Hierarchical perception of melody

Jens Hjortkjær[*1]

[*]Department of Arts and Cultural Studies, University of Copenhagen, Denmark
[1]jensh@hum.ku.dk

ABSTRACT

Two experiments were designed to investigate the hierarchical perception of melodies. The hierarchical structure of tonal melodies is formally described in Lerdahl and Jackendoff's *Generative Theory of Tonal Music* (GTTM). The experiments were designed to assess different aspects of GTTM time span reduction quantitatively. Subjects were presented with pairs of melodies in a same-different task. A short novel and a longer original tonal melody could differ within the harmonic context at single tones on one of four time span reduction levels. It was found that detection rate of the melodic deviances correlated strongly with time span level. The same method was employed on an atonal and a poorly composed tonal melody. Detection rates were lower for the atonal but not for the ill-formed melody, and responses still varied significantly with position of the deviant tone. Generally, the correlation of responses with reductional level was found to be much stronger than with alternative musical parameters that could be thought to influence detection rate, such as metrical position. The obtained quantitative measures of hierarchical levels suggest that these are subject to variations not normally envisaged by music analysis. It is suggested that hierarchical levels may be graded in nature, with the perceived hierarchical depth depending on the nature of the melodic sequence.

I.   INTRODUCTION

The ability to perceive musical melodies as meaningful sequences is thought to depend on the formation of a mental hierarchy of events. This is not only the case with music - understanding the narrative of a novel or a film depends on the mental ability to distinguish the importance of perceived events in relation to the plot. Structural hierarchies of melodies have been described formally within music theory. The music theory of H. Schenker and the *Generative Theory of Tonal Music* (GTTM) by F. Lerdahl and R. Jackendoff (1983) are prominent approaches. 'Time span reduction' is the analytical procedure in GTTM by which the structural weight of tones in a melody is assigned. Tones in the musical surface that are considered ornamental are reduced out to arrive at a more fundamental structural level of the melody. This analytical process is iterated to elucidate still deeper structural levels.

There has been some experimental support for the cognitive reality of hierarchical perception in tonal music (Deliège, 1987; Castellano *et al.* 1984; Bharucha, 1987; Large *et al.,* 1995, Bharucha & Krumhansl, 1983; Palmer & Krumhansl 1987). Serafine *et al.* (1989) presented melodies and performances of structural reductions and asked subjects to match a melody with its reduction. They found that listeners were 'moderately able' to identify the correct reduction. Dibben (1994) used a similar method with similar results. The relatively low recognition rate could be a consequence of methodology. Structural reductions may describe cognitive processes without being perceivable as an instance of the melody when performed. Both GTTM and Schenkerian theory claim that structural levels are themselves music (in schenkerian theory reduced levels must adhere to voice leading principles in the same way that actual melodies do). At deeper reductional levels the reduction is descriptive of a range of tonal melodies (the deepest levels of reduction - the Schenkerian *Urlinie* - is precisely a description of the abstract structure that define the work as tonal). It will therefore become increasingly difficult to match deeper reductions with a specific melody.

To overcome this problem a different experimental approach is suggested by the present studies. Instead of presenting musical reductions, melodies were manipulated at individual notes in a same-different task. Analytical reductions of the presented melodies were done and within-harmony deviations were inserted at different time span reduction levels. It was hypothesized that deviations in the melody would be easier to recognize if they belong to deeper time span levels. If the memory representation of a tonal melody constitutes a hierarchical reduction then changes to the reduced levels would change the mental representation of it at a more fundamental level. For instance, tones at or near the beginning or end of a melody usually play an important structural role (like the beginning or end of a narrative), and a change of the melody at such positions would be perceived more readily than changes at other positions. We also speculated that an effect of different recognition rates at different time span levels would increase with longer sequences. Short sequences can be held within short-term memory that is known to represent material in a more direct manner than long-term memory. In a same-different task subjects may try to focus on individual pitches rather than the melodic whole, which will become increasingly difficult for longer melodies.

