In : Plack, Oxenham, Fay & Popper A. (Eds) Pitch Perception, Springer Verlag, SHAR (in press)

Chapter 9

Effect of Context on the Perception of Pitch Structures

E. Bigand (1) & B. Tillmann (2)

(1) Laboratoire d'Etude de l'Apprentissage et du Développement

CNRS UMR 5022, Université de Bourgogne

Dijon, France

(2) Neurosciences et Systèmes Sensoriels

CNRS UMR 5020

Lyon, France

Running title. Context effects on pitch perception

Emmanuel Bigand (corresponding author)

Université de Bourgogne

CNRS UMR 5022 Laboratoire d'Etude de l'Apprentissage et du Développement

Boulevard Gabriel; F-21000 Dijon, France

Tel. +33 (0)380395782; Fax. +33 (0)380395767; bigand@u-bourgogne.fr

Barbara Tillmann, CNRS UMR 5020 Neurosciences et Systèmes Sensoriels

50 Av. Tony Garnier, F-69366 Lyon Cedex 07 France

Tel: +33 (0) 4 37 28 74 93; Fax: +33 (0) 4 37 28 76 01;

barbara.tillmann@olfac.univ-lyon1.fr

1. Introduction

Our interaction with the natural environment involves two broad categories of processes to which cognitive psychology refers as *sensory-driven processes* (also called *bottom-up processes*) and *knowledge-based processes* (also called *top-down processes*). Sensory-driven processes extract information relative to a given signal by considering exclusively the internal structure of the signal. Based on these processes, an accurate interaction with the environment supposes that external signals contain enough information to form adequate representations of the environment and that this information is neither incomplete nor ambiguous. Several models of perception have attempted to account for human perception by focusing on sensory-driven processes. Some of these models are well known in visual perception (Marr 1982; Biederman 1987), as well as in auditory perception (see de Cheveigné, Chapter 6) and, more specifically, music perception (Leman 1995; Carreras et al. 1999; Leman et al. 2000). For example, Leman's model (2000) describes perceived musical structures by considering uniquely auditory images associated with the musical piece. The model comprises of a simulation of the auditory periphery, including outer and middle ear filtering and cochlea's inner hair cells, followed by a periodicity analysis stage that results in pitch images, and which are stored in short-term memory. These pitch patterns are then fed into a self-organizing map that infers musical structures (i.e., keys).

Sensory-driven models have been largely developed in artificial systems. They capture important aspects of human perception. The major problem encountered by these models is that environmental stimuli generally miss some crucial information required for adapted behavior. Environmental stimuli are usually incomplete, ambiguous, always changing from one occurrence to the next, and their psychological meaning changes as a function of the overall context in which they occur. For example, a small round orange object would be identified as a tennis ball in a tennis court, but as a fruit in a kitchen, and the other way round as an orange in a tennis court when the tennis player starts to peel it, or as a tennis ball in a kitchen when a child plays with it. A crucial problem for artificial systems of perception consists in formalizing these effects of context on object processing and identification. A fast and accurate adaptation to the everyday-life environment requires the human brain to analyze signals on the basis of what is known about the regular structures of this environment. The cognitive system needs to be flexible in order to recognize a signal despite several modifications of its physical features (as is the case for spoken word comprehension), to anticipate the incoming of future events, to restore missing information and so on. From this point of view, human brains differ radically from artificial systems by their considerable power to integrate contextual information in perceptual processing. Most of the involved processes are knowledge-driven, which results in a smooth interaction with the environment. A further example that highlights the importance of top-down processes is given by considering what happens when something unexpected suddenly occurs in the environment. In some situations, top-down processes are so strong that the cognitive system fails to accomplish a correct analysis of the situation ("I cannot believe my eyes or my ears"). In some contexts, this failure to interpret unexpected events risks being detrimental and may have dramatic consequences (e.g., in industrial accidents).

No doubt, both bottom-up and top-down processes are indispensable for a complete adaptation to the environment. Sensory-driven processes ensure that the cognitive system is informed about the objective structure of the environmental signals, sometimes in a quite automatic way. Top-down processes, by contrast, contribute to facilitate the processing of signals from very low levels (including signal detection) to more complex ones (such as perceptual expectancies or object identification). It is likely that the contribution of both groups of processes depends on several factors relating to the external situation and to the psychological state of the perceiver. For example, in contrast to a silent perceptual setting with clear signals, a noisy environmental situation would encourage top-down process to intervene in order to compensate for the deterioration of the signals. Projective tests used in clinical psychology (e.g., Rorschach test) may be seen as powerful methods to provoke top-down processes for analyzing ambiguous visual figures with the goal of discovering aspects of the individual's personality. If the visual figures were clearly representing environmental scenes, top-down processes would be less activated.

Although the contribution of top-down processes has been well documented in several domains, including speech perception and visual perception, much remains to be understood about how exactly these processes work in the auditory domain, specifically in non-verbal audition (see McAdams and Bigand 1993). The relatively small part devoted to top-down processes in text books on human audition is rather surprising since no obvious arguments lead us to believe that human audition is more influenced by sensory-driven processes than by top-down processes. The aim of the present and final chapter of this book is to consider some studies that provide convincing evidence about the role played by top-down processes on the processing of pitch structures in music perception. We start by considering some basic examples in the visual domain, which differentiate both types of processes (section 2). We then consider how similar top-down processes influence the perception as well as the memorization of pitch structures (section 3) and govern perceptual expectancies (section 4). Most of these examples were taken from the music domain. As will become evident in what follows, it is likely that Western composers have taken advantage of the fundamental characteristic of the human brain to process pitch structures as a function of the current context and have thus developed a complex musical grammar based on a very small set of musical notes. The next section (5) summarizes some of the neurophysiological bases of top-down processes in the music domain. The last two sections of the chapter analyze the acquisition of knowledge and top-down processes as well as their simulation by artificial neural nets. In section 6, we argue that regular pitch structures from environmental sounds are internalized through passive exposure and that the acquired implicit knowledge then governs auditory expectations. The way this implicit learning in the music domain may be formalized by neural net models is considered in section 7. To close this chapter, we put forward some implications of these studies on context effects for artificial systems of pitch processing and for methods of training hearing-impaired listeners (section 8).[1]

2. Bottom-up versus Top-down Processes

A first example illustrating the importance of top-down processes in vision is shown in Figure 1 and was given by Fisher (1967). Start looking at the left drawing of the first line while masking the second line of the figure. You will identify the face of a man. If now, you look to the other drawings on the right, your perception remains unchanged and the drawing on the extreme right will be perceived as the face of a man. Present now the second line of drawings to another person and require her/him to identify the first drawing on the right, while masking those of the first line. She or he will identify the body of a woman. This perception will not change for the drawings on the left, including the one of the extreme left. The critical point of this demonstration is that the last

---

[1] Music theoretic concepts and basic aspects of pitch processing in music necessary for the understanding of this chapter are introduced in the following sections. Readers interested in more extensive presentations may consult the excellent chapters in Deutsch (1982, 1999) and Dowling and Harwood (1986).

drawing on the right of the first line is identical to the last drawing on the left of the second line. Nevertheless, the same drawing has been perceived completely differently as a function of the context in which it has been presented. After a set of drawings representing a face, it is identified as a man's face. After a set of drawings representing the body of a woman, it is identified as a body. Since the sensory information is strictly identical in both situations, this difference in perception can be explained by the intervention of top-down, context-dependent, processes that determine perception.

Similar examples are numerous in cognitive psychology, and two further examples are presented here. Just consider the sentence displayed in Figure 2 top If you read « my phone number is area code 603, 6461569, please call » without any difficulty, some of the letters have been identified differently depending on the word context in which they appear: with the verb « is » being identified as 15 in the code number, the letter « b » as « h » in phone and as « b » in number, and the letter « l » as « d » in code, and as « l » in please. Similar context effects on letter processing have been reported in reading experiments showing that letter identification and memorization is better when letters form meaningful words (word superiority effect). In a related vein, in Figure 2 bottom you are more likely to interpret the sign in the middle of the two triplets as a B in the sequence on the left and as the number 13 in the sequence on the right. The way a stimulus evolves in space constitutes a further contextual factor that can influence perceptual identification as illustrated by the following example: a hand-drawing of a duck can be perceived as representing the flight of a duck when moving from right to left, but as a flight of plane when moving from left to right. Effects of context are not specific to language or vision, and other examples can be found in tasting (Chollet 2001). For example, changing the color of wine is sufficient to identify the wine as red while being white wine and vice versa, even in expert wine tasters (Morrot et al. 2001).

Some effects of context have been reported in the auditory domain as well. For example, Ballas and Mullins (1991) reported that the identification of an environmental sound (e.g., a burning detonator) that is acoustically similar to another sound (e.g., food cooking) is weaker when it is presented in a context that biases its identification toward the meaning of the other sound (peeling vegetables / cutting food / a burning detonator) than in a context that is consistent with its meaning (lighting matches / burning detonator / explosion). In a well-known experiment, Warren (1970; Warren and Sherman 1974) reported phonemic restoration effects that depend on the semantic context of the spoken sentence. A phoneme was either removed or replaced by white noise bursts in spoken sentences (indicated by *). For example: "It was found that the *eel was on the orange", "It was found that the *eel was on the table" or "It was found that the *eel was on the axle". As a function of the surrounding sentence, listeners reported hearing 'peel', 'meal' or 'wheel' in the three examples. Interestingly, the phenomenon of phonemic restoration only takes place when a noise burst replaces the missing signal. Warren (see Warren 1999 for a review of his work) suggests that a listener hears a sound as being present (participants actually report hearing the phoneme as superimposed on the noise) when there is contextual evidence that the sound may have been present, but has been potentially masked by another sound. Perceptual restoration is not specific to the language domain, and similar effects have been reported in the music domain (Sasaki 1980; DeWitt and Samuels 1990). Sasaki (1980) for example, reported that notes replaced by noise in familiar melodies were 'filled in' by the listener. These outcomes suggest that the cognitive system anticipates specific auditory signals on the basis of the previously heard context (either linguistic or musical). This expectancy is strong enough to restore incomplete or missing information. In some cases, the auditory expectations also influence very peripheral auditory processes. Howard et al. (1984) reported that the detection threshold for auditory signals is influenced by a preceding context, even without an explicit signal indicating the pitch height of the to-be-detected target. In their study, a series of sounds constantly decreased in pitch height with the target being the last event. The contextual movement created the expectation that the target would be placed in the continuity, and participants were more sensitive in detecting a target at that expected pitch height.

The influence of context on the processing of pitch structures was reported as early as 1958 by Francès (1958). In one of his experiments, Francès required musicians to detect mistuned notes in piano pieces. This mistuning was performed in different ways. In one condition, some musical notes were mistuned in such a way that the pitch interval between the mistuned notes and those to which they were anchored was reduced[2]. For example, the leading note (the note B in a C major key) is generally anchored to the tonic note (the note C in a C major key). Francès mistuned the note B by increasing its fundamental frequency (F0) so that the pitch interval between the notes B and C (an ascending semitone or half-step) was reduced. In the other experimental condition, this mistuning was performed in the opposite way (the frequency of the B leading tone was decreased). When played without musical context, participants easily perceived both types of mistuning. Placed in a musical context, only the second type of mistuning (which conflicted with musical anchoring) was perceived. This outcome shows that the perceptual ability to perceive changes in pitch structures (in this study, the shift of the F0 of a musical note) is modulated by top-down processes that integrate the function of the note in the overall musical context. It is likely that the effect of top-down processes reported by Francès in this study was driven by listeners' knowledge of Western tonal music. If this experiment was run with listeners who have never been

---

[2] In Western tonal music, unstable musical tones instill a tension that is resolved by other specific musical tones in very constrained ways (see Bharucha 1984b, 1996). Unstable tones are said to be anchored to more stable ones.

exposed to Western tonal music, these exact context effects may probably not have occurred, or, at least, may have been different (see Castellano et al. 1984).

Since Francès (1958), numerous studies have been performed to further understand the role played by knowledge-driven processes on the perception of pitch structures. Some of these studies demonstrated that the perception and memorization of pitches depend on the musical context in which the pitches appear (Section 3). During the last decade, several studies provided further evidence that the ease with which we process pitch structures mostly depends on knowledge-driven expectations (Section 4). We start by reviewing these studies, and we will then consider in more detail whether context effects are hardwired or develop in the brain (Section 5).

3. Effects of Context on the Perception and Memorization of Pitch Structures in Western Tonal Music

Music is a remarkable medium illustrating how top-down and bottom-up processes may be intimately entwined. It is likely that composers initially developed musical syntactic-like rules that took advantage of the psychoacoustic properties of musical sounds. However, these structures have been influenced by centuries of spiritual, ideological, patriotic, social, geographic and economic practices that are not necessarily related to the physical structure of the sound. The music theorist, Rosen (1971), noted that it can be asked whether Western tonal music is a natural or an artificial language. It is obvious that on the one hand, it is based on the physical properties of sound, and on the other hand, it alters and distorts these properties with the sole purpose of creating a language with rich and complex expressive potential. From a historical perspective, the Western harmonic system can be considered as the result of a long theoretical and empirical exploration of the structural potential of sound (Chailley 1952). The challenge for cognitive psychology is to understand how listeners today grasp a system in which a multitude of psychoacoustic constraints and cultural conventions are intertwined. Is the ear strongly influenced by the acoustic foundations of musical grammar, mentally reconstructing the relationship between the initial material and the final system? Or are the combinatorial principals only internal, without a perceived link to the subject matter heard at the time? In the latter case, the perception of pitch (the only musical dimension of interest in this chapter) seems to depend on top-down rather than bottom-up processes. Consider, for example, musical dissonance: Helmholtz (1885/1954) postulated that dissonance is a sensation resulting from the interference of two sound waves close in frequency, which stimulate the same auditory filter in conflicting ways. Although it is linked to a specific psychoacoustic phenomenon, this sensation of dissonance relies on a relative concept that cannot explain the structure of Western music on its own (cf. Parncutt 1989). The idea of dissonance has evolved during the course of musical history: certain musical intervals (e.g., the third) were not initially considered as consonant. Each musical style could use these sensations of dissonance in many ways. For example, a minor chord with a major 7th is considered to be perfectly natural in jazz, but not in classical music. Similarly, certain harmonic dissonances of Beethoven, whose musical significance we now take for granted, were once considered to be harmonic errors that required correction (cf. Berlioz 1872). Even more illustrative examples of the cultural dimension of dissonance are innumerable when considering contemporary music or the different musical systems of the world. These few preliminary notes show that sensory qualities linked to pitch cannot be understood outside of a cultural reference frame.

It is actually well established in the music cognition domain that a given auditory signal (a musical note) can have different perceptual qualities depending on the context in which it appears. This context dependency of musical note perception was exhaustively studied by Krumhansl and collaborators from 1979 to 1990 (for a summary of this research see Krumhansl 1990). In order to understand the rationale of these studies, let us consider shortly the basic structures of the Western musical system.

Two aspects of the notion of pitch can be distinguished in music: one related to the fundamental frequency F0 of a sound (measured in Hertz), which is called pitch height, and the other related to its place in a musical scale which is called pitch chroma. Pitch height varies directly with frequency over the range of the audible frequencies. This aspect of pitch corresponds to the sensation of high and low. Pitch chroma embodies the perceptual phenomenon of octave equivalence by which two sounds separated by an octave are perceived as somewhat equivalent. Pitch chroma is organized in a circular fashion, with octave-equivalent pitches considered to have the same chroma. Pitches having the same chroma define pitch classes. In Western music, there are 12 pitch classes referred to with the following labels: C, C# or D*b*, D, D# or E*b*, E, F, F# or G*b*, G, G# or A*b*, A, A# or B*b*, and B. All musical styles of Western music (from baroque music to rock' n roll and jazz music) rest on possible combinations of this finite set of 12 pitch classes. Figure 3 illustrates the most critical features of these pitch classes combined in the Western tonal system.

