Musical Change Deafness:
The Inability to Detect Change in a Non-speech Auditory Domain

Kat R. Agres (kra9@cornell.edu)
Department of Psychology, Cornell University, Uris Hall
Ithaca, NY 14853 USA

Carol L. Krumhansl (elk4@cornell.edu)
Department of Psychology, Cornell University, Uris Hall
Ithaca, NY 14853 USA

Abstract

This article presents two experiments investigating the degree to which listeners can detect changes in melodies. In both studies, pairs of melodies were presented to a group of professional musicians and a group of non-musicians. In Experiment 1, musical structure and musical expertise were explored with stylistic, non-stylistic, and random melodies. Experiment 2 utilized a full-factorial design to examine tonality, musical interval, metrical position, note duration, and musical expertise. Significant effects were found for several variables, but tonality had a particularly large effect on performance. Under some conditions, large changes between the melodies went undetected even by professional musicians. These results suggest that listeners form a memory representation for schematically consistent tones, which we refer to as the "musical gist". These results also suggest a comparison with change blindness, in which viewers can fail to notice salient changes in a visual scene, raising the question of whether similar processing operates in both modalities.

Keywords: change deafness; change detection; auditory gist; musical memory

Introduction

In this article, we present two experiments demonstrating that large changes in music can go undetected. Intuition suggests that our auditory perception of the world is not always robust and complete. Picture a scenario in which you are so involved in a conversation that it is not until several minutes have passed that you notice music playing in the background. It is not difficult, either, to imagine being oblivious to auditory changes in the environment, such as the sound of cars braking at a traffic light. It is likely that we often selectively process and remember what might be called the 'auditory gist' of a scene (see Harding, Cooke, & Konig, 2007).

There is evidence that listeners also form a gist when listening to music. Although some musical features seem to be encoded veridically, like the particular quality (or timbre) of a singer’s voice, other musical features, such as pitch and rhythm, are not always remembered in detail (e.g., Levitin, 2002). A professional musician, for example, is not always able to reproduce a novel melody immediately after hearing it. Instead, musical information is often encoded schematically (Snyder, 2000), with only general characteristics and certain salient features remembered.

One result of an efficient, schematic memory of music is that certain types of changes can go unnoticed. This draws a parallel to work done in vision concerning the inability to detect a change within a visual scene, called change blindness (for a review, see Simons & Rensink, 2005). Research in this area has demonstrated that changes to more salient aspects of a scene (such as those that may comprise a gist) are detected more frequently than peripheral changes (Rensink, 2002).

Though the change blindness paradigm has recently been extended to the tactile modality (Gallace, Tan, & Spence, 2006; Gallace, Auvray, Tan, & Spence, 2006), surprisingly little research has investigated change deafness, the auditory analog of this phenomenon, despite clear evidence that memory for auditory details is fallible (Vitevitch, 2003; Eramudugolla, Irvine, McAnally, Martin, & Mattingley, 2005). Because change blindness is a robust phenomenon that addresses perception, memory, and attention, the present research sought to examine whether similar processing occurs within the non-speech auditory modality.

Of the few studies examining change deafness in the auditory domain, none have systematically investigated music. Vitevitch (2003) demonstrated that listeners display significant levels of change deafness for detecting a different speaker during a lexical shadowing task. Speech can be useful for studying the mechanisms of change deafness, but instrumental music has the advantage that it is free of explicit semantic content. As a consequence, studying music has the potential to uncover basic auditory mechanisms without the influence of external reference. Thus, music is an ideal stage for testing whether the mechanisms underlying change blindness extend more broadly to auditory perception.

When an acoustic event is actively given attention, change deafness should be less frequent than when the sound is in the perceptual background. In instrumental music, certain notes comprise the foundation of the melody, while others are ‘passing tones’, generally of less emphasis and importance to the musical phrase. The following studies provide evidence that foundational and emphasized elements of music are more likely to be stored in short-term
memory, while less important elements are not encoded in the gist.

In addition to the influence of sensory information, in which one processes the low-level, psychophysical features of a stimulus, high-level top-down processing also plays a role in change detection. For example, domain-specific expertise seems to facilitate the detection of change within the learned domain. In a study reminiscent of Chase and Simon’s (1973) seminal chess expertise study, Werner and Thies (2000) showed that a group of experts in American football were more successful at detecting changes in football scenes than a non-expert comparison group. We extend this notion by testing whether a group of trained musicians will be more likely to detect changes in short, novel melodies than non-musicians.