The same-different task has been used a variety of studies on music perception. In classic studies Deutsch *et al.* demonstrated that judgment about the identity of tones in a sequence is heavily affected by the harmonic or melodic nature of other interpolating tones in the sequence (Deutsch & Roll, 1974; Deutsch, 1982). If a tone is perceived as part of a larger gestalt (such as a melodic phrase) then changing the tone may be perceived as a change in the global pitch pattern. This is true of simple pitch patterns that are not part of a tonal melody - deviation to a pattern of pitches ascending stepwise on a diatonic scale is easily perceived. Changes to tones within a melody may also be perceived as a function of such low level auditory processing. However, within the context of a melodic gestalt some elements are more essential in defining the mental representation. An analogy may be drawn to visual perception: in perceiving another person, movements in many parts of the body are processed but even small movements of the face (defining a facial expression) are easily perceived

since they are often essential in defining the visual gestalt of the person (rather than e.g. movements of the shoulders at a similar physical magnitude). In detecting differences in a melody it may be expected that low and high level cognitive processes interact. However, if structural high-level features are strong enough they should outweigh the influence of low-level features in the experimental procedure.

II. METHOD

A. Stimuli

The two experiments used similar setups and differed primarily in the presented melodies. Four melodies were presented in total, two in each experiment (figure 1). In the first experiment two novel melodies were created (melody 1 and 2). Melody 1 was tonal and melody 2 was atonal. Both melodies were 8 seconds long and had identical rhythms.

In the second experiment a precomposed 'real' tonal melody (melody 3) was used (an extract from Isaac Albéniz' *España*). Before the beginning of the experiment subjects had to indicate whether they knew the melody in advance. Only subjects indicating that they did not know the pre-composed melody in advance were used for analysis.

A second ill-formed 'tonal' melody (melody 4) was constructed for experiment 2 by permutation of the order of tones from melody 3. The permutation was created by a random reordering of tones that resulted in correlations with the key profiles given by Krumhansl and Kessler (Krumhansl & Kessler, 1982; Krumhansl, 1990) nearly identical to that of melody 3. Consequently, melody 4 had an identical estimate of tonality by this correlational measure, but lacked the sense of meaningful construction of the precomposed melody 3. A tonal centre could clearly be perceived, but the melody lacked a sense of musical coherence. Melody 3 and 4 used in experiment 2 were both 18 seconds long and were also rhythmically identical.

Each melody could differ in one of four places. In melodies 1 and 3 (tonal) these positions corresponded to four different time span levels of the reductional analysis (figure 1). Time span reduction of the atonal and ill-formed melody (2 and 4) was not attempted, and these melodies deviated in similar positions as those in melodies 1 and 3 respectively. The positions of deviations in the tonal melodies were chosen so as to reduce the influence of surface salience due to low level auditory processing (repeated notes, parallelism etc.). In experiment 1 only one deviation was defined at each position, whereas the melodies in experiment 2 could deviate to two alternative tones for each position. This was done to ensure that heard deviations were perceived as a function of position in the melodic hierarchy, rather than as a change in melodic contour. The melodies in experiment 1 were all presented in the same key, whereas each consecutive presentation was transposed to a new key in experiment 2. The stimuli were generated using frequency modulated tones and presented to subjects in headphones in an isolated room.

B. Subjects

26 subjects participated in experiment 1 and 11 subjects participated in experiment 2. All participants had either some or much practical musical training. In the experiments, no significant differences in the detection rates were found between subjects as a function of level of training. Consequently, subjects are not differentiated according to levels of musical training in the analysis.

C. Procedure

In both experiments subjects were presented with either unchanged pairs or the original and deviation melody on consecutive trials. For all melodies each deviation were presented 4 times in total. After each pair, subjects had to judge whether the melodies were identical or different by pressing one of two buttons. Half of the presentations were identical pairs in experiment 1, and only 4 presentations were identical in experiment 2.

In experiment 1, subjects were presented with the two melodies in separate blocks, whereas the presentations in experiment 2 alternated between the two melodies within the block.