The specific constraints to combine these pitch classes have evolved through centuries and vary as a function of stylistic periods. The basic constraints that are common to most Western musical styles are described in textbooks of Western harmony and counterpoint. A complete description of these constraints is beyond the scope of this chapter, and we will simply focus on those features that are indispensable for understanding the basis of context effects in Western tonal music. For this purpose, it is sufficient to understand that the 12 pitch classes are combined into two categories of musical units: chords and keys. The musical notes (i.e., the twelve chromatic notes) are combined to define musical chords. For example, the notes C, E and G define a C major chord, and the notes F, A and C define an F major chord. The frequency ratios between two notes define musical pitch intervals and are expressed in the music domain by the number of semitones (for a presentation of intervals in terms of frequency ratios see Burns 1999, Table 1). For example, the distance in pitch between the notes C and E is 4 semitones and defines the pitch interval of a major third. The pitch interval between the notes C and E*b* is three semitones, and defines a minor third. The pitch interval between the notes C and G is 7 semitones, and defines a perfect fifth. A diminished fifth is defined by two musical notes separated by 6 semitones (e.g., C and G*b*). Musical chords can be major, minor or diminished depending on the types of interval they are made of. A major chord is made of a major third and a perfect fifth (e.g., C-E, and C-G, respectively). A minor chord is made of a minor third and a perfect fifth (e.g., C-E*b* and C-G). A diminished chord is made of a minor third (C-E*b*) and diminished fifth (e.g., C-G*b*). A critical feature of Western tonal music is that a musical note (say C) may be part of different chords (e.g., C, F and A*b* major chords, c, a and f minor chords), and its musical function changes depending on the chord in which it appears. For example, the note C acts as the root, or tonic, of C major and c minor chords, but as the dominant note in F major and f minor chords.

The 12 musical notes are combined to define 24 major and minor chords that, in turn, are organized into larger musical categories called musical keys. A musical key is defined by a set of pitches (notes) within the span of an octave that are arranged with certain pitch intervals among them. For example, all major keys are organized with the following scale: two semitones (C-D in the case of the C major key), two semitones (D-E), one semitone (E-F), two semitones (F-G), two semitones (G-A), two semitones (A-B) and one semitone (B-C'). The scale pattern repeats in each octave. By contrast, the minor keys (in its minor harmonic form) are organized with the following scale: two semitones (C-D, in the case of the C minor key), one semitone (D-E*b*), two semitones (E*b*-F), two semitones (F-G), one semitone (G-A*b*), three semitones (A*b*-G), and one semitone (B-C). On the basis of the twelve musical notes and the 24 musical chords, 24 musical keys can be derived (e.g., 12 major and 12 minor keys)[3]. For example, the chords C, F, G, d, e, a and b° belong to the key of C major, and the chords F#, C#, B, g#, a#, d# and e#° define the key of F# major. Further structural organizations exist inside each key (referred to as tonal-harmonic hierarchy in Krumhansl, 1990) and between keys (referred to as inter-key distances). The concept of tonal hierarchy designates the fact that some musical notes have more referential functions inside a given key than others. The referential notes act in the music domain like cognitive reference points act in other human activities (Rosch 1975, 1979). Human beings generally perceive events *in relation to* other more referential ones. As shown by Rosch and others, we perceive the number 99 as being almost 100 (but not the reverse), and we prefer to say that basketball players fight like lions (but not the reverse). In both examples, "100" and "lion" act as cognitive reference points for mental representations of numbers and fighters (see also Collins and Quillian 1969). Similar phenomena occur in music. In Western *tonal* music, the tonic of the key is the most referential event in relation to which all other events are perceived (Schenker 1935; Lerdahl and Jackendoff 1983, for a

---

[3] The first attempt to musically explore all of these keys was done by JS Bach in the Well-tempered clavier. Major, minor and diminished chords are defined by different combinations of three notes. Minor chords and minor keys are indicated by lower case letters, and major chords and major keys by upper case letters. The symbol ° refers to diminished chords.

formal account)[4]. Supplementary reference points exist, as instantiated by the dominant and mediant notes[5]. These differences in functional importance define a within-key hierarchy for notes. A similar hierarchy can be found for chords: chords built on the first degree of the key (the tonic chord) act as the most referential chord of Western harmony, followed by the chords built on the fifth and fourth scale degrees (called dominant and subdominant, respectively).

Intra-key hierarchies are crucial in accounting for context effects in music. Indeed, a note (and also a chord) has different musical functions depending on the key context in which it appears. For example, the note C acts as a cognitive reference note in the C major and c minor keys, as the less referential dominant note in the F major and minor keys, as a moderate referential mediant note in the A*b* major key and the a minor key, as weakly referential notes in the major keys of B*b*, G and E*b* as well as in the minor keys of b*b*, g and e, as an unstable leading note in the major and minor keys of D*b* and as non-referential, non-diatonic note in all remaining keys. As the 12 pitch classes have different musical functions depending on the 12 major and 12 minor key contexts in which they can occur, there are numerous possibilities to vary the musical qualities of notes in Western tonal music. The most critical feature of the Western musical system is thus to compensate for the small number of pitch classes (12) by taking advantage of the influence of context on the perception of these notes. In other words, there are 12 physical event classes in Western music, but since these events have different musical

---

[4] The tonal system refers to a set of rules that characterize Western music since the baroque (seventeenth century), classical, and romantic styles. This system is still quite prominent in the large majority of traditional and popular music (rock, jazz) of the Western world as in Latin America.

[5] Western music is based on an alphabet of twelve tones, known as the chromatic scale. This system then constitutes subsets of seven notes from this alphabet, each subset being called a scale or key. The key of C major (with the tones C, D, E, F, G, A, B) is an example of one such subset. The first, third and fifth notes of the major scale (referred to as tonic, median and dominant notes) act as cognitive references notes. Musical chords correspond to the simultaneous sounding of 3 different notes. A chord is built on a root, which gives its name to the chord. So that the C major chord correspond to a major chord built on the tone C. In a given key, the chords built on the first, fourth and fifth notes of the scale (i.e., C, F and G, in a C major scale, for example) are referred to as Tonic, Subdominant and Dominant chords. These chords act as cognitive references events in Western music (see Krumhansl, 1990, Bigand 1993, for a review).

functions depending on the context in which they occur, the Western tonal system has a great number of possible musical events.

A further way to understand the importance of this feature for music listening is to consider what would happen if the human brain were not sensitive to contextual information. All the music we listen to would be made of the same 12 pitch classes. As a result, there would be a huge redundancy in pitch structures inside a given musical piece as well as across all Western musical pieces. As a consequence, we may wonder whether someone would enjoy listening to Beethoven's 9th symphony, Dvorak's Stabat mater or Verdi's Requiem until the end of the piece (with a duration of about 90 minutes) and whether someone would continue to enjoy listening to these musical pieces after having perceived them once or twice[6]. This problem would be even more crucial for absolute pitch listeners who are able to perceive the exact pitch value of a note without any reference pitch. It is likely that composers have used the sensitivity of the human brain for context effects in order to reduce this redundancy. Indeed, Western musical pieces rarely remain in the same musical key. Most of the time, several changes in key occur during the piece, the number of changes being related to the duration of the piece. These key changes modify the musical functions of the notes and result in noticeable changes of the perceptual qualities of the musical flow. For a very long time, Western composers have used the psychological impact of these changes in perceptual qualities for expressive purposes (see Rameau 1721, for an elegant description). Expressive effects of key changes or modulations are stronger when the second key is musically distant from the previous one. For example, the changes in perceptual qualities of the musical flow resulting from the modulation from the key of C major to the key of G major will be

---

[6] To some extent, twelve-tone music of Schoenberg, Webern and Berg faces this difficult problem when using rows of 12 pitch classes for composing long musical pieces without the possibility to manipulate their musical function. Not surprisingly, the first dodecaphonic pieces were of very short duration (see Webern pieces for orchestra).

moderate and less salient than those resulting from a modulation from the C major key to the F# major key.

The musical distances between keys are defined in part by the number of notes (and chords) shared by the keys. For example, there are more notes shared by the keys of C and G major than by the keys of C and F# major. A simplified way to represent the inter-key distances is to display keys on a circle (Fig. 2, bottom), which is called the circle of fifths. Major keys are placed on this circle as a function of the number of shared notes (and chords), with more notes and chords in common between adjacent keys on the circle. Inter-key distances with minor keys are more complex to represent because the 12 minor keys share different numbers of notes and chords with major keys. Moreover, the number of shared notes and chords defines only a very rough way to describe musical distances between keys. A more convincing way to compute these distances considers the strength of the changes in musical functions that occurs for each note and chord when the music modulates from one key to another (see Lerdahl 1988; Krumhansl 1990; Lerdahl 2001). A complete account of this computation is beyond the scope of this chapter, but one example is sufficient to explain the underlying rationale. The number of notes shared by the C major key and the c minor key is 5 (i.e., the notes C, D, F, G and B). The number of notes shared by the C major key and the B*b* major key is also 5 (i.e., C, D, F, G and A). Nevertheless, the musical distance between the former keys is less strong than between the latter keys. This is because the change in musical functions are less numerous in the former case than in the latter. Indeed, the cognitive reference points (tonic and dominant notes) are the same (C and G) in the C major and c minor key contexts. By contrast, these two notes are not referential in the key context of B*b* major (in which the notes B*b* and F act as the most referential notes). As a consequence, a modulation from the C major key to the B*b* major key has more musical impact than a modulation toward the c minor key. More generally, by choosing to modulate from one key to another, composers modify the musical functions of notes, which results in expressive effects for Western listeners: the more distant the musical keys are, the stronger the effect of the modulation. Composers of the Romantic period (e.g., Chopin) used to modulate more often toward distant keys than did composers of the Baroque (e.g., Vivaldi, Bach) and Classical periods (e.g., Haydn, Mozart). If human brains were not integrating contextual information for the processing of pitch structures, all these refinements in musical styles would probably have never been developed.

To summarize, the most fundamental aspect of Western music cognition is to understand the context dependency of musical notes and chords and of their musical functions. Krumhansl's research provides a deep account of this context dependency of musical notes for both perception and memorization. In her seminal experiment, she presented a short tonal context (e.g. seven notes of a key or a chord) followed by a probe note (defining the "probe-note" method). The probe note was one note of the 12 pitch classes. Participants were required to evaluate on a 7-point scale how well each probe note fit with the previous context. As illustrated in Figure 4, the goodness-of-fit judgments reported for the 12 pitch classes varied considerably from one key context to another. Musical notes receiving higher ratings are said to be perceptually stable in the current tonal context. Krumhansl and Kessler's (1982) tonal key-profiles demonstrated that the same note results in different perceptual qualities, referred to as musical stabilities, depending on the key of the tonal context in which it appears. These changes in musical stability of notes as a function of key contexts can be considered as the cognitive foundation of the expressive values of modulation.

Krumhansl also demonstrated that within-key hierarchies influence the perception of the relationships between musical notes. In her experiments, pairs of notes were presented after a short musical context and participants rated on a scale from 1 to 7 the degree of similarity of the second note to the first note, given the preceding tonal context. All possible note pairs were constructed with the 12 pitch classes. The note pairs were presented after short tonal contexts that covered all 24 major and minor keys. The similarity judgments can be interpreted as an evaluation of the psychological distance between musical notes with more similarly judged notes corresponding to psychologically closer notes. The critical point of Krumhansl's finding was that the psychological distances between notes depended on the musical context as well as on the temporal order of the notes in the pair. For example, the notes G and C were perceived as being closer to each other when they were presented after a context in the C major key than after a context in the A major key or the F# major key. In the C major key context, the G and C notes both act as strong reference points (as dominant and tonic notes, respectively) which is not the case in the A and F# major keys to which these notes do not belong.

This finding suggests that musical notes are perceived as more closely related when they play a structurally significant role in the key context (i.e. when they are tonally more stable). In other words, tonal hierarchy affects psychological distances between musical pitches by a principle of *contextual distance*: the psychological distance between two notes decreases as the stability of the two notes increases in the musical context. The temporal order of presentation of the notes in the pair also affected the psychological distances between notes. In a C major context for example, the psychological distance between the notes C and D was greater when the C note occurred first in the pair than the reverse. This *contextual asymmetry principle* highlights the importance of musical context for perceptual qualities of musical notes and shows the influence of a cognitive representation on the perception of pitch structures.

A further convincing illustration of the influence of the temporal context on the perception of pitch structures was reported by Bharucha (1984a). In one experimental condition, he presented a string of musical notes, such as B3-C4-D#4-E4-F#4-G4, to the participants. In the other experimental condition, the temporal order of these notes was reversed leading to the sequence G4-F#4-E4-D#4-C4-B3. In the musical domain, this sequence is as ambiguous as the well-known Rubin figure in the visual domain, which can be perceived either as a goblet or two faces. Indeed, the sequence is based on the three notes of the C major chord (C-E G) that are interleaved with the three notes of the B major chord (B-D#-F#). Interestingly, these chords do not share a parent key, and are thus somewhat incompatible. Bharucha demonstrated that the perception of this pitch sequence depends on the temporal order of the pitches. Played in the former order, the sequence is perceived as being in C major; played in the latter order, it is perceived in B major. In other words, the musical interpretation of an identical set of notes changes with the temporal order of presentation. This effect of context might be compared with the context effect described above concerning the influence of stimulus movement on visual identification (duck versus planes).

The context effects summarized above have also been reported for the memorization of pitch structures. For example, Krumhansl required participants to compare a standard note played before a musical sequence to a comparison note played after this musical sequence. The performance in this memorization task depended on the musical function of both standard and comparison notes in the interfering musical context. When standard and comparison notes were identical (i.e., requiring a *same response*), performance was best when the notes acted as the tonic note in the interfering musical context (e.g. C in the C major key), it diminished when the notes acted as mediants (e.g., E in the C major key) and was worst when they did not belong to the key context. This finding underlines the role of the *contextual identity principle*: The perception of identity between two instances of the same musical note increases with the musical stability of the note in the tonal context. When standard and comparison notes were different (i.e., requiring a *different response*), the memory errors (confusions) also depended on the musical function of these notes in the interfering musical context, as well as on the temporal order. For example, when the comparison note acted as a strong reference note in the context (e.g., a tonic note) and the standard as a less referential note, memory errors were more numerous than when the comparison note acted as a less referential note and the standard as a strong reference note in the context. This finding cannot be explained by sensory-driven processes. It suggests that in the auditory domain, as in other domains (see for example Rosch for the visual domain), some pitches act as cognitive reference points in relation to which other pitches are perceived. It thus provides a further illustration of the principle of contextual asymmetry described above. Consistent support for contextual asymmetry effects on memory was reported by Bharucha (1984) with a different experimental setting.

Several attempts have been made to challenge Krumhansl and colleagues' demonstration of the cognitive foundation of musical pitch. For example, Huron and Parncutt (1993) argued that most of Krumhansl's probe-note data may be accounted for by a sensory model and can emerge from an echoic memory model based on pitch salience and including a temporal decay parameter. More recently, Leman (2000) provided a further challenge to these data arguing that none of the previously reported context effects occur at a cognitive level but may simply be explained by some sort of sensory priming. Notably, Leman (2000) simulated data with the help of a short-term memory model based on echoic images of periodicity pitch only.

