter (Eds.), *Handbook of mathematical psychology, Vol.II.* New York: Wiley, pp. 419-491.

Miyazaki, K. (1993a). Absolute pitch as an inability: Identification of musical intervals in a tonal context. *Music Perception*, 11, 55-72.

Miyazaki, K. (1993b). Recognition of transposed melodies by absolute pitch listeners. Proceedings of the 3rd ICMPC, I. Deliège (Ed.).

Mozer, M.C. (1991). Connectionist music composition based on melodic, stylistic, and psychological constraints. In P.M.Todd & D.G.Loy (Eds)., *Music and connectionism.* Cambridge: The MIT Press.

Narmour, E. (1983). Some major theoretical problems concerning the concept of hierarchy in the analysis of tonal music. *Music Perception*, 1(2), 129-199.

Narmour, E. (1990). The analysis and cognition of basic melodic structures: The implication-realization model. Chicago: University of Chicago Press.

Narmour, E. (1991). The Top-Down and Bottom-Up systems of musical implication: building on Meyer's theory of emotional syntax. Music Perception, 9(1), 1-26.

Narmour, E. (1992). The analysis and cognition of melodic complexity: The implication-realization model. Chicago: University of Chicago Press.

North, A.C., & Hargreaves, D.J. (1995). Subjective complexity, familiarity, and liking for popular music. Psychomusicology, 14, 77-93.

Panksepp, J. (1995). The emotional sources of 'chills' induced by music. *Music Perception*, 13(2), 171-207.

Parker, C. (1950). Leap frog. On *bird and diz* [CD]. New York: Verve. (issued 1956; reissued 1986). "Leap frog" (1950, track 8).

Patel, A.D., & Peretz, I. (1997). Is music autonomous from language? A neuropsychological appraisal. In *Perception and cognition of music*, Chapter 10, eds I. Deliège & J. Sloboda. Psychology press, Hove.

Plomp, R., & Levelt, W.J.M. (1965). Tonal consonance and critical bandwidth. *The journal of the Acoustical Society of America*, 38, 548-560.

Randel, D.M. (Ed.). (1978). *Harvard Concise Dictionary of Music.* Cambridge, Mass: Harvard University Press.

Repp, B.H. (1996). The Art of inaccuracy: Why pianists' errors are difficult to hear. *Music Perception*, 14(2), 161-184.

Roediger III, H.L. (1990). Implicit memory: retention without remembering. *American Psychologist*, 45, 1043-56.

Sano, H, & Jenkins, B.K. (1989). A neural network model for pitch perception. *Computer Music Journal*, 13(3).

Sayegh, S.I. (1989). Fingering for string instruments with the optimum path paradigm. *Computer Music Journal*, 13(3).

Scarborough, D.L., Miller, B.O., & Jones, J.A. (1989). Connectionist models for tonal analysis. *Computer Music Journal*, 13(3).

Schoenberg, A. (1969). *Structural functions of harmony* (Rev. ed.). New York: Norton.

Serafine, M.L., Glassman, N., & Overbeeke, C. The cognitive reality of hierarchic structure in music. *Music Perception*, 1989, 6(4), 397-430.

Shepard, R.N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological review*, 91(4), 417-447.

Shepard, R.N. (1964). Circularity in judgments of relative pitch. *The journal of the Acoustical Society of America*, 36, 2346-53.

Simon, H.A. & Sumner, R.K. (1968). Pattern in music. In B. Kleinmuntz (Ed.), Formal representation of human judgment. New York: Wiley.

Sloboda, J.A. (1985). The musical mind. Oxford psychology series, No.5.

Sloboda, J.A. (1991). Music structure and emotional response: some empirical findings. Psychology of music, 19, 110-120.

Sloboda, J.A., & Edworthy, J. (1981). Attending to two melodies at once: the effect of key relatedness. Psychol. Mus. 9, 39-43.

Smith, J.D., Kemler Nelson, D.G., Grohskopf, L.A., & Appleton, T. (1994). What child is this? What interval was that? Familiar tunes and music perception in novice listeners. Cognition, 52, 23-54.

Smith, J.D. & Melara, R. (1990). Aesthetic preference and syntactic prototypicality in music: 'Tis the gift to be simple. Cognition, 34, 279-298.

Speer, J.R.,& Meeks, P.U. (1985). School children's perception of pitch in music. *Psychomusicology*, 5, 49-56.

Terhardt, E. (1984). The concept of musical consonance: A link between music and psychoacoustics. *Music perception*, 1(3), 276-295.

Todd, P.M. (1989). A connectionist approach to algorithmic composition. *Computer Music Journal*, 13(4).

Toiviainen, P. (1995). Modeling the target-note technique of Bebop-style Jazz improvisation: an artificial neural network approach. *Music Perception*, 12(4), 399-413.

Trainor, L.J., & Trehub, S.E. (1992). A comparison of infants' and adults' sensitivity to Western musical structure. *Journal of Experimental Psychology: Human Perception and Performance*, 18(2), 394-402.

Trainor, L.J., & Trehub, S.E. (1993). Musical context effects in infants and adults: Key distance. *Journal of Experimental Psychology: Human Perception and Performance*, 19(3), 615-626.

Trainor, L.J., & Trehub, S.E. (1994). Key membership and implied harmony in Western tonal music: Developmental perspectives. *Perception & Psychophysics*, 56(2), 125-132.

Trehub, S.E., Cohen, A.J., Thorpe, L.A., & Morrongiello, B.A. (1986). Development of the perception of musical relations: Semitone and Diatonic structure. *Journal of Experimental Psychology: Human Perception and Performance*, 12(3), 295-301.

Trehub, S.E., Thorpe, L.A., & Trainor, L.J. (1990). Infant's perception of *good* and *bad* melodies. *Psychomusicology*, 9, 5-19.

Trehub, S.E., & Unyk, A.M. (1991). Music prototypes in developmental perspective. *Psychomusicology*, 10, 31-45.

Vos, P.G., & Troost, J.M. (1989). Ascending and descending melodic intervals: Statistical findings and their perceptual relevance. *Music Perception*, 6(4), 383-396.

Ward, W.D. (1954). Subjective musical pitch. *Journal of the Acoustical Society of America*, 26, 369-380.

Watkins, M. & Watkins, O. (1974). Processing of recency items for free recall. *Journal of Experimental Psychology*, 102, 488-493.

Winograd, T. (1968). Linguistics and the computer analysis of tonal harmony. Journal of Music Theory, 12, 3-49.

Youngblood, J.E. (1958). Style as information. Journal of Music Theory, 2, 24-35.

Younger, B., & Gotlieb, S. (1988). Development of categorization skills: Changes in the nature or structure of infant form categories. *Developmental Psychology*, 24, 611-619.

VITA

Frédéric Georges Paul Piat was born on the 7th of August 1968 in Dijon, France, the son of Christian Binot and Jacqueline Piat. He obtained a M.S. in Computer Sciences at the University of Burgundy in Dijon in 1991 and a DEA in Artificial Intelligence and Production Sciences at the University of Besancon in 1992 while working as a programmer for the educational software company InovaSys. He started as a teaching assistant in 1993 at the University of Texas in Dallas, where he obtained a M.S. in Applied Cognition and Neuroscience in 1998 and where he is a PhD candidate in the Human Development and Communication Disorders program. His son Guilhem Piat was born August 15, 1995. He will hold a post-doctoral position at the National Technical University of Athens, Greece, in 1999.