Figure 1. Presented melodies. Melodies 1 and 2 were presented in experiment 1; melodies 3 and 4 were presented in experiment 2. Numbers above deviating tones correspond to analysed time span level (for melodies 1 and 3) of the melodic position (level 1 is the musical surface, and level 4 is the deepest level of reduction)

III. RESULTS

A. Experiment 1

As indicated in figures 2 and 3 below, responses on different-trials are highly dependent on the deviant tone in both melodies. Two-way analysis of variance (deviations, subjects) showed that the difference in response to individual deviations were highly significant between subjects in both melodies (melody 1 (tonal): $F_{(4,26)}=38.42$, $p<.0001$, melody 2 (atonal): $F_{(4,26)}=9.1$, $p<.0001$), although less so for the atonal melody. As can be observed, the correct response rate is generally lower for the atonal melody, suggesting that deviations are easier to detect in the tonal melody.
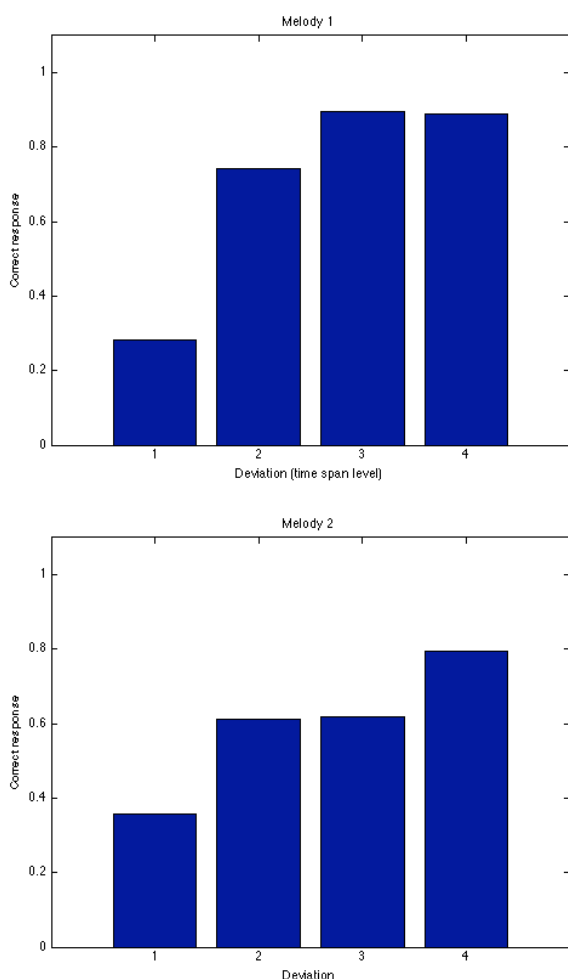
Figures 2-3. Mean correct responses for deviations in melodies 1, tonal (above) and 2, atonal (below) presented in experiment 1. Deviation numbers correspond to time span levels for the tonal melody (see figure 1 for details)

Since only responses to different-trials are analysed a measure of discriminability that considers responses to same pairs was obtained. To this end, d' indexes were calculated with a signal detection model assuming equal-variance Gaussian distribution of the perceived signal strength (McNicol, 1972). A d' mean of 1.6 for the tonal melody and 1.0 for the atonal indicate that subjects were sensitive to the task of detecting deviant tones (but less so for the atonal melody).

The mean correct responses for the tonal melody also indicate that deviant tones at deeper time span levels are better detected. The overall linear correlation between time span levels and correct responses was found to be strong ($r = .88$). Tones at levels 3 and 4 are roughly at the same level of detection though. This is most likely explained by the fact that many subjects were able to detect every deviation at these levels, and would imply the need to increase the difficulty of the task. In Experiment 2 longer melodies was used to achieve this.

Positions of deviance for the atonal melody were modified so as to observe the effect of changing the first and last tone, changing the peak note (deviation 3) and changing the direction of melodic movement (deviation 2), as in melody 1. The results indicate that surface saliences are perceived, but to a perhaps noticeable low degree. Change to a marked peak tone (deviation 3) or a dramatic change of melodic direction (deviation 2) was only perceived in a little more than half of the presentations. Changes to the endnote, however, are perceived almost at the level of the tonal melody. This may indicate that the final note is structurally important to the perception of any tone sequence, irrespective of the hierarchical structure of the melodic sequence.

B. Experiment 2

Experiment 1 showed a strong effect of time span level on detection rate in the tonal melody, but an effect of position could also be observed in the atonal melody. To further assess whether the results in fact indicate an effect due to hierarchical levels in the mental representation, the same method was used on a different set of melodies. Melodies used in experiment 2 were longer in order to reduce the possible influence of short-term memory.