Given that both top-down and bottom-up processes are intimately entwined in Western music, a critical issue remains to assess the strength of each type of process for music perception. Dowling's remarkable work has demonstrated how both processes may contribute to melodic perception and memorization (Dowling, 1972, 1978, 1986, 1991; Bartlett and Dowling, 1980, 1988; Dowling and Bartlett, 1981; Dowling et al. 1995). The influence of bottom-up processes is reflected by listeners' sensitivity to the melodic contour (that is the up-and-down of pitch intervals in the melody). Top-down influences are reflected by the importance of the position of the notes in the musical scale (e.g., tonic or dominant). One critical feature of Dowling's experiments was to demonstrate that a change in melodic contour was more difficult to perceive when the comparison melody was played in a far rather than a close key. A further fascinating finding of Dowling was to show that a given melody played in two different harmonic contexts was not easily perceived as having exactly the same melodic contour. The change in scalar position of the melodic notes from one musical key context to the other interfered with the ability to perceive the melodic contour.

One of our experiments on melody perception directly addressed the strength of top-down processes in a very similar way (Bigand 1997). The study involved presenting 29-note sequences (Figure 5) to participants. The challenge was to modify the perception of these note sequences by changing only a few pitches (i.e., five pitches between melody T1 and melody T2). On music theoretical grounds, these few pitch changes should be sufficient to make participants perceive the melody T1 in the context of an a minor key and the melody T2 in the context of a G major key. Given that the musical stability of individual notes changes as a function of key, the profile of perceived musical stability was supposed to vary strongly from T1 to T2, even though both melodies shared a large set of pitches, the same contour and the same rhythm. For example, stop note 2 is a strong referential tonic note in T1, but a weak referential subtonic note in T2. Similarly, stop note 4 is a rather referential mediant note in T1 and a less referential subdominant note in T2. By contrast, stop note 3 is a weak referential supertonic in T1, but a rather strong referential mediant in T2. Readers familiar with music can observe that notes that are referential in one melodic context are less referential in the other, and this is valid up to the last note. Indeed, stop note 23 is a referential tonic in T1, but a less referential supertonic in T2. As a consequence, melody T1 sounds complete, but melody T2 does not. The experimental method to measure perceived musical stability consisted in breaking the melody into 23 fragments, each starting from the beginning of the melody and ending on a different note of the melody (i.e., incremental method). As in Krumhansl and Palmer's (1987a,b) studies, participants were required to evaluate the degree of completeness of each fragment. Fragments ending on a stable musical note were supposed to result in stronger feelings of musical completion than those ending on a musically instable note. As a consequence, we predicted musical stability profiles to vary strongly from T1 to T2.

The observed stability profiles of the two melodies were negatively correlated in both musicians' and nonmusicians' data (see Fig. 5 bottom for musicians' data). This outcome shows that listeners (musician and nonmusicians) perceived the pitch structure of the two melodies differently, even though they largely contained the same set of pitches and pitch intervals, and had identical melodic contours and rhythms. Moreover, when these melodies were used in a memorization task, participants estimated on average that about 50% of the pitches of the T2 melodies had been changed to create the T1 melodies (Bigand and Pineau 1996). Surprisingly, musicians did not outperform nonmusicians in this task suggesting that for both groups of listeners the musical functions of melodic notes contributed more strongly to defining the perceptual identity of a melody than the actual pitches, pitch intervals, melodic contour and rhythm. Both studies underline the strength of cognitive top-down processes on the perception and memorization of melodic notes.

As explained above, musical notes define the smallest building block of Western tonal music. Musical chords define a larger unit of Western musical pitch structures. A musical chord is defined by the simultaneous sounding of at least three notes, one of these notes defining the root of the chord. Other notes may be added to this triadic chord, which results in a large variety of musical chords. The influence of musical context on the perception of the musical qualities of these chords, as well as the perceptual relations between these chords has been largely investigated by Krumhansl and collaborators (see Krumhansl 1990 for a summary). The rationale of these studies follows the rationale of the studies briefly summarized above for musical notes (see Krumhansl 1990).

For example, in Bharucha and Krumhansl (1983), two chords were played after a musical context, and participants rated on a 7-point scale the similarity of the second chord to the first one given the preceding context. The pairs of chords were made of all combinations of chords belonging to two musical keys that share only a few pitches (C and F# major). In other words, these keys are musically very distant. If the perception of harmonic relations was not context-dependent, the responses of participants would not have been affected by the context in which these pairs were presented. Figure 6 demonstrates that the previous musical context had a huge effect on the perceived relationships of the two chords. When the context was in the key of C major, the chords of the C major key were perceived as more closely related than those of the F# major key. When the F# major key defined the context, the inverse phenomenon was reported. The most critical finding was that when the musical key of the context progressively moved from the C major key to the F# major key through the keys of G, A and B (see the positions of these keys on the Cycle of Fifths, Figure 3), the perceptual proximity between the chord pairs progressively changed, so that C major chords progressively were perceived as less related, and F# major chords more related (cf. Krumhansl et al. 1982). Similar context effects have also been reported in memory experiments, suggesting that it is unlikely that these context effects are caused by sensory-driven processes solely (Krumhansl 1979; Bharucha and Krumhansl 1983).

It is difficult to rule out entirely the influence of sensory-driven processes on the perception of Western harmony in these experiments. This restriction applies even though the authors carefully used Shepard tones (Shepard 1964)[7] and provided converging evidence from perceptual and memory tasks, which suggests that the reported context effects occurred at a cognitive level. The purpose of one of our studies was to contrast sensory and cognitive accounts of the perception of Western harmony (Bigand et al. 1996). Participants listened to triplets of chords with the first and third chords being identical (e.g. X-C-X). Only the second chord was manipulated and participants evaluated on a 10-point scale the musical tension instilled by the second chord. The manipulated chord was either a triad (i.e., the 12 major and 12 minor triads) or a triad with a minor seventh (i.e., 12 major chords with minor seventh, and 12 minor chords with a minor seventh). The musical tensions were predicted by Lerdahl's cognitive Tonal Pitch Space theory (Lerdahl 1988) and by several psychoacoustical models, including Parncutt's theory (Parncutt 1988). One of the main outcomes was that all models contributed to predicting the perceived musical tension, with albeit a stronger contribution of the cognitive model. This outcome suggests that the abstract knowledge

---

[7] Shepard tones consist, for example, of five sine wave components spaced at octave frequencies in a five-octave range with an amplitude envelope being imposed over this frequency range so that the components at low and high ends approach hearing threshold. These tones have an organ-like timbral quality and minimize the perceived effect of pitch height.

of Western pitch regularities constitutes some kind of cognitive filter that influences how we perceive musical notes and chords. A further influence of this knowledge is documented in the next section by showing that internalized pitch regularities also result in the formation of perceptual expectancies that can facilitate (or not) the processing of pitch structures.

4 Influence of Knowledge-Driven Expectancy on the Processing of Pitch Structures

Once we are familiarized with a given environment, we process environmental stimuli in a highly constrained way. For example, we are not able to ignore linguistic information displayed in our native language, and we automatically anticipate from a previous context the type of events that are likely to occur next. Irrepressible processing and perceptual anticipation have been documented in a variety of domains, including language, face processing and vision. During the last decade, numerous studies have been devoted to investigating the influence of auditory expectations on the processing of pitch structures in the music domain. The seminal studies on harmonic expectancies involved very short contexts. For example, in Bharucha and Stoeckig (1987), participants were required to perform a simple perceptual task on a target chord that was preceded by a prime chord. The harmonic relationship between the prime chord and the target chord defined the variable of interest, and the critical point was to assess whether this relation influenced the processing of the target. For the purpose of the experimental task, the target chord was either in tune or out of tune, and participants had to decide quickly and accurately whether the target was in tune or out of tune. The principal outcome was that the processing of in-tune targets (e.g., a C major chord) was easier and faster when the target was preceded by a musically related prime chord (e.g., a G major chord) than by a musically unrelated prime chord (e.g., an F# major chord). In the research of Bharucha and collaborators, the effect of context was reversed for out-of-tune targets (with better identification of out-of-tune targets when preceded by a musically unrelated prime). These findings provided evidence for the anticipatory processes that occur from chord to chord when listening to music.

Further experiments were performed to confirm that priming effects mostly occur at a cognitive level and cannot result only from sensory priming. Bharucha and Stoeckig (1987) reported priming effects even when prime and target chords did not share any component notes. Tekman and Bharucha (1992) reported priming effects even when prime and target were separated by long silent intervals, and when white noise was introduced between prime and target. Moreover, in a recent study, we observed that harmonic relatedness resulted in a stronger priming effect than chord repetition (Bigand et al, in preparation). In the harmonic priming condition, the target chord (say a C major chord) was preceded by a musically highly related prime chord (a G major chord in this case). In the repetition priming condition, prime and target chords were identical (a C major chord followed by a C major chord). Repetition priming involves a strong component of sensory priming since the two chords are identical. Harmonic priming involves strong top-down influences since the harmonic relation between prime and target corresponds to the most significant musical relationship in Western tonal music (i.e., an authentic cadence, which is a harmonic marker of phrase endings). In a set of five experiments, we never observed stronger priming effects in the repetition condition. Moreover, significantly stronger priming was observed in the harmonic priming condition in most of the experiments. This finding raises considerable difficulties for sensory models of music perception as the processing of a musical event is more facilitated when it is preceded by a different, but musically related chord than when it is preceded by an identical (repeated) chord.

These studies suggest that a single prime chord manages to activate an abstract knowledge of Western harmonic hierarchies. This activation results in the expectation that harmonically related chords should occur next. The present interpretation does not imply that sensory priming never affects chord processing. Indeed, Tekman and Bharucha (1998) showed that cognitive priming failed to overrule sensory priming when Stimulus-Onset-Asynchrony (SOA) between chords was as short as 50ms. In this experiment, the authors contrasted two types of prime and target relations. In one type of chord pair, the target shared one note with the prime (C and E major chords)[8], but shared no parent major key. The other type of pair represented the opposite situation with the target sharing no note with the prime (C and D major chords), but both sharing a parent key (i.e., the key of G Major). Consequently, the first pair favors sensory priming, while the second pair favors cognitive priming. The authors demonstrated that the processing of the target chord was facilitated in the second pair only for SOAs longer than 50ms. This outcome suggests that top-down influences need some time to be instilled, while sensory priming occurs very quickly.

The influence of longer musical contexts on the processing of target chords has been addressed in several ways. In Bigand and Pineau (1997), eight-chord sequences were used with the last chord defining the target. The harmonic function of the target chord was varied by manipulating the first six chords of the sequence (Fig. 7). In the strongly expected condition, the target chord acted as a tonic chord (I). In the less expected condition, the target acted as a subdominant chord (IV), which was musically congruent with the context, but less expected. In order to reduce sensory priming effects,

---

[8] The major chords C, D and E consist of the tones (C-E-G), (D,F# A) and (E-G#-B), respectively.

the chord immediately preceding the target was identical in both conditions. For the purpose of the experimental task, the target chord was rendered acoustically dissonant in half of the trials by adding a note to the chord. As a consequence, 25% of the trials ended on a consonant tonic chord, 25% on a consonant subdominant chord, 25% on a dissonant tonic chord, and 25% on a dissonant subdominant chord. Participants were required to indicate as accurately and as quickly as possible whether the target chord was acoustically consonant or dissonant. The critical finding of the study was to show that this consonant/dissonant judgment was more accurate and faster when targets acted as a tonic rather than as a subdominant chord. This suggests that the processing of harmonic spectra is facilitated for events that are the most predictable in the current context. Moreover, this study provided further evidence that musical expectancy does not occur from chord to chord, but also involves higher levels of musical relations.

This last issue was further investigated in Bigand et al. (1999) by using 14-chord sequences. As illustrated in Figure 7 (b), these chord sequences were organized into two groups of seven chords. The first two conditions replicated the conditions of Bigand and Pineau (1997) with longer sequences: chord sequences ended on either a highly expected tonic target chord or a weakly expected subdominant target chord. The third condition was new for this study and created a moderately expected condition. This third group of sequences was made out of the sequences in the first two conditions: The first part of the highly expected sequences (chords 1 to 7) defined the first part of this new sequence type and the second part of the weakly expected sequences (chords 8 to 14) defined their second part. The critical comparison was to assess whether the processing of the target chord is easier and faster in the moderately expected condition than in the weakly expected condition. This facilitation would indicate that the processing of a target chord has been primed in this third sequence by the very beginning of the sequence (the first seven chords which are highly related). The behavioral data confirmed this prediction. For both musician and nonmusician listeners, the processing of the target was most facilitated in the highly expected condition, followed by the moderately expected condition and then by the weakly expected condition. This finding further suggests that context effects can occur over longer time spans and at several hierarchical levels of the musical structure (see also Tillmann et al. 1998).

The effect of large musical contexts on chord processing has been replicated with different tasks. For example, in Bigand et al. (2001), chord sequences were played with a synthesized singing voice. The succession of the synthetic phonemes did not form a meaningful, linguistic phrase (e.g., /da fei ku ʃo fa to kei/). The last phoneme was either the phoneme /di/ or /du/. The harmonic relation of the target chord was manipulated so that the target acted either as a tonic or as a subdominant chord. The experimental session thus consisted of 50% of the sequences ending on a tonic chord (25% being sung with the phoneme *di*, 25% with the phoneme *du*) and 50% of sequences ending with a subdominant chord (25% sung with the phoneme *di*, 25% with the phoneme *du*). Participants performed a phoneme-monitoring task by identifying as quickly as possible whether the last chord was sung with the phoneme *di* or *du*. Phoneme-monitoring was shown to be more accurate and faster when the phoneme was sung on the tonic chord than on the subdominant chord. This finding suggests that the musical context is processed in an automatic way - even when the experimental task does not require paying attention to the music. As a result, the musical context induces auditory expectations that influence the processing of phonemes. Interestingly, these musical context effects on phoneme monitoring were observed for both musically trained and untrained adults (with no significant difference between these groups), and have recently been replicated with 6-year-old children. The influence of musical contexts was replicated when participants were required to quickly process the musical timbre of the target (Tillmann in preparation) or the onset asynchrony of notes in the target (Tillmann and Bharucha 2002).

These experiments differ from those run by Bharucha and collaborators not only by the length of the musical prime context, but also because complex musical sounds were used as stimuli (e.g., piano-like sounds in Bigand et al. 1999; singing voice-like sounds in Bigand et al. 2001) instead of Shepard notes. Given that musical sounds have more complex harmonic spectra than do Shepard notes, sensory priming effects should have been more active in the studies by Bigand and collaborators. A recent experiment was designed to contrast the strength of sensory and cognitive priming in long musical contexts (Bigand et al. 2003). Eight-chord sequences were presented to participants who were required to make a fast and accurate consonant/dissonant judgment on the last chord (the target). For the purpose of the experiment, the target chord was rendered acoustically dissonant in half of the trials by adding an out-of-key note. As in Bigand and Pineau (1997), the harmonic function of the target in the prime context was varied so that the target was always musically congruent: in one condition (highly expected condition), the target acted as the most referential chord of the key (the tonic chord) while in the other (weakly expected condition) it acted as a less referential subdominant chord. The critical new point was to simultaneously manipulate the frequency of occurrence of the target in the prime context. In the *no-target-in-context condition*, the target chords (tonic, subdominant) never occurred in the prime context. In this case, the contribution of sensory priming was likely to be neutralized. As a consequence, a facilitation of the target in the highly expected condition over the weakly expected condition could be attributed to the influence of knowledge-driven processes. In the *subdominant-target-condition*, we attempted to boost the strength of sensory priming by increasing the frequency of occurrence of the subdominant chord only in the prime context (the tonic chord never occurred in the context). In this condition, sensory priming was thus expected to be stronger, which should result in facilitated processing for subdominant targets.

In Experiment 1, the consonant/dissonant task was performed more easily and quickly for tonic targets, and there was no effect of the frequency of occurrence. This finding suggests that top-down processes (cognitive priming) are more influential than sensory-driven process (sensory priming) in large musical contexts even though complex piano-like sounds were used. In Experiment 2, the same sequences were used, but the tempo at which the sequences were played was increased. The slowest tempo was two times faster than in Experiment 1 (i.e., 300ms per chord) and the highest tempo was 8 times faster (i.e., 75ms per chord). The tempo variable was manipulated in blocks, with half of the participants starting the experiment with the slowest tempo and ending with the fastest tempo (group Slow-Fast). The other half of the participants started with the fastest tempo and ended with the slowest tempo (group Fast-Slow). On the basis of Tekman and Bharucha (1998), we expected that sensory priming would become more influential than cognitive priming with increasing tempo.

Our findings globally confirmed this hypothesis, with an interesting data pattern. At tempi of 300ms and 150ms per chord, priming effects were always stronger for tonic chords, irrespective of the target's frequency of occurrence. This data pattern changed at the fastest tempo (75ms per chord), and there was a significant interaction with the temporal order at which the tempi were presented in the experimental session (i.e., groups Fast-Slow versus Slow-Fast). At this extremely fast tempo, sensory priming overruled cognitive priming only in the Fast-Slow group, and cognitive priming continued to be more influential in the Slow-Fast group. This second experiment sheds new light on the working of top-down processes in music by demonstrating that these processes continue to be more influential than sensory-driven processes even at a tempo as fast as 150ms per chord.

This outcome highlights the speed at which the cognitive system manages to process abstract information (e.g., the musical function of a chord). At the tempo of 75ms, sensory-driven processes overrule cognitive processes only in listeners who started to process musical sequences presented at this extremely fast tempo. The fact that cognitive priming continued to be more influential than sensory priming in the Slow-Fast group suggests that, once activated, the cognitive component continues to overrule sensory priming even at this extremely fast tempo. Once again, this complex pattern of data was observed for both musically trained and untrained listeners. This finding demonstrates that the auditory perception of musically untrained listeners is more sophisticated than generally assumed, at least for tasks involving the processing of complex pitch structures (e.g. musical chords). The weak difference observed in most of the studies cited above suggests that context effects in music involve robust, cognitive mechanisms.

5 Neurophysiological Bases of Context Effects in the Music Domain

Neurophysiological studies investigate the functioning of top-down processes by analyzing event-related potentials (ERPs) following contextually unexpected events, and by describing the cortical areas involved in these processes with the help of imaging techniques such as functional Magnetic Resonance Imaging (fMRI). Different techniques allow the analysis of different aspects of the neurophysiological bases due to their inherent methodological advantages and limitations, which are notably linked to their temporal and spatial resolution. While electrophysiological methods, which are based on direct mapping of transient brain electric dipoles generated by neuronal depolarization (electroencephalography, EEG) and the associated magnetic dipoles (magnetoencephalography, MEG), provide fine temporal resolution of the recorded signal without precise spatial resolution, fMRI and Positron Emission Tomography (PET) provide increased anatomical resolution of the implied brain structures, but the length of the measured temporal sample is rather long. Griffiths (Chapter 5) describes how these methods allow further understanding of processes linked to different pitch attributes and low-level perceptual processes. The present section focuses on the contribution of these techniques to our understanding of higher-level cognitive processes involved in auditory perception.

Numerous neurophysiological studies investigating top-down processes have used linguistic stimuli and visual stimuli (for a recent review of functional neuroimaging in cognition see Cabeza and Kingstone, 2001). For context effects in language perception, evoked potentials following semantic and syntactic violations have been distinguished. At the end of a sentence (e.g., "The pizza was too hot to …"), the processing of a semantically unexpected word (e.g., "cry") in comparison to an expected word (e.g., "eat") evokes an N400 component (i.e., a negative evoked potential with a maximum amplitude 400ms after the onset of the target word; Kutas and Hillyard 1980). By contrast, a syntactically incorrect sentence construction evokes a late positive potential (with a maximum amplitude 600ms after the onset of the target word defining a P600 component) that has a larger amplitude than the potential evoked by a complex, but correct sentence structure (Patel et al. 1998). Moreover, in simple syntactic sentences, no P600 was observed. This outcome suggests that the amplitude of the P600 is inversely related to the ease of integrating a word into the previous context, with complex syntax and syntactic violation having a cost in terms of structural integration processes.

Over the last few years, a growing number of studies have used musical stimuli (e.g., Besson and Faïta 1995; Janata 1995; Koelsch et al. 2000; Regnault et al. 2001). Interestingly, the influence of a musical context has been shown to be associated with similar electrophysiological reactions as those observed in language perception: a given musical event evokes a stronger P300 (i.e.. a positive evoked potential with a maximum amplitude 300ms after the onset of the target) or a late positive component (LPC peaking around 500 and 600ms) when it is unrelated to the context than when it is related. Besson and Faïta (1995) used familiar and unfamiliar melodies ending on either a congruous diatonic note[9], an incongruous diatonic note or a nondiatonic note. At the onset of the last note of the melodies, the amplitude of the LPC component was stronger for the nondiatonic note than for the incongruous diatonic ones and the weakest for the congruous diatonic notes. Other studies have analyzed the event-related potentials consecutive to a violation of harmonic expectancies (i.e., for chords). Consistent with Besson and Faïta (1995), it was shown that the amplitude of the LPC increases with increasing harmonic violation: the positivity was larger for distant-key chords than for closely related or in-key chords (Janata 1995; Patel et al. 1998). In Patel et al. (1998) for example, target chords that varied in the degree of their harmonic relatedness to the context occurred in the middle of musical sequences: the target chord was either the tonic chord of the established context key, belonged to a closely related key or belonged to a distant, unrelated key. The target evoked an LPC with largest amplitude for distant-key targets, and with decreasing amplitude for closely related key targets and tonic targets. Patel et al. (1998) compared directly the evoked potentials due to syntactic

---

[9] Diatonic notes correspond to notes that belong to the key context.

relations and harmonic relations in the same listeners: both types of violations evoked an LPC component suggesting that a late positive evoked potential is not specific to language processing, but reflects more general structural integration processes based on listeners' knowledge.

The neurophysiological correlates of musical context effects are reported also for finer harmonic differences between target chords. Based on the priming material of Bigand and Pineau (1997), Regnault et al. (2001) attempted to separate two levels of expectations – one linked to the context (related versus less-related targets) and one linked to the acoustic features of the target in the harmonic priming situation (consonant versus dissonant targets). Related targets and less-related targets correspond to the tonic and subdominant chords represented in Figure 6. In half of the trials, these targets were rendered acoustically dissonant by adding an out-of-key note in the chord (e.g., a C# to a C major chord). The experimental design allows an assessment of whether violations of cognitive and sensory expectancies are associated with different components in the event-related potentials. For both musician and nonmusician listeners, the violation of cognitive and sensory expectancy was shown to result in an increased positivity at different time scales. The less-related, weakly expected target chords (i.e., subdominant chords) evoked a P3 component (200-300ms latency range) with larger amplitude than that of the P3 component linked to strongly related tonic targets. The dissonant targets elicited an LPC component (300-800ms latency range) with larger amplitude than the LPC of consonant targets. This outcome suggests that violations of top-down expectancies are detected very quickly, and even faster than violations of sensory dissonance. The observed fast-acting, top-down component is consistent with behavioral measures reported in a recent study designed to trace the time course of both top-down and bottom-up processes in long musical contexts (Bigand et al. 2003, and see section 4, above). In addition, the two components (P3, LPC) were independent; notably the difference in P3 amplitude between related and less-related targets was not influenced by the acoustic consonance/dissonance of the target. This outcome suggests that musical expectancies are influenced by two separate processes. Once again, this data pattern was reported for both musically trained and untrained listeners: both groups were sensitive to changes in harmonic function of the target chord due to the established harmonic context.

Nonmusicians' sensitivity to violations of musical expectancies in chord sequences has been further shown with ERPs (Koelsch et al. 2000) and MEG (Maess et al. 2001) for the same harmonic material. In the ERP study, an early right-anterior negativity (named ERAN, maximal around 150ms after target onset) reflected the harmonic expectancy violation in the tonal contexts. The ERAN was observed independently of the experimental task: e.g., the detection of timbral deviances while ignoring harmonies (Experiments 1 and 2) or the explicit detection of chord structures (Experiments 3 and 4). Unexpected events elicited both an ERAN and a late bilateral frontal negativity, N5, (maximal around 500-550ms). This latter ERP component N5 was interpreted in connection with musical integration processes: its amplitude decreased with increasing length of context and increased for unexpected events. A right-hemisphere negativity (N350) in response to out-of-key target chords has been also reported by Patel et al. (1998, right antero-temporal negativity, RATN) who suggested links between the RATN and the right fronto-temporal circuits that have been implicated in working memory for tonal material (Zatorre et al. 1994). It has been further suggested by Patel et al. (1998) and Koelsch et al. (2000) that the right early frontal negativities might be related to the processing of syntactic-like musical structures. They compared this negativity with the left early frontal negativity ELAN observed in auditory language studies for syntactic incongruities (e.g., Friederici 1995; Friederici et al. 2000). This component is thought to arise in the inferior frontal regions around Broca's area.

The implication of the prefrontal cortex has also been reported for the manipulation and evaluation of tonal material, notably for expectancy violation and working memory tasks (Zatorre et al. 1992, 1994; Patel et al. 1998; Koelsch et al. 2000). Further converging evidence for the implication of the inferior frontal cortices in musical context effects has been provided by Maess et al.'s (2001) study using magneto-encephalography measurements on the musical sequences of Koelsch et al. The deviant musical events evoked an increased bilateral mERAN (the magnetic equivalent of the ERAN) with a slight asymmetry to the right for some of the participants. The generators of this MEG signal were localized in Broca's area and its right hemisphere homologue. Koelsch et al. (2002) investigated with fMRI the neural correlates of musical sequences similar to previously used material (Koelsch et al. 2000; Maess et al. 2001): chord sequences contained infrequently presented unexpected musical events. The observed activation patterns confirmed the implication of Broca's area (and anterior superior insular cortices) in the processing of musical violations. The reported network further included Wernicke's area as well as superior temporal sulcus, Heschl's gyrus and both planum polare and planum temporale.

A recent fMRI study investigated neural correlates of target chord processing in a musical priming paradigm (Tillmann et al. 2003). In 8-chord sequences, the last chord defined the target that was either strongly related (a tonic chord) or unrelated (a chord belonging to a different, unrelated key). As in previous musical priming studies, half of the targets were rendered acoustically dissonant for the experimental task. Participants were scanned with fMRI while performing speeded intonation judgments (consonant versus dissonant) on the target chords. Behavioral results acquired in the scanner replicated the facilitation effect of related over unrelated consonant targets. The overall activation pattern associated with target processing showed commonalities with networks previously described for target detection and novelty processing (Linden et al. 1999; Kiehl et al. 2001). This network included activation in frontal areas (inferior, middle and superior frontal gyri, insula, anterior cingulate) and posterior areas (inferior parietal gyri, posterior cingulate) as well as in the thalamic nuclei and the cerebellum. The characteristics of the targets, notably in how far the chord fit or violated the expectations built up by the prime context, influenced the activation levels of some of these network components. Increased activation was observed for targets that violated expectations based on either sensory-acoustic or harmonic relations. For example, the activation in bilateral inferior frontal regions (i.e., inferior frontal gyrus, frontal operculum, insula) was stronger for unrelated than for related (consonant) targets. The strength of activation in these areas also indicated the detection of dissonant targets in comparison to consonant targets.

The manipulation of harmonic relations in this fMRI study was extremely strong: in the related condition, the target played the role of the most important, stable chord (i.e., the tonic) and in the unrelated condition the target did not even belong to the key of the prime context. Consequently, the two targets had either strong or weak association strengths to the other chords of the prime context. When analyzing musical pieces of the Western tonal repertoire, it will become evident that the related target chord is frequently associated with chords of the prime context, while the unrelated target chord is not. The musical priming study reported increased activation in (bilateral) inferior frontal areas for targets weakly associated to the prime events (the unrelated targets). Interestingly, language studies that manipulated associative strengths between words also reported increased inferior frontal activation for weakly associated words (Wagner et al. 2001) or semantically unrelated word pairs (West et al. 2000). The strong manipulation of the harmonic relations has a second consequence: the notes of the related target occurred in the prime context while the notes of the unrelated target did not. In other words, in these musical sequences sensory and cognitive priming worked in the same direction and favored the related target. It is interesting to make the link with other functional imaging data reporting the phenomenon of repetitive priming for the processing of objects and words: decreased inferior frontal activation is observed for repeated items in comparison to novel items (Koustaal et al. 2001). This finding suggests that weaker activation for musically related targets might also involve repetition priming for neural correlates in musical priming. This hypothesis, which needs further investigation, is very challenging as behavioral studies (reported above) provide evidence for strong cognitive priming (Bigand et al. 2003).

The outcome of the musical priming study is convergent with Maess's source localization of the MEG signal after a musical expectancy violation. The present data sets on musical context effects can be integrated with other data showing that Broca's area and its right homologue participate in nonlinguistic processes (Pugh et al. 1996; Griffiths et al. 1999; Linden et al. 1999; Müller et al. 2001; Adams and Janata 2002) besides their roles in semantic (Poldrack et al. 1999; Wagner et al. 2000), syntactic (Caplan et al. 1999; Embick et al. 2000) and phonological functions (Pugh et al. 1996; Fiez et al. 1999; Poldrack et al. 1999). Together with the musical data, current findings point to a role of inferior frontal regions for the integration of information over time (cf. Fuster 2001). The integrative role includes storing previously heard information (e.g., a working memory component) and comparing the stored information with further incoming events. Depending on the context, listener's long-term memory knowledge about possible relations and their frequencies of occurrence (and co-occurrence) allows the development of expectations for typical future events. The comparison of expected versus incoming events allows the detection of a potential deviant and incoherent event. The processing of deviants, or more generally of less frequently encountered events, may then require more neural resources than processing of more familiar or prototypical stimuli.

6. Implicit Learning of Pitch Regularities

One finding reported in most of the studies described above may have surprised the reader. Top-down influences on perception, memorization and processing of pitch structures were consistently shown to depend only weakly on the extent of musical expertise. This finding contradicts the common belief that musical experts should perceive music differently than musically untrained (supposedly naive) listeners. In the reported experimental studies, musically untrained listeners are sensitive to the same contextual factors as musician listeners, and these factors influence perceptual behavior (and neurophysiological correlates) in roughly the same way as for musician listeners. This outcome suggests that top-down processes are acquired through robust processes that do not require explicit training. This conclusion raises an intriguing question: how can the pitch structure regularities of our environment be internalized by the human brain? In this section, we argue that implicit learning processes that have been investigated in several domains in cognitive psychology are likely to occur as well in the auditory domain and particularly in the music domain. The last section (7) then proposes how these processes might be formalized in a neural net model.

Implicit learning describes a form of learning in which subjects become sensitive to the structure of a complex environment through simple, passive exposure to that environment. Reber (1992) considers this type of learning to be a fundamental cognitive process that permits the acquisition of complex information, which is inaccessible to deductive reasoning. Implicit learning has some specific characteristics that distinguish it from explicit learning processes: implicitly acquired knowledge remains longer in memory (Allen and Reber 1980), is less sensitive to interindividual differences (Reber et al. 1991) and is more resistant to cognitive and neurological disorders (Abrams and Reber 1988).