Subjects were presented with either a precomposed tonal melody (melody 3) or and an ill-formed tonal melody (melody 4) formed by reordering the tones of melody 3. Again, the detection rate of the individual melodic deviations differed greatly in both melodies. Analysis of variance (deviations, subjects) again showed the difference in responses to different deviations to be significant (melody 3: $F_{(4,10)}=15.52$, $p<.0001$; melody 4: $F_{(4,10)}=8.89$, $p<.0002$), but less so for the ill-formed melody.

As indicated in figures 4 and 5 there is, again, an apparent correlation between the stipulated time span levels and response rates in the precomposed melody (melody 3). Surprisingly, correct response rate for deviation 4 (the end note of the melodic phrase) was significantly ($p<.003$) lower than for deviation 3 (an intermediate cadential note). This suggests that even though there is an overall correlation between analytical time span levels and detection of melodic differences ($r = .83$), there may be considerable variation within this correlation.
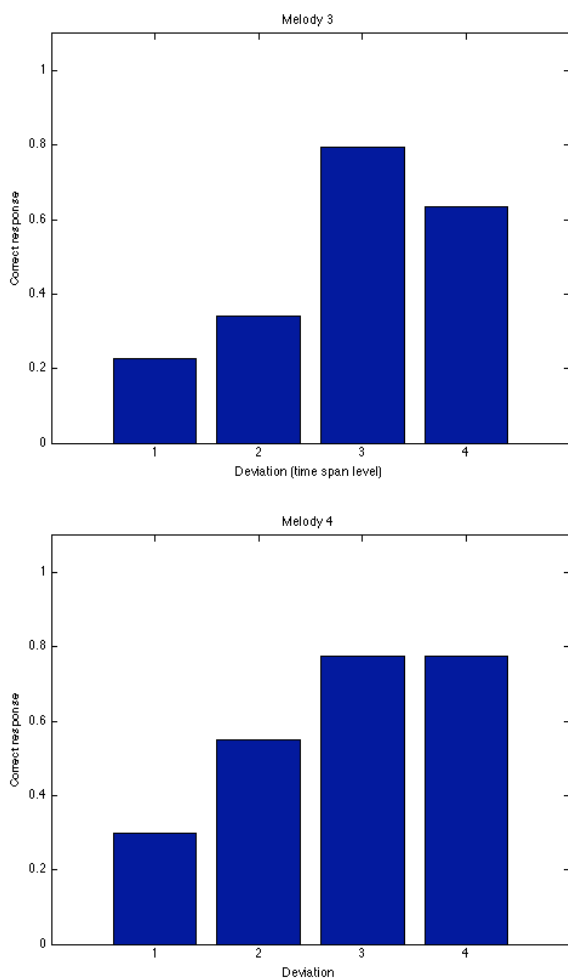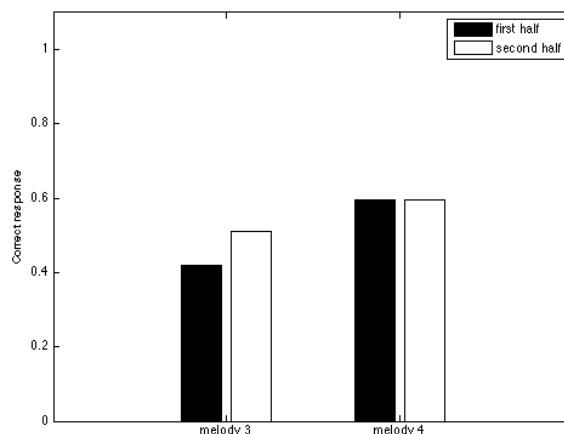
Figures 4-5. Mean correct responses for deviations in melodies 3, precomposed tonal (above) and 2, ill-formed tonal (below) presented in experiment 2

Also surprisingly, there was a relatively high detection rate and a strong linear correlation between deviations and responses for the ill-formed melody as well ($r = .97$). This indicates that even though there is a low degree of melodic coherence (which may be thought to provide the hierarchical structure) subjects attend selectively to different tones of the melody in a similar way as for a well-formed melody. This could mean either that the tones are perceived by some surface salience factor without being hierarchically linked, or that some sense of tonal hierarchy is still present even in the absence of a strict hierarchical structure that can be observed analytically. It is conceivable that the melody activates a tonal schema that implies a hierarchical organization. That is, the tonal character of the melody makes it possible to abstract some degree of hierarchical structure, even though the melodic whole is not a well-defined hierarchy of tones. For instance, deviances 3 and 4 may still be perceived as intermediate endings of a melodic sequence beginning with the first note without the melodic notes in between being hierarchically differentiated.