The most famous experimental protocols to study implicit learning consist of presenting participants with sequences of events (e.g., letters, light positions, sounds) generated by an artificially defined grammar. Figure 8 displays a sample grammar similar to the grammar first used by Reber (1967, 1989). The arrows represent legal transitions between the different letters (X-S-J-Q-W), and a loop indicates possible repetitions of a letter (X or S in this case). During the first phase of the experiment, participants were exposed to sequences of letters that conform to the rules of the grammar (e.g., WJSSX; XSWJSX). One group of participants was asked to discover the rules that generate the grammar (Explicit Condition), while the other group was asked to memorize the sequences and was unaware that any rules existed (Implicit Condition). During the second phase of the experiment, the participants were informed that the sequences of the first phase had been produced by a rule system (which was not described to them). The participants were then asked to judge the grammaticality of new letter sequences. Half of these sequences were ungrammatical (e.g., XSQJ, WSQX for example) and half were new grammatical exemplars. In general, participants in the Implicit Condition performed better than those in the Explicit Condition (varying between 60% and 80% of correct responses). Only a few participants of the implicit group were able to describe aspects of the rules used to generate the letter sequences. As stated initially by Reber (1967, 1989), participants acquired an implicit knowledge of the abstract rules of the grammar. The very nature of the knowledge acquired in these experimental situations, as well as the complete implicit nature of this knowledge has been a matter of debate and still is now (see Perruchet and Pacteau 1990; Perruchet et al. 1997; Perruchet and Vinter 2001), but it is largely admitted that passive exposure results in the internalization of regularities underlying the variations of the external environment.

Although auditory stimuli were rarely used in the domain of implicit learning, some empirical findings demonstrate that regular structures of the auditory environment can also be internalized through passive exposure. A strict adaptation of Reber's study to the auditory domain was realized by Bigand, Perruchet and Boyer (1998), with letters being replaced by musical sounds of different timbres (e.g., gong, trumpet, piano, violin, voice). In the first phase of the experiment, participants listened to sequences of timbres that obeyed the rules of an artificial grammar. The Implicit group was asked to memorize the sequences and to indicate whether a particular timbre sequence was heard for the first or the second time. The Explicit group was required to memorize the timbre sequences and was told that these sequences had been produced by a computer program. Participants of this group were encouraged to try to identify these rules and were told that discovering these rules would contribute to better memory performance. After this first exposition phase, both groups were required to differentiate grammatical and ungrammatical sequences of timbres. A control group was added that performed this last phase without having been exposed to the grammatical sequences. Explicit and Implicit groups performed better than the control group in the grammatical task, with the performance of the Implicit group being slightly better than that of the Explicit group. This outcome suggests that prior exposure to a small number of timbre sequences governed by an artificial rule system was sufficient to enable participants to determine the new sequences that broke one or more of these rules. The internalization of the timbre grammars may therefore result from the simple exposure to sequences generated by the system without the necessity to implement any explicit process of analysis.

A very elegant demonstration of the strength of implicit learning in the auditory domain was provided by Saffran and collaborators. In their initial experiments (Saffran et al. 1996; Saffran et al. 1997), meaningless phonemes were presented to adults, children and infants in a continuous sequence (e.g., bupadapatubitutibu...). The phoneme sequence was constructed with several artificial three-syllable words (e.g., bupada, patubi) chained together without pauses or other surface cues. Consequently, the transition probabilities between two syllables[10] allowed finding word boundaries: transition probabilities inside a word were high, but transition probabilities across word boundaries were weak. If listeners became sensitive to these statistical regularities, they would be able to extract the words from this artificial language. The experiments consisted of two phases. In a first exposition phase, participants listened to the continuous stream for about 20 minutes (Saffran et al. 1996 for adults) while performing either a coloration task or doing nothing. In the second phase of the experiment, participants were tested with a two-alternative forced-choice task: a real word of the artificial language and a non-word (three syllables that do not create a word) were presented in pairs, and participants had to indicate which one belonged to the previously heard sequence. Participants performed above chance in this task, even when words were contrasted to so called *part-words* in which two syllables were part of a real word, but the association with the third syllable was illegal[11]. In infant experiments, the testing phase was based on novelty preferences (and the dishabituation effect): infants' looking

---

[10] The transition probability that A is followed by B is defined by the frequency of the pair AB divided by the frequency of A (Saffran et al. 1996).

[11] For example, for the word "bupada" a part-word would contain the first two syllables followed by a third different syllable "bupaka" (with the constraint that this association does not form another word).

times were longer for the loudspeaker emitting nonwords than for the loudspeaker emitting words. The simple exposure to the sequence of phonemes results in the internalization of artificial words even for 8-month-old infants. With the goal to show that the capacity to extract these statistical regularities is not restricted to linguistic material, Saffran et al. (1999) replaced the syllables by pure tones in order to create *words of tones*, which, once again, are concatenated continuously to each other to create a sequence. The tones were carefully chosen in such a way that the tone words and the chaining of these words in the sequence did not create a specific key context, and overall, they did not respect tonal rules nor did they resemble familiar three-tone sequences (e.g. the NBC television network's chimes). After exposition, both adults and 8-month-old infants performed above chance in the testing phase and performed as well as for linguistic-like sequences of syllables. Listeners thus succeeded in segmenting the tone stream and in extracting the *tone units*. Overall, Saffran et al.'s data suggest that statistical learning of different materials can be based on similar knowledge-acquisition processes.

To some extent, this finding can be considered as illustrating in the laboratory the processes that actually occur in real life for extensive exposure to environmental sounds, including music. It is obvious that a musical system such as the Western tonal system is more complex than the artificial grammar exposed in Figure 8. However, the opportunities to be exposed to sequences obeying this system from birth (and probably three or four months before birth) are so numerous that most of the rules of Western tonal music may be internalized through similar processes. Following this hypothesis, Western listeners may have acquired a sophisticated knowledge about Western tonal music, even though this knowledge remains at an implicit level of representation. A large set of empirical studies has actually demonstrated that musically untrained listeners (even young children) have internalized several aspects of the statistical regularities underlying pitch combinations that are specific to Western tonal music (Francès 1958; Thompson and Cuddy 1989; Krumhansl 1990; Cuddy and Thompson 1992a, 1992b; see Bigand 1993, for a review). Some extensions to other musical cultures have been realized in single studies (Castellano et al. 1984; Krumhansl et al. 1999). Once acquired, this implicit knowledge induces fast and rather automatic top-down influences on the perception and processing of Western pitch structures and renders musically untrained listeners "musically expert" for the processing of these pitch structures. One critical issue that remains is to formalize the functioning of these implicit learning processes in the auditory domain. The last section provides some first insights into this issue.

7. Neural Net Modeling of Implicit Learning of Western Pitch Structures

Pitch models and models of basic processes of pitch perception have been presented by de Cheveigné (Chapter 6). The present section focuses on models of music perception, and particularly artificial neural networks that simulate the learning and perceiving of musical structures. One of the principal advantages of artificial neural networks (e.g., connectionist models) is their capacity to learn representations, categorizations or associations between events. In these networks, the rules governing the material are not stored in an explicit (symbolic) way, but emerge from multiple constraints represented by the connections of the network, which have been learned by repeated exposure. In the following, some basics of neural net modeling will be reviewed first, followed by applications of neural nets to music perception. In this line, a model using Self-Organizing Maps will be presented as one example of neural nets simulating the learning and perception of musical structures.

An artificial neural network consists of units linked via synaptic connections of different strengths. The units are generally arranged into layers, with an input layer coding the incoming information. The input units are activated when a stimulus is presented to the network. This activation is sent via the connections to units in other layers. The strength of the transmitted activation is determined by the strengths of the connections (i.e.. weights of the connections). At the outset, a network does not incorporate any knowledge of the material, and this ignorance is reflected by connection weights set to random values. In parallel with biological networks, the learning process is defined as a modification of connection weights (Hebb 1949). Over the course of learning, the neural net units gradually become sensitive to different input events or categories. The learning process can be either supervised by an external teaching exemplar (e.g., the delta rule, McClelland and Rumelhart 1986) or unsupervised via passive exposure (e.g., competitive learning, Rumelhart and Zipser 1985). In supervised learning algorithms, an external teaching instance prescribes the target output that has to be reached and the weights of the connections are modified so that the model's output matches this target. In unsupervised learning algorithms, the network adapts its connections in such a way that it becomes sensitive to the underlying correlational structure between events of the training set: statistical regularities of the input material are extracted and events that often occur together are encoded and represented by the net units. As acculturation to musical structures presumably occurs without supervision in listeners, unsupervised learning algorithms seem to be well suited to modeling music cognition. The present section thus focuses on unsupervised learning algorithms, notably the competitive learning algorithm that provides the basis for learning in Self-Organizing Maps (SOMs, Kohonen 1995) and ART networks (ART stands for Artificial Resonance Theory, see Grossberg 1970, 1976).

For the competitive learning process, a set of training stimuli is presented repeatedly to the network and the learning takes place by competition among the units (Rumelhart and Zipser 1985). When an input is presented to the network, the input layer sends activation via the random connection weights to the units of the next layer. The unit receiving the maximum activation is defined as the 'winner' of the competition (e.g., best representing the current input) and is allowed to learn the representation of this input even better. Following the learning rule, the weights of the connections are updated in such a way that the links coming from active input units are reinforced and links coming from inactive input units are weakened. In other words, the response of the winning unit will subsequently be stronger for this same input pattern (or similar ones) and weaker for other patterns. In a similar way, other units learn to specialize their responses to other input patterns. The competitive learning algorithm represents the basis for learning in SOMs. In a network using an SOM, the units that are connected to the input layer follow a spatial layout: units are arranged in the form of a map and neighborhood relationships can be defined between map units as a function of the distance between these units. For learning in an SOM, not only the winning unit, but also the neighboring units are allowed to learn. At the beginning of learning, the size of the neighborhood is broad and over the course of learning its radius decreases. This learning process leads to topological mappings between input data and neural net units on the map: units that respond maximally for similar input patterns are located near each other on the map. Topological organization conforms to principles of cortical information processing, such as spatial ordering in sensory processing areas (e.g., somatosensory, vision, audition). Winter (Chapter 4) and Griffiths (Chapter 5) review the tonotopic organization of the auditory system that can be found at almost all major stages of processing (i.e., inner ear, auditory nerve, cochlear nucleus and auditory cortex).

Neural nets based on unsupervised learning algorithms are helpful in understanding how we learn musical patterns by mere exposure, how these patterns might be represented, and how this knowledge arising from acculturation influences perception. Recently, we used the SOM algorithm to simulate the cognitive capacity to extract underlying regularities and to become sensitive to musical structures via implicit and unsupervised learning processes (Tillmann et al. 2000). Western tonal musical pieces are based on a three-level organizational system containing notes, chords and keys (cf. Section 2). For the simulation of the implicit learning of tonal regularities, a hierarchical network with two SOMs was defined. The units of the input layer coded the incoming twelve pitch classes taking into consideration octave equivalence. Each unit of the input layer was connected to the units of the first SOM that in turn were connected to the units of the second SOM. Before learning, the weights of all connections were set to random values. During learning, chords and chord sequences were presented repeatedly to the input layer of the network. The connectionist algorithm changed connections in order to allow units to become specific detectors of combinations of events over short temporal windows. The structure of the system adapted to the regularities of tonal relationships through repeated exposure to musical material. Over the course of learning, the weights of the connections changed to reflect the regularities of co-occurrences between notes and between chords. The first connection matrix reflects which pitch (or virtual pitch) is part of a chord; the second matrix reflects which chord is part of a key. The units of the first SOM became specialized for the detection of chords and the units of the second SOM for the detection of keys. Both SOM layers showed a topological organization of the specialized units. In the chord layer, units representing chords that share notes (or subharmonics) were located close to each other on the map, but chords not sharing notes were not represented by neighboring units. In the key layer, the units specialized in the detection of keys were organized in a circle: keys sharing numerous chords and notes were represented close to each other on the map and the distance between keys increased with decreasing number of shared events. The organization of key units reflects the music theoretic organization of the circle of fifths: the more the keys are harmonically related, the closer they are on the circle (and on the network map). The learnability of this kind of higher-level topological map (cf. also Leman 1995) has led to the search for neural correlates of key maps (Janata et al. 2002).

The hierarchical SOM thus managed to learn Western pitch regularities via mere exposure. The entire learning process is guided by bottom-up information only and takes place without an external teacher. Furthermore, there are no explicit rules or concepts stored in the model. The connections between the three layers extract via mere exposure how the events appear together in music. The overall pattern of connections reflects how notes, chords and keys are interrelated. Just as for nonmusician listeners, the tonal knowledge is acquired without explicit instruction or external control. The input layer of the present network was based on units coding octave equivalent pitch classes. This model can be conceived as being on the top of other networks that have learned to extract pitch height from frequency (Sano and Jenkins 1991; Taylor and Greenhough 1994; Cohen et al. 1995) and octave-equivalent pitch classes from spectral representations of notes (Bharucha and Mencl 1996).

The SOM model integrates three levels of organization of the musical system. Other neural net models have been proposed in the literature that focused on either one or two organizational levels of music perception as for example pitch perception (Sano and Jenkins 1991; Taylor and Greenhough 1994), chord classification (Laden and Keefe

1991) or melodic sequence learning (Bharucha and Olney 1989; Page 1994; Krumhansl et al. 1999). More complex aspects of musical learning that are linked to the perception of musical style have been simulated by Gjerdingen (1990) using an ART network. Other models focused more strongly on the preprocessing of the auditory signal by auditory modules and on the bottom-up processes involved in learning and perception (Leman 1995, 2000; Leman and Carreras 1998).

As presented up to this point, one characteristic of neural networks is the adaptation to environmental structures and the learning of a representation of the regularities inherent in the environment. Another attractive characteristic of neural networks is the possibility of accounting for top-down influences and for the way they combine with bottom-up influences. In the language domain, neural net models of word recognition (McClelland and Rumelhart 1981; Rumelhart and McClelland 1982) and of speech recognition (Elman and McClelland 1984; McClelland and Elman 1986) simulate the top-down influences of the knowledge representation via activation reverberating between layers, notably interactive activation between higher level units (words) and lower level units (letters or phonemes). When, for example, part of the written word is missing, the reverberating activation helps to select possible candidates and to restore information in order to recognize the word. In music perception, Bharucha (1987) proposed a model (referred to as MUSACT) that relies on a comparable architecture including a mechanism of spreading activation. In MUSACT, note units are connected to chord units that in turn are connected to key units. When a stimulus is presented to the model, note units are activated and activation reverberates in the system until equilibrium is reached. This reverberation mechanism simulates the top-down influences and changes the activation patterns in favor of culturally defined relationships. For example, when a C Major chord (i.e., consisting of the notes C-E-G) is presented to the network, the activation pattern of the chord layer at the beginning reflects bottom-up influences only, notably the chord unit of E Major will be more activated than the chord unit of D Major because it shares one note with the stimulus chord (e.g. the note E), even if the chords C Major and D Major are harmonically more closely related. After reverberation, activation patterns change qualitatively and mirror theoretic Western harmonic hierarchies: the chord unit of D Major now receives stronger activation than the chord unit of E Major. The model thus predicts sensory priming for extremely short time spans with a facilitation of the E Major chord over the D Major chord, and cognitive priming for longer time spans with a facilitation of the D Major chord over the E Major chord. The model thus succeeds in simulating the time course of bottom-up and top-down activation as reported in short context priming by Tekman and Bharucha (1998, cf. Section 4). The MUSACT model has also simulated a set of priming data showing an effect of cognitive top-down influences in chord processing (Tillmann et al. 1998; Bigand et al. 1999; Tillmann and Bigand 2001; Bigand et al. 2002; Tillmann et al. 2003).

However, MUSACT represents an idealized end-state of an implicit learning process as it is based on music theoretic constraints and neither connections nor weights resulted from a learning process. As reported above, a representation of pitch regularities (as implemented by MUSACT) can be learned by passive self-organization (cf. Tillmann et al. 2000). In addition to testing this learned model with priming material (as was done with the MUSACT model), the SOM model has been tested for its capacity to simulate a variety of empirical data on the perceived relationships between and among notes, chords and keys. For these simulations, the experimental material of behavioral studies was presented to the network and the activation levels of the network units were interpreted as levels of tonal stability. The more a unit (i.e., a chord unit or a note unit) is activated, the more stable the musical event is in the corresponding context. For the experimental tasks, it was hypothesized that the level of stability affects performance (e.g., a more strongly activated, stable event is more expected or judged to be more similar to a preceding event). The simulated data covered a range of experimental tasks, notably similarity judgments, recognition memory for notes and chords, priming, electrophysiological measures for chords, and perception and detection of modulations and distances between keys. Overall, the simulations showed that activation in the learned SOM model mirrored the data of human participants in a range of experiments on the perception of tonality (cf. Tillmann et al. 2000, for more details of individual results).

The SOM simulations provide an example of the application of artificial neural networks to increasing our understanding of learning and representing knowledge about the tonal system and the influence of this knowledge on perception and processing. The learning process can be simulated by passive exposure to musical material, just as it is supposed to happen in nonmusician listeners. Once acquired, the knowledge influences perception. It is worth underlining that the SOM model simulates a set of context effects linked to the perception of notes and of chords: the same chord unit is activated with different levels of activation depending on the tonality of the preceding context. For example, the model simulates the principles of Contextual Distance and Contextual Asymmetry observed for human participants in the similarity judgments of chord pairs presented above in Section 3 (Krumhansl, Bharucha and Castellano 1982; Bharucha and Krumhansl 1983): the activation level of a chord unit changes as a function of the harmonic distance to the preceding key context and of the temporal order of presentation in the pair. The learned musical SOM network thus provides a low-dimensional and parsimonious representation of tonal knowledge: the contextual dependency of musical functions of an event emerges from the activation reverberating in the system, and the important stable events (e.g., musical prototypes and anchor points of a key) do not have to be stored separately in different units for each of the possible keys.

8 Conclusion

Throughout this chapter, we have documented that the processing of pitch structures is strongly context dependent. These context effects have been shown for the perception of specific attributes of musical sounds (such as musical stability), for the memorization of pitch (Section 3), as well as for the speed and accuracy of processing perceptual attributes related to the pitch dimension (e.g., sensory dissonance, musical timbre, phoneme, Section 4). These top-down influences involve rather specific electrophysiological responses and cortical areas that, interestingly, seem not to differ radically between language and musical domains. This suggests that some brain structures may be specialized in the integration of contextual information. The ecological interest of this specialization might be, notably, to enhance the processing of the pitch dimension. Most of the examples reported to illustrate context effects come from the music domain (similar examples are, of course, numerous in spoken language). Probably, composers have intuitively developed a musical system that taps into this incredible flexibility of the auditory system to attribute perceptual sound qualities as a function of the context in which they appear. The Western musical system takes advantage of this fundamental feature of the human brain: the ability to interpret sensory input differently depending on the current context. The Western tonal system is remarkable from this point of view. Despite a very small number of pitch classes (12), an infinite number of musical sequences can be composed by taking advantage of context effects in perception and by modifying the perceptual qualities of musical sounds as a function of the current context.

Of course, the question arises as to whether this feature of context dependency is unique for Western tonal music or whether other musical systems use it. In section 6, we argue that the observations made with musical material are just one example manifesting the broad competence of the human brain to internalize statistical regularities of environmental structures. Attempts to confirm implicit learning processes in the auditory domain have been presented with different sets of artificial materials. It is likely that new musical grammars will be internalized through passive exposure in roughly the same way. On the basis of this internalized knowledge, similar context effects will probably be reported in the future for contemporary music, as well as for other artificial sound structures that are derived from similar principles. Given the strength of the implication of implicit learning in auditory processing, we addressed in the previous section how these processes may be formalized in a neural net model.

We hope the present chapter will encourage new researchers to spend more time investigating the role played by implicit learning and top-down processes in the auditory domain. It is striking that learning and top-down processes are concepts that are missing in most current textbooks on audition (but see McAdams and Bigand, 1993 for auditory cognition and SHAR textbooks for learning, plasticity and development, Rubel Popper and Fay 1997; Parks Rubel Popper and Fay in press). Interestingly, audition is almost missing in the literature on implicit learning as well as on perceptual learning. As a consequence, the role of a listener's knowledge on auditory perception remains unclear and its importance is often disregarded or not even acknowledged. A better understanding of context effects in auditory perception has two possible main implications for the future. The first one is that adding knowledge and top-down processes in artificial models of auditory perception (including models of pitch processing) is likely to improve the models (see Carreras et al. 1999). There is strong evidence showing that the human brain manages to process pitch in a sophisticated way with the help of these top-down processes. The way this knowledge is represented in the mind, as well as the way this knowledge is acquired through exposure needs to be documented in more detail. Our preliminary findings on pitch structures suggest that similar processes of learning may then be implemented in artificial systems so that they manage to simulate top-down processes (for a discussion of this issue in visual perception see Herzog and Fahle 2002).

The second main implication, which is only beginning to be considered, concerns the rehabilitation of hearing-impaired listeners. Over the last year, research projects are emerging that investigate learning processes in hearing-impaired listeners and patients with cochlear implants. However, up to now, this research mainly focuses on perceptual processes in audition, as for example loudness perception (Philibert et al. 2002), sound localization and binaural hearing cues (Moore 2002), or on phoneme processing and single word processing without considering extended contexts (Clark 2002).

Numerous research has now generally established that top-down processes result in perceptual expectancies that enhance signal detection (e.g., Howard et al., 1984 for pitch detection threshold) and signal processing in all sensory modalities. Reinforcing these top-down processes in hearing-impaired listeners represents a key concern since the top-down processes should contribute to a compensation of the failure of sensory processes. Of course, this kind of strategy occurs naturally in hearing-impaired listeners and is usually developed by auditory teaching methods. However, it is obvious that the more the scientific community knows about the functioning of top-down processing as well as the functioning of learning processes in the auditory domain, the more efficient such teaching methods will be. Several factors that influence implicit auditory learning need to be studied in the auditory domain, and the benefits drawn from implicit versus explicit training should be evaluated. The outcome of this research will also have implications for research in technical engineering devoted to the remediation of hearing-impaired listeners. Up to now, considerable effort has been made for the investigation and the improvement of reception and coding of auditory signals at peripheral levels of processing. It is now necessary to invest in technical support favoring the development and improvement of higher level, cognitive processes and perceptual top-down strategies that will then help the listener to restore missing or deficient sensory signals, to the extent that such is possible. During the last decade, cognitive engineering devoted to training techniques has been developed and strongly improved in several domains. These new technologies offer considerable possibilities to define auditory learning programs that will encourage implicit learning of auditory sound and scene structures. In order to take best advantage of these new technologies for hearing-impaired listeners, it is important that the scientific community involved in audition reinforces considerably the research programs on perceptual and statistical learning in audition.

9 Summary

This chapter focused on the effect of listeners' knowledge on the processing of pitch structures. In section 2, several examples taken from vision and audition illustrated the differences between sensory processes and knowledge-driven processes (also referred to as bottom-up and top-down processes). Empirical evidence for top-down effects on the processing of pitch structures (perception and memorization) was presented in sections 3 and 4. It has been shown that a long series of musical notes can be perceived differently as a function of the musical key context in which the notes occur, and that the speed and accuracy with which some qualities of musical chords (e.g., consonance versus dissonance, harmonic spectra) are processed depends on the musical function of the chord in the current context. The neurophysiological structures implied in top-down processes in music perception were reviewed in section 5. Sections 6 and 7 addressed the origins of knowledge-driven processes. It was argued that a fundamental characteristic of the human brain is to internalize the statistical regularities of the external environment. In the case of music, intense passive exposure to Western musical pieces results in an implicit knowledge of Western musical regularities, which, in turn, govern the processing of pitch structures. The way implicit learning processes might be formalized by neural net models was developed in section 7. In conclusion, it was emphasized that the context effects observed in music perception reflect the considerable importance of top-down processes in the auditory domain. This conclusion has several implications, notably for artificial models of pitch processing as well as for auditory training methods designed for hearing-impaired listeners.

References

Abrams M, Reber AS (1988) Implicit learning: robustness in the face of psychiatric disorders. J Psycholinguist Res 17: 425-439.

Adams RB, Janata P (2002) A Comparison of Neural Circuits Underlying Auditory and Visual Object Categorization. NeuroImage 16: 361–377.

Allen R, Reber AS (1980) Very long term memory for tacit knowledge. Cognition 8: 175-185.

Ballas JA, Mullins T (1991) Effects of context on the identification of everyday sounds. Hum Perfor 4: 199-219.

Bartlett JC, Dowling WJ (1980a) The recognition of transposed melodies: A key-distance effect in developmental perspective. J Exp Psychol Hum Percept Perform 6: 501-515.

Bartlett JC, Dowling WJ (1988) Scale structure and similarity of melodies. Music Percep 5: 285-314.

Berlioz H (1872) Mémoires. Paris, France: Flammarion.

Besson M, Faïta F (1995) An event-related potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. J Exp Psychol Hum Percept Perform 21: 1278-1296.

Bharucha JJ (1984a) Event hierarchies, tonal hierarchies, and assimilation: A reply to Deutsch and Dowling. J Exp Psychol Gen 113: 421-425.

Bharucha JJ (1984b) Anchoring effects in music: The resolution of dissonance. Cognit Psychol 16: 485-518.

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

Bharucha JJ (1996) Melodic anchoring. Music Perception 13: 383-400.

Bharucha JJ, Krumhansl CL (1983) The representation of harmonic structure in music: Hierarchies of stability as a function of context. Cognition 13: 63-102.

Bharucha JJ, Olney KL (1989) Tonal cognition, artificial intelligence and neural nets. Contemporary Music Review 4: 341-356.

Bharucha JJ, Mencl WE (1996) Two Issues in auditory cognition: Self-organization of octave categories and pitch-invariant pattern recognition. Psychological Science 7: 142-149.

Bharucha JJ, Stoeckig K (1987) Priming of chords: Spreading activation or overlapping frequency spectra? Perception & Psychophysics 41: 519-524.

Biederman I (1987) Recognition-by-components: A theory of human image understanding. Psychol Rev 94: 115-147.

Bigand E (1993) Contributions of music to research on human auditory cognition. In McAdams S and Bigand E (eds), Thinking in Sound: the cognitive psychology of human audition. Oxford: Claredon Press, pp. 231-273.

Bigand E (1997) Perceiving musical stability: the effect of tonal structure, rhythm, and musical expertise. J Exp Psychol Hum Percept Perform 23: 808-822.

Bigand E, Pineau M (1996) Context effects on melody recognition: A dynamic interpretation. Current Psychology of Cognition 15: 121-134.

Bigand E, Pineau M (1997) Global context effects on musical expectancy. Perception & Psychophysics 59: 1098-1107.

Bigand E, Parncutt R, Lerdahl F (1996) Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. Perception & Psychophysics 58: 125-141.

Bigand E, Perruchet P, Boyer M (1998) Implicit learning of an artificial grammar of musical timbres. Current Psychology of Cognition 17: 577-600.

Bigand E, Madurell F, Tillmann B, Pineau M (1999) Effect of global structure and temporal organization on chord processing. J Exp Psychol Hum Percept Perform 25: 184-197.

Bigand E, Tillmann B, Poulin B, D'Adamo DA (2001) The effect of harmonic context on phoneme monitoring in vocal music. Cognition 81: B11-B20.

Bigand E, Poulain B, Tillmann B, D'Adamo D (2003) Cognitive versus sensory components in harmonic priming effects. J Exp Psychol Hum Percept Perform 29: 159-171.

Burns EM (1999) Intervals, scales and tuning In Deutsch D (ed) The Psychology of Music, 2nd edition, Academic Press: San Diego, pp. 215-264.

Cabeza R, Kingstone A (2001) Handbook of functional neuroimaging in cognition. Cambridge MA: MIT.

Caplan D, Alpert N, Waters G (1999) PET Studies of Syntactic Processing with Auditory Sentence Presentation. NeuroImage 9: 343–351.

Carreras F, Leman M, Lesaffre M (1999) Automatic harmonic description of musical signals using schema-based chord decomposition. Journal of New Music Research 28: 310-333.

Castellano MA, Bharucha JJ, Krumhansl CL (1984) Tonal hierarchies in the music of North India. J Exp Psychol Gen 113: 394-412.

Chailley J (1951) Traité historique d'analyse musicale. Paris: Leduc.

Chollet SDV (2001) Impact of training on beer flavor perception and description: Are trained and untrained subjects really different? J of Sensory Studies 16: 601-618.

Clark GM (2002) Learning to understand speech with cochlear implant. In Fahle M and Poggio T (eds), Perceptual Learning. Cambridge, MA: MIT, pp. 147-160.

Cohen MA, Grossberg S, Wyse LL (1995) A spectral network model of pitch perception. J Acoust Soc Am 98: 862-879.

Collins AM, Quillian MR (1969) Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior 8: 241-248.

Cuddy LL, Thompson WF (1992a) Asymmetry of perceived key movement in chorale sequences: Converging evidence from a probe-note analysis. Psychological Research 54: 51-59.

Cuddy LL, Thompson WF (1992b) Perceived key movement in four-voice harmony and single voices. Music Perception 9: 427-438.

Deutsch, D (1982) (ed) The Psychology of Music, New York: Academic Press.

Deutsch, D (1999) (ed) The Psychology of Music, Second Edition, San Diego: Academic Press.

DeWitt LA, Samuels AG (1990) The role of knowledge-based expectation in music perception: Evidence from musical restoration. J Exp Psychol Gen 119: 123-144.

Dowling WJ, Harwood DL (1986) Music Cognition. Orlando, Florida: Academic Press.

Dowling WJ (1972) Recognition of melodic transformations: Inversion, retrograde, and retrograde inversion. Perception & Psychophysics 12: 417-421.

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

Dowling WJ (1986) Context effects on melody recognition: Scale-step versus interval representations. Music Perception 3: 281-296.

Dowling WJ (1991) Tonal strength and melody recognition after long and short delays. Perception & Psychophysics 50: 305-313.

Dowling WJ, Bartlett JC (1981) The importance of interval information in long-term memory for melodies. Psychomusicology 1: 30-49.

Dowling WJ, Kwak S, Andrews MW (1995) The time course of recognition of novel melodies. Perception & Psychophysics 57: 136-149.

Elman JL, McClelland JL (1984) The interactive activation model of speech perception. In N Lass N (ed), Language and speech. New York: Academic Press, pp. 337-374.

Embick E, Marantz A, Miyashita Y, O'Neil W, Sakai KL (2000) A syntactic specialization for Broca's area. Proceedings of the New York Academy of Science 97: 6150-6154.

Fiez JA, Balota DA, Raichle ME, Petersen SE (1999) Effects of Lexicality, Frequency and Spelling-to-Sound Consistency on the Functional Anatomy of Reading. Neuron 24: 205-218.

Fisher GH (1967) Perception of ambiguous stimulus materials. Perception and Psychophysics 2 :421-422.

Francès R (1958) La perception de la musique (2° ed.) Paris: Vrin [1984 The perception of music (Dowling trans.), Hillsdale, NJ: Earlbaum].

Friederici AD (1995) The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysicological data. Brain and Language 50: 259-281.