Also, the possibility that there is an interaction between the two melodies of experiment 2, presented in alternation, must be taken into account. Some subjects reported that they gradually formed an idea of the melodies across presentations

and perceived differences as deviations from this learned standard rather than within the individual pair. This could imply that subjects gradually become aware of deviance positions, and direct their attention selectively to these positions. To examine this we compared responses to the first half of the presentations to the second half. As can be observed in figure 6 there is a small increase in correct responses on the second half of the presentations compared to the first half, but this was in fact greater for the well-formed melody (melody 3) than for the ill-formed one (melody 4). If learning the well-formed melody enhances detection of deviant tones in the ill-formed melody then the reverse picture would be expected.

Figure 6. Comparisons of mean correct responses to all deviations between the first and last half of the presentations

The melodies in experiment 2 could deviate to one of two notes. As can be seen in figure 7 the responses for melody 3 are at similar levels for the higher and lower deviation at each position, suggesting that detection of melodic change did in fact depend on the structural position of the tone, rather than on change of the exact interval. Responses to the two deviations were less similar for the ill-formed melody (figure 8). This may suggest that perception of the ill-formed melody rely more on specific melodic intervals than the well-formed melody 3, and that specific salient intervals are responsible for part of the variation in the responses.
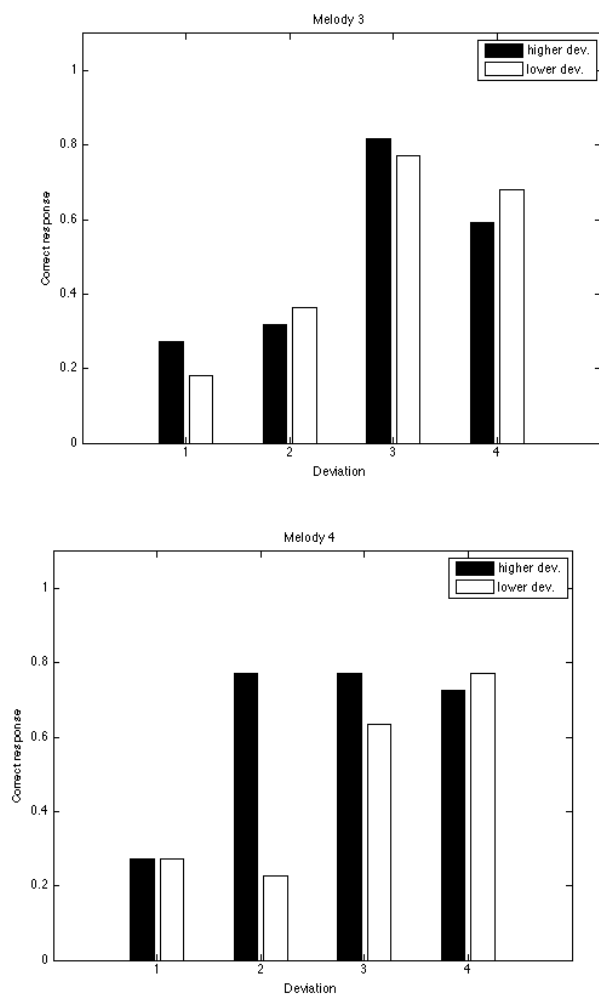
Figures 7-8. Mean correct responses to each of the two alternative deviations (higher/lower) at each position in melody 3 (above) and melody 4 (below)

It is likely to suspect that deviant melodic tones at strong metrical beats will be noticed more easily, and that metrical weight are responsible for variations in detection rate. To investigate this, the linear correlation between response rate and position in the metrical hierarchy (as described in GTTM) was calculated. Since structurally important tones often occur at strong beats there is generally a correlation between time span level and metrical position. This was the case in experiment 1, but was avoided for the melodies in experiment 2 (the deviation at time span level 2 occur at a strong beat). Correlation between response rate and metrical level was found to be low for the precomposed melody ($r = .19$), much weaker than the correlation with time span levels ($r = .83$).