Friederici AD, Meyer M, van Cramon DY (2000) Auditory language comprehension: An event-related fMRI study on the processing of syntactic and language information. Brain and Language 74: 289-300.

Fuster JM (2001) The prefrontal cortex - An update: Time is of the essence. Neuron 30: 319-333.

Gjerdingen RO (1990) Categorization of musical patterns by self-organizing neuronlike networks. Music Perception 8: 339-370.

Griffiths TD, Johnsrude I, Dean JL, Green GGR (1999) A common neural substrate for the analysis of pitch and duration pattern in segmented sound? Neuroreport 10: 3825-3830.

Grossberg S (1970) Some networks that can learn, remember and reproduce any number of complicated space-time patterns. Studies in Applied Mathematics 49: 135-166.

Grossberg S (1976) Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. Biological Cybernetics 23: 121-134.

Hebb DO (1949) The organization of behavior. New York: Wiley.

Helmholtz HL v (1885/1954) On the sensations of tone as a physiological basis for the theory of music (A. J. Ellis, Trans.) London: Longmans, Green.

Herzog M, Fahle M (2002) Top-down information and models of perceptual learning. In Fahle M and Poggio T (eds), Perceptual Learning. Cambridge, MA: MIT, pp. 367-380.

Howard JH, O'Toole AJ, Parasuraman R, Bennett KB (1984) Pattern-directed attention in uncertain-frequency detection. Perception & Psychophysics 35: 256-264.

Huron D, Parncutt R (1993) An improved model of tonality perception incorporating pitch salience and echoic memory. Psychomusicology 12: 154-171.

Janata P (1995) ERP measures assay the degree of expectancy violation of harmonic contexts in music. J Cogn Neurosci 7: 153-164.

Janata P, Birk J, Van Horn JD, Leman M, Tillmann B, Bharucha JJ (2002) The cortical topography of tonal structures underlying Western music. Science 298: 2167-2170.

Kiehl KA, Laurens KR, Duty TL, Forster BB, Liddle PF (2001) Neural sources involved in auditory target detection and novelty processing: An event-related fMRI study. Psychophysiology 38: 133-142.

Koelsch S, Gunter T, Friederici AD (2000) Brain indices of music processing: "nonmusicians" are musical. J Cogn Neurosci 12: 520-541.

Koelsch S, Gunter TC, v Cramon DY, Zysset S, Lohmann G, AD. F (2002) Bach speaks: a cortical "language-network" serves the processing of music. NeuroImage 17: 956-966.

Kohonen T (1995) Self-Organizing Maps. Berlin: Springer.

Koustaal W, Wagner AD, Rotte M, Maril A, Buckner RL, Schacter DL (2001) Perceptual specificity in visual object priming: functional magnetic resonance imaging evidence for a laterality difference in fusiform cortex. Neuropsychologia 39: 184-99.

Krumhansl CL (1990) Cognitive Foundations of Musical Pitch. Oxford: Oxford University Press.

Krumhansl CL, Bharucha JJ, Castellano M (1982) Key distance effects on perceived harmonic structure in music. Perception & Psychophysics, 32, 96-108.

Krumhansl CL, Bharucha JJ, Kessler EJ (1982) Perceived harmonic structures of chords in three related keys. J Exp Psychol Hum Percept Perform, 8, 24-36.

Krumhansl CL, Kessler E (1982) Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. Psychol Rev, 89, 334-368.

Krumhansl CL, Louhivuori J, Toivianinen P, Jarvinen T, Eerola T (1999) Melodic expectation in Finnish spiritual folk hymns: Converging evidence of statistical, behavioral and computational analyses. Music Perception, 17, 151-195.

Kutas M, Hillyard SA (1980) Event-related brain potentials to semantically inappropriate and surprisingly large words. Biological Psychology, 11, 99-116.

Laden B, Keefe DH (1991) The representation of pitch in a neural net model of chord classification. In P Todd and G Loy (Eds.), Music and Connectionism (pp. 64-83) Cambridge MA: MIT Press.

Leman M (1995) Music and Schema Theory. Berlin: Springer.

Leman M, Carreras F (1998) Schema and Gestalt: Testing the Hypothesis of Psychoneural Isomorphism by Computer Simulation. In M Leman (Ed.), Music, Gestalt, and Computing (pp. 144-168) Berlin: Springer.

Leman M. (2000). An auditory model of the role of short-term memory in probe tone rating. Music Perception, 17, 437-460.

Leman M, Lessaffre M, Tanghe K (2000) The IPEM Toolbox Manual. University of Ghent, IPEM-Dept. of Musicology: IPEM.

Lerdahl F (1988) Tonal Pitch Space. Music Perception, 5, 315-345.

Lerdahl F (2001) Tonal Pitch Space: Oxford University Press.

Lerdahl F, Jackendoff R (1983) A generative theory of tonal music. Cambridge, MA: MIT press.

Linden DEJ, Prvulovic D, Formisano E, Vollinger M, Zanella FE, Goebel R, Dierks T (1999) The functional neuroanatomy of target detection: An fMRI study of visual and auditory oddball tasks. Cerebral Cortex 9: 815-823.

Maess B, Koelsch S, Gunter T, Friederici AD (2001) ‘Musical syntax’ is processed in the Broca’s area: An MEG-study. Nature Neuroscience 4: 540-545.

Marr D (1982) Vision: A computational investigation into the human representation and processing of visual information. San Francisco: Freeman.

McAdams S, Bigand E (1993) Thinking in sound. Oxford: Claredon Press.

McClelland JL, Rumelhart DE (1981) An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. Psychol Rev 86: 287-330.

McClelland JL, Elman JL (1986) The TRACE model of speech perception. Cognit Psychol 18: 1-86.

McClelland JL, Rumelhart DE (1986) Parallel distributed processing: Exploration in the Microstructure of Cognition (Vol. 2) Cambridge, MA: MIT Press.

Moore DR (2002) Auditory development and the role of experience. Br Med Bull 63: 171-181.

Morrot G, Brochet F, Dubourdieu D (2001) The colors of odors. Brain and Language 78: 309-320.

Müller R-A, Kleinhans N, Courchesne E (2001) Broca's area and the discrimination of frequency transitions: A functional MRI study. Brain and Language 76: 70-76.

Page MA (1994) Modeling the perception of musical sequences with Self-Organizing neural networks. Connection Science 6: 223-246.

Palmer C, Krumhansl CL (1987a) Independent temporal and pitch structures in determination of musical phrases. J Exp Psychol Hum Percept Perform 13: 116-126.

Palmer C, Krumhansl CL (1987b) Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing, and familiarity. Perception & Psychophysics 41: 505-518.

Parks TN, Rubel AW, Popper AN, Fay RR (in press) Plasticity of the auditory system. SHAR, Springer.

Parncutt R (1988) Revision of Terhardt's psychoacoustical model of the roots of a musical chord. Music Perception 6: 65-94.

Parncutt, R (1989) Harmony: A psychoacoustical approach. Berlin: Springer.

Patel AD, Gibson E, Ratner J, Besson M, Holcomb PJ (1998) Processing syntactic relations in language and music: An event-related potential study. J Cogn Neurosci 10: 717-733.

Perruchet P, Pacteau C (1990) Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? J Exp Psychol Gen 119: 264 - 275.

Perruchet P, Vinter A, Gallego J (1997) Implicit learning shapes new conscious percepts and representations. Psychonomic Bulletin and Review 4: 43-48.

Philibert B, Collet L, Vesson J, Veuillet E (2002) Intensity-related performances are modified by long-term hearing aid use: a functional plasticity? Hear Res 165: 142-151.

Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, Gabrieli JDE (1999) Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. Neuroimage 10: 15-35.

Pugh KR, Shaywitz BA, Fulbright RK, Byrd D, Skudlarski P, Katz L, Constable RT, Fletcher J, Lacadie C, Marchione K, Gore JC (1996) Auditory selective attention: An fMRI investigation. NeuroImage 4: 159-173.

Rameau J-P (1721) Treatise on harmony (P Gosset, Trans.) (1971 ed) New York: Dover.

Reber AS (1967) Implicit learning of artificial grammars. Journal of Verbal Learning and Verbal Behavior 6: 855-863.

Reber AS (1989) Implicit learning and tacit knowledge. J Exp Psychol Gen 118: 219-235.

Reber AS (1992) The cognitive unconscious: An evolutionary perspective. Consciousness and Cognition 1: 93-133.

Reber AS, Walkenfeld F, Hernstadt R (1991) Implicit and explicit learning: individual differences and IQ. J Exp Psych Learning Memory Cognition 17: 888-896.

Regnault P, Bigand E, Besson M (2001) Event-related brain potentials show top-down and bottom-up modulations of musical expectations. J Cogn Neurosci 13: 241-255.

Rosch E (1975) Cognitive reference points. Cognit Psychol 7: 532-547.

Rosch E (1979) On the internal structure of perceptual and semantic categories. In Moore TE (ed) Cognitive development and the acquisition of language. New York: Academic Press.

Rosen, C (1971) Le style classique, Haydn, Mozart, Beethoven (M. Vignal, Trans.) Paris: Gallimard.

Rubel EW, Popper AN, Fay RR (eds) (1997) Development of the auditory system. SHAR, vol. 9. Springer.

Rumelhart DE, McClelland JL (1982) An interactive activation model of context effects in letter perception. Part 2. Psychol Rev 89: 60-94.

Rumelhart DE, Zipser D (1985) Feature discovery by competitive learning. Cognitive Science 9: 75-112.

Saffran J, Aslin R, Newport E (1996) Statistical learning by 8-month-old infants. Science 274: 1926-1928.

Saffran JR, Johnson EK, Aslin RN, Newport EL (1999) Statistical learning of tone sequences by human infants and adults. Cognition 70: 27-52.

Saffran JR, Newport EL, Aslin RN, Tunick RA, Barrueco S (1997) Incidental language learning. Psychological Science 8: 101-105.

Sano H, Jenkins BK (1991) A neural network model for pitch perception. In Todd N and Loy G (eds), Music and Connectionism. Cambridge, MA: MIT Press, pp. 42-49.

Sasaki T (1980) Sound restoration and temporal localization of noise in speech and music sounds. Tohoku Psychological Folia 39: 70-88.

Schenker H (1935) Der Freie Satz. Neue musikalische Theorien und Phantasien (N Meeùs, Trans.) Liège: Margada.

Shepard RN (1964) Circularity in judgments of relative pitch. J Acoust Soc Am 36: 2346-2353

Taylor I, Greenhough M (1994) Modeling pitch perception with adaptive resonance theory artificial neural networks. Connection Science 6: 135-154.

Tekman HG, Bharucha JJ (1992) Time course of chord priming. Perception & Psychophysics 51: 33-39.

Tekman HG, Bharucha JJ (1998) Implicit knowledge versus psychoacoustic similarity in priming of chords. J Exp Psychol Hum Percept Perform 24: 252-260.

Thompson WF, Cuddy LL (1989) Sensitivity to key change in chorale sequences: A comparison of single voices and four-voice harmony. Music Perception 7: 151-168.

Tillmann B, Bigand E (2001) Global relatedness effect in normal and scrambled chord sequences. J Exp Psychol Hum Percept Perform 27: 1185-1196.

Tillmann B, Bharucha JJ (2002) Harmonic context effect on temporal asynchrony detection. Perception & Psychophysics 64: 640-649.

Tillmann B, Bigand E, Pineau M (1998) Effects of global and local contexts on harmonic expectancy. Music Perception 16: 99-117.

Tillmann B, Bharucha JJ, Bigand E (2000) Implicit learning of tonality: A Self-Organizing approach. Psychol Rev 107: 885-913.

Tillmann B, Bharucha JJ, Bigand E (2001) Implicit Learning of Regularities in Western Tonal Music by Self-Organization. In: French R, Sougné, J (eds) Proceedings of the Sixth Neural Computation and Psychology Workshop: Evolution, Learning, and Development. Perspectives in Neural Computing series, London: Springer, pp. 175-184.

Tillmann B, Janata P, Bharucha JJ (2003) Inferior frontal cortex activation in musical priming. Cognitive Brain Research 16: 145-161.

Wagner AD, Paré-Blagoev EJ, Clark J, Poldrack RA (2001) Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. Neuron 31: 329–338.

Wagner AD, Koustaal W, Maril A, Schacter DL, Buckner RL (2000) Task-specific repetition priming in left inferior prefrontal cortex. Cerebral Cortex 10: 1176-1184.

Warren RM (1970) Perceptual restoration of missing speech sounds. Science 167: 392-393.

Warren RM (1999) Auditory Perception: A new analysis and synthesis. Cambridge, UK: Cambridge University Press.

Warren RM, Sherman GL (1974) Phonemic restoration based on subsequent context. Perception & Psychophysics 16: 150-156.

West WC, Dale AM, Greve D, Kuperberg G, Waters G, Caplan D (2000) Cortical activation during a semantic priming lexical decision task as revealed by event-related fMRI. Paper presented at the Human Brain Mapping Meeting.

Zatorre RJ, Evans AC, Meyer E (1994) Neural mechanisms underlying melodic perception and memory for pitch. J Neuroscience 14: 1908–1919.

Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch processing in speech perception. Science 256: 846–849.

Figure Captions

Figure 1. Example of the importance played by top-down process in vision by Fishler (1967, Psychonomic Society, reproduced with permission). See explanations in the text (section 2).

Figure 2. Examples of the importance played by top-down process in reading. See explanations in the text (section 2). (the top figure is adapted from Figure 3.41 CRIDER, ANDREW B., PSYCHOLOGY, 4th Edition, © 1993. Reprinted by permission of Pearson Education, Inc., Upper Saddle River, NJ.).

Figure 3. Schematic representation of the three organizational levels of the tonal system. Top) 12 pitch classes, followed by the diatonic scale in C Major. Middle) construction of three major chords, followed by the chord set in the key of C Major key. Bottom) relations of the C Major key with close major and minor keys (left) and with all major keys forming the circle of fifths (right). (Tones are represented in italics, minor and major chords/keys in lower and upper case respectively) (from Tillmann et al. 2001, Implicit Learning of Regularities in Western Tonal Music by Self-Organization (pp. 175-184), Figure 1, in: Proceedings of the Sixth Neural Computation and Psychology Workshop: Evolution, Learning, and Development, © Springer)

Figure 4. Probe tone ratings for the 12 pitch classes in C Major and F# Major contexts from Krumhansl and Kessler (1982, American Psychological Association, adapted with permission).

Figure 5. Top) The two melodies T1 and T2 used in Bigand (1996) with their 23 stop notes on which musical stability ratings were given by participants. Bottom) Musical stability ratings from musician participants superimposed on the two melodies T1 and

T2 (from Bigand 1996, Fig. 2, American Psychological Association, adapted with permission)

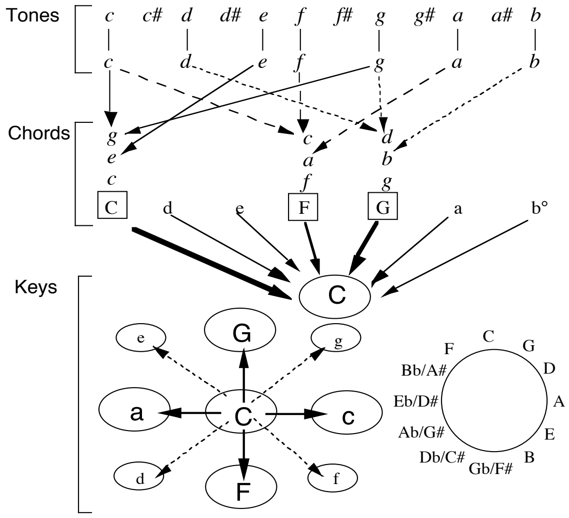
Figure 6. Representations based on chord similarity ratings in the contexts of C Major, F# Major and A Major (Reprinted from Cognition, 13, Bharucha and Krumhansl, The representation of harmonic structure in music: Hierarchies of stability as a function of context, 63-102, Copyright (1983), with permission from Elsevier; and from Perception & Psychophysics, 32, Krumhansl et al. Key distance effects on perceived harmonic structure in music, 96-108 Copyright (1982) with permission from Psychonomic Society). The closer chords are in the plane, the more similar they are rated to be. Roman numbers refer to the functions of the chords in the key. They reflect the degree of the scale on which the chords are constructed, e.g. I for tonic, IV for subdominant, V for dominant, and ii, iii, vi and vii for chords constructed on second, third, sixth and seventh degree of the scale.

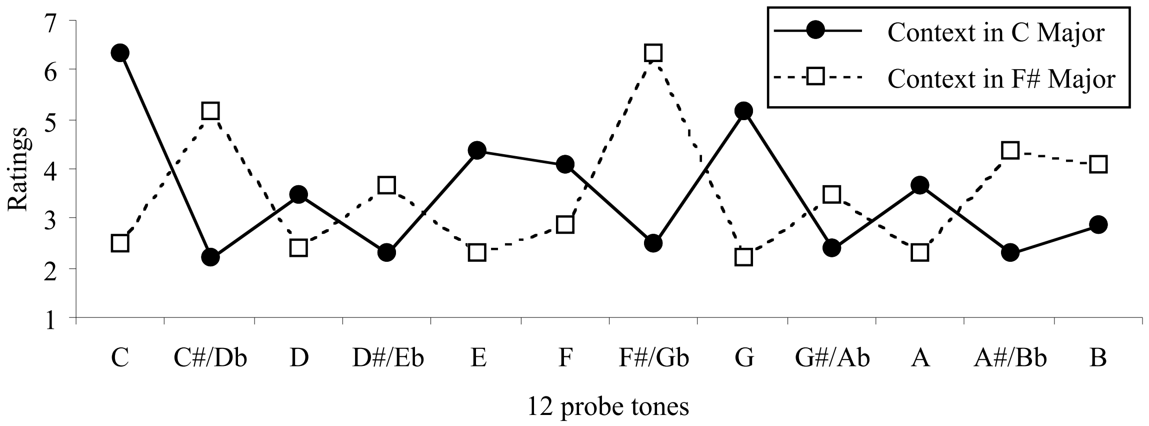
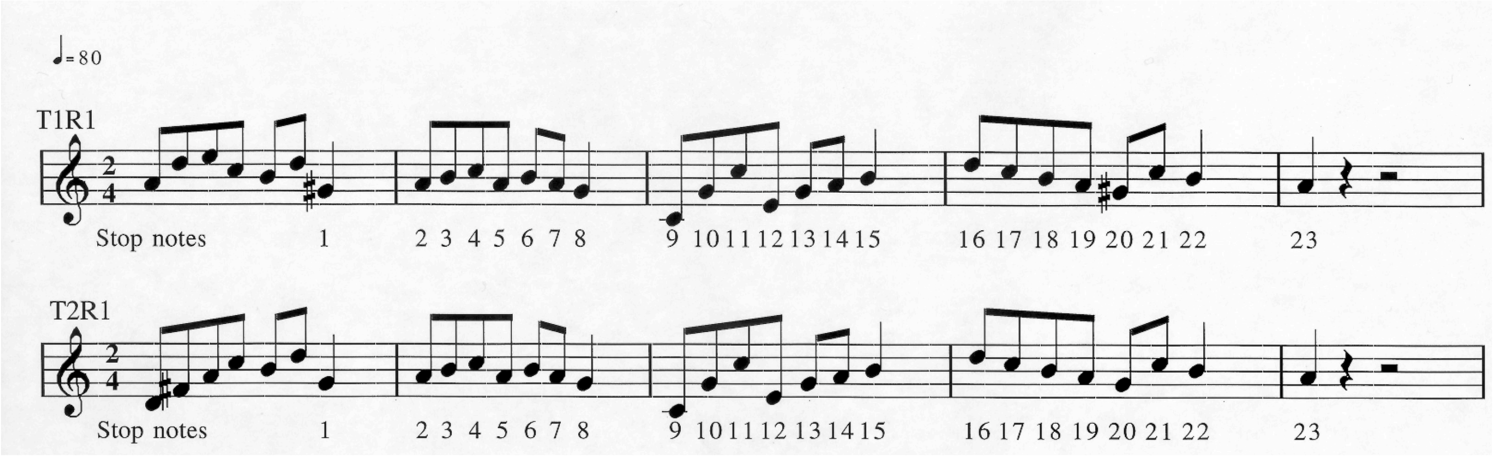
Figure 7. Top) One example of the eight-chord sequence used by Bigand and Pineau (1997) for the highly expected condition ending on the tonic chord (I) and the weakly expected condition ending on the subdominant chord (IV) (from Bigand 1999, Fig. 1, American Psychological Association, adapted with permission). Bottom) An example of the 14-chord sequences in the highly expected condition, the weakly expected condition and the moderately expected condition (adapted with permission from Bigand 1999, Fig. 6, American Psychological Association, adapted with permission)

Figure 8. Example of a finite state grammar generating letter sequences. The sequence XSXXWJX is grammatical whereas the sequence XSQSW is not.

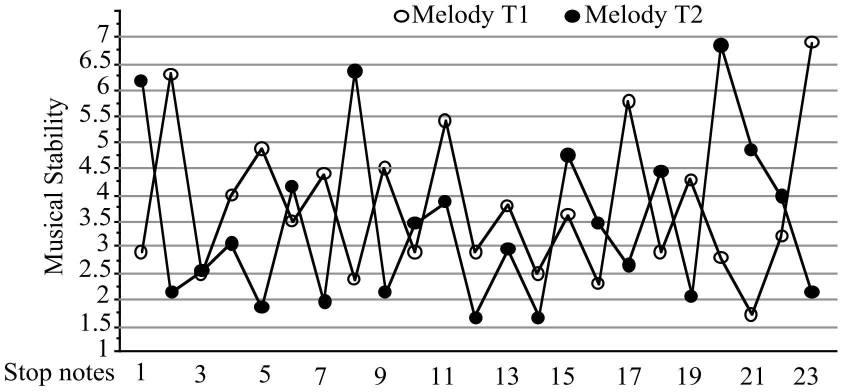
my phone number is
area code 603, 6461569, please call !

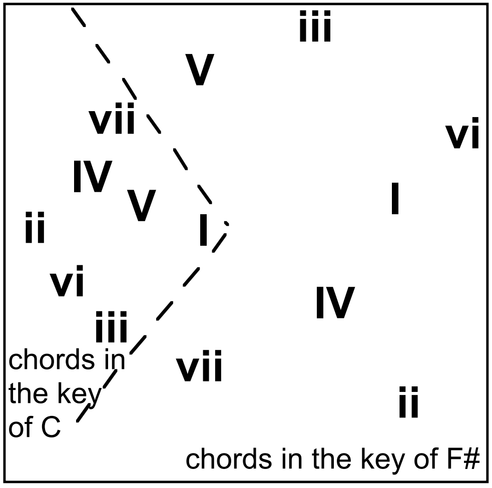
ABC 12 13 14


Tones
c c# d d# e f f# g g# a a# b
c d e f g a b
Chords
g e c
c a f
d b g
C d e F G a b°
Keys
C
e G g
a C c
d F f
C G D A E B Gb/F# Db/C# Ab/G# Eb/D# Bb/A# F

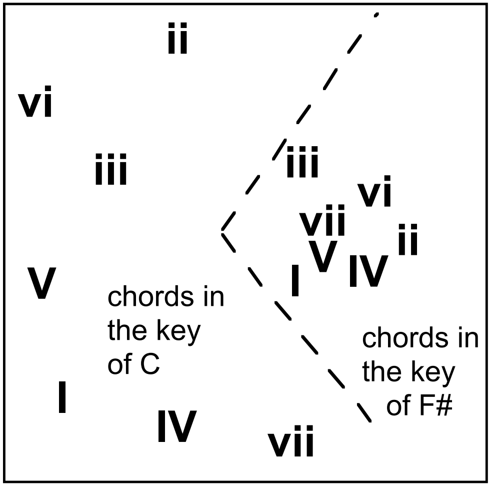

Ratings
7
6
5
4
3
2
1
C
C#/Db
D
D#/Eb
E
F
F#/Gb
G
G#/Ab
A
A#/Bb
B
12 probe tones
Context in C Major
Context in F# Major


♩= 80
T1R1
Stop notes 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
T2R1
Stop notes 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

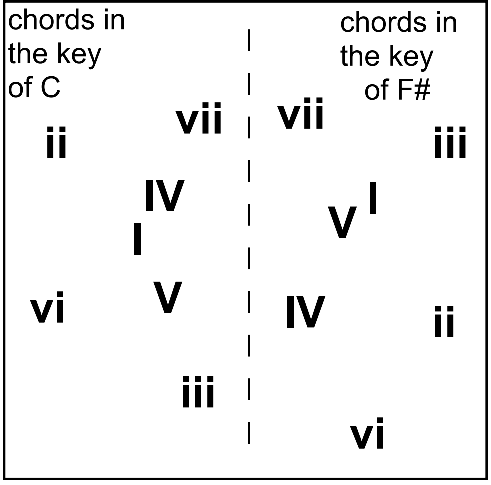

○Melody T1
●Melody T2
Musical Stability
7
6.5
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
Stop notes
1
3
5
7
9
11
13
15
17
19
21
23



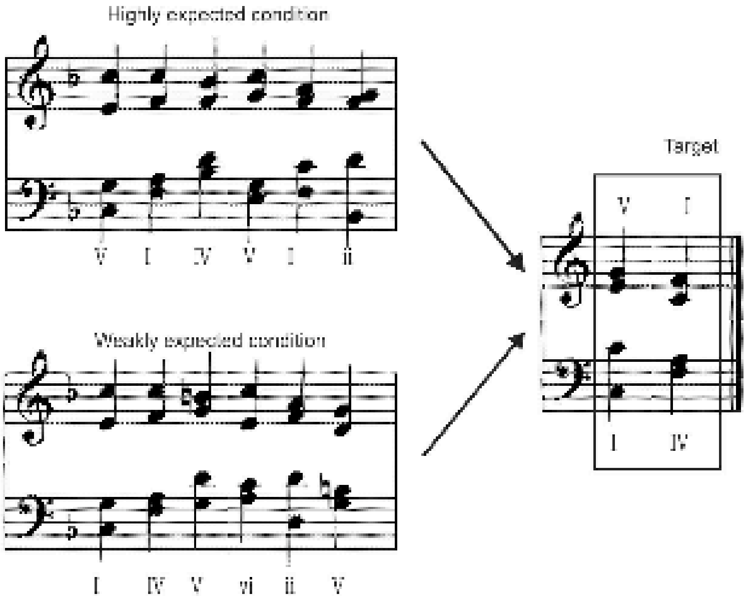
C Major Context

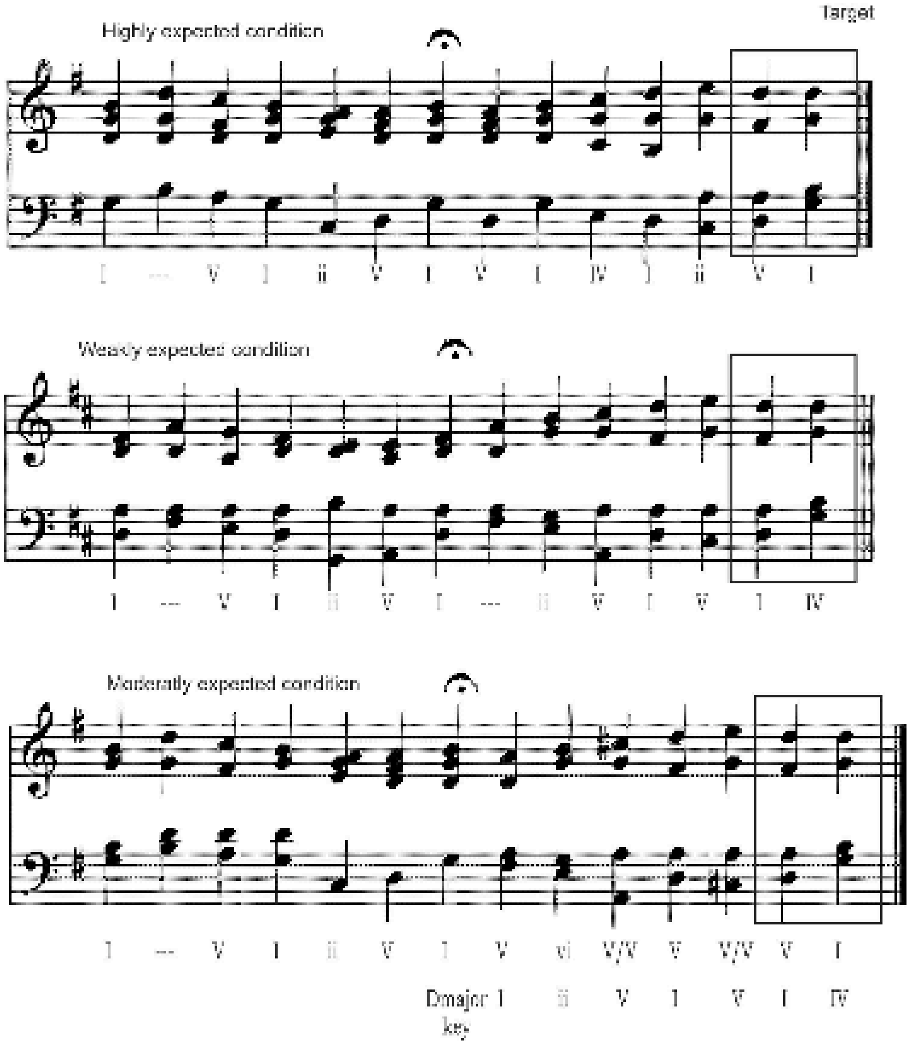


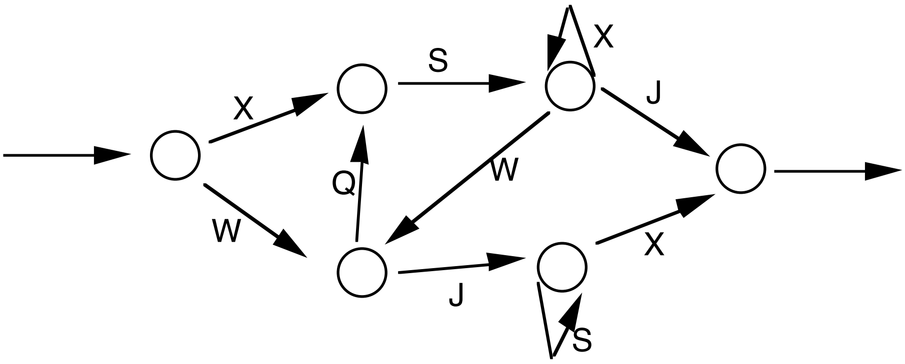
F# Major Context



A Major Context


Highly expected condition
V I IV V I ii
Weakly expected condition
I IV V vi ii V
Target
V I
I IV


Target
Highly expected condition
I --- V I ii V I V I IV I ii V I
Weakly expected condition
I --- V I ii V I --- ii V I V I IV
Moderatly expected condition
I --- V I ii V I V vi V/V V V/V V I
Dmajor key I ii V I V I IV


X
W
S
Q
W
X
J
J
X
S