The increased length of the melodies, compared to those of experiment 1, did lower the correct response rate, by which the (surprising) differences in sensibility to structurally important notes (time span level 3 and 4) become observable. However, it remains unclear whether some iconic memory imprint, even for these relatively long sequences, causes the relatively high detection rates for the ill-formed melody. It is conceivable that even longer sequences would lower the detection rate for ill-formed more than for well-formed

melodies, since long term memory could be expected to rely on hierarchical organization.

IV. DISCUSSION

Since the perception of melody depends on a range of perceptual and cognitive processes, the detection of tone differences within a melody must be thought to depend on many of these factors. The observed detection rates reflect a sum of interacting parameters, including surface salience, metrical organization, and structural levels. This makes the current methodology somewhat sensible and care must be taken to dissociate the structural levels from other factors. The strong correlations found between time span level and detection rate found in different tonal melodies in this study, suggest that the structural levels do in fact influence detection of melodic differences to a degree that is empirically observable.

Measures of linear correlation give a general indication of a relationship between hierarchical levels of musical perception and detection of difference, but a general assumption of linearity may turn out to be questionable. In melody 3, deviations at reductional level 3 were recognized at a higher rate than deviations at level 4. Analytically, the melodic tone ending a phrase (as deviation 4) clearly defines the tonal structure of the phrase at a more fundamental level than an intermediate resting note (as deviation 3).

More importantly, it is not necessarily clear in what way analytically derived levels of structure relate to cognitive levels of representation. Although analytical levels reflect some cognitive reality at a general level, as shown by the linear correlations, music analysis evokes an abstract formalism. If detection of melodic deviance is a valid indicator of structural levels, then the current study demonstrate that the perceptual hierarchy are subject to variations that are not fathomed by the formal architecture of music analysis. At a given level of analytical reduction, tones of the melodic structure are reduced or elaborated to arrive at deeper or higher levels, but there is no notion of different distances between these levels. Music analysis is most often analysis of well-formed musical works, and the relationship between levels of structure can be defined. In perceiving music, perhaps only heard partially or perhaps less well-formed, an abstraction of structural relations at different levels may still take place, but to a lesser degree. The present study indicates that mental hierarchies are subject to considerable variation depending on the nature of the musical material. Whereas analytical levels of musical structure are present or not, an empirical measure, such as the present, suggest that mental hierarchies are graded. Differences in the ability to memorize different melodies may reflect differences in the depth of their hierarchical organization. Melodic hierarchies stipulated in music theory are not ultimately falsified or verified by their correlation with detection of differences. Rather, the experimental approach demonstrates variation in the cognitive reality of these hierarchies, and serves as a fruitful supplement to the analytical one.

ACKNOWLEDGMENT

REFERENCES

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

Bharucha, J.J. & Krumhansl, C.L. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition*, 13, 63-102

Castellano, M.A., Bharucha, J.J. & Krumhansl, C.L. (1984). Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General*, 113, 394-412

Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl & Jackendoff's Grouping Preference Rules. *Music Perception*, 4, 325-359

Deutsch, D. & Roll, P.L. (1974). Error patterns in delayed pitch comparison as a function of relational context. *Journal of Experimental Psychology*, 103, 1027-1034

Deutsch, D. (1982). The influence of melodic context on pitch recognition judgment. *Perception & Psychophysics*, 31, 407-410

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

Krumhansl, C.L. (1990). *Cognitive foundations of musical pitch*. Oxford: Oxford University Press

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

Large, E., Palmer, C. & Pollack, J.B. (1995). Reduced memory representations for music. *Cognitive Science*, 19, 53-96

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

McNicol, D. (1972). *A primer of signal detection theory*. Sydney: George Allen & Unwin Ltd.

Palmer, C. & Krumhansl, C.L. (1987). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 116-126

Serafine, M.L., Glassman, N. & Overbeeke, C. (1989). The cognitive reality of hierarchic structure in music. *Music Perception*, 6, 397-429