Across modalities and paradigms, a number of studies have shown that a primary goal of perception is to rapidly understand our environment. Conventional approaches in perceptual research typically assume hierarchical processing of visual and auditory scenes, with processing of local details occurring before global characteristics. Current research, however, suggests the opposite may occur, as purported by Harding, Cooke, and Konig (2007; see also Hochstein & Ahissar, 2002; Navon, 1977; but see Friedman, 1979, for the contrary viewpoint).

In music, global context plays a significant role, especially the overall tonality or key (e.g., Krumhansl, 1990). Memory for tones and chords depends on their roles in the key, that is, whether they are or are not structurally significant in the key. Thus, we expect that the musical key is a salient characteristic of the gist of a melody, and that changes to structurally significant tones should be detected more frequently than changes to less significant tones. Additionally, if the tonal structure itself is ambiguous, schematic processing will be more difficult, and less detail should be encoded. Therefore, we also predict that the less musical structure, the more change deafness.

**Experiment 1: Musical Structure**

In the following study, listeners heard a short (two-measure long) melody followed by another short melody that may or may not have contained a changed tone. On trials containing a changed tone, the smallest interval of change was one semitone, which is at least seven to ten times larger than the threshold for hearing differences in pitch (Wier, Jesteadt, & Green, 1977; Tervaniemi, Just, Koelsch, Widmann, & Schröger, 2004). Thus, the tones in isolation would never be confused with one another. This study was directed at understanding the properties of melodic contexts that prevent listeners from hearing these relatively large pitch changes.

The effects of two factors were explored on the ability to detect changes of a tone within a melody. One factor was musical structure, in which some melodies were stylistic and conformed to musical conventions, some melodies were non-stylistic, and others were generated randomly. The second factor was musical expertise; the performance of non-musicians was compared to professional musicians.

**Method**

**Participants**

A group of 15 Cornell undergraduates volunteered to participate in the experiment for extra credit in a psychology course. This Non-musician group had little musical training (average training = 2.9 years, std = 3.1 yrs). Most of those participants who had once received musical training had not played an instrument in over 9 years.

A group of 11 Professional Musicians from the Indianapolis Symphony Orchestra were recruited and paid $20 to participate in the experiment. They comprised the Professional musician group, and had an average training of 44.9 years (std = 8.6 yrs).

**Materials**

Seventy-two stimuli were composed for this experiment. Each stimulus trial consisted of a 4-sec long melody, 500 ms of white noise, and then another 4-sec long melody that was either an exact repetition of the melody or an altered version of the melody with one tone changed. The change was in pitch, and varied from one semitone (a minor 2nd) to seven semitones (a Perfect 5th); rhythm was never altered. The changed tone could occur anywhere within the two measures, but always preserved the contour of the original melody (the pattern of rising and falling in pitch). A brief pause followed the stimulus to allow for a response. Trials with two identical melodies comprised the ‘Same’ stimuli, and those with a changed tone were the ‘Change’ stimuli.

All of the melodies were two measures long (in 4/4 time), and varied in rhythm to avoid monotony. To ensure consistency across stimuli, two quarter notes (long notes) and four eighth notes (short notes) were included per measure. Each melody was in the musical key of C, G, D, or F Major, and fell into one of three categories: Stylistic, Non-Stylistic, and Random. Stylistic melodies followed all of the normal constraints of Western classical music. Non-stylistic melodies sounded awkward, often featuring strange melodic jumps or unusual tonal progressions. The pitches of the Random melodies were determined using a random number generator where 1 equaled the lowest note in the two-octave scale being used, 2 equaled the second note in the scale, etc. The random melodies did, however, adhere to the rhythmic constraints of the Stylistic and Non-stylistic melodies (of two long notes and four short notes per measure).

The melodies were created in Digital Performer 4.5, saved as MIDI (Musical Instrument Digital Interface) files, and then converted into .wav files using a MIDI to .wav converter. Five hundred ms of white noise was then inserted between the two melodies of each stimulus using Cool Edit 2000. The total stimulus, including the initial melody, white noise, and comparison melody, was 8.5 seconds in length.
**Procedure**

Participants wore Bose noise canceling headphones and listened to the stimuli at a comfortable listening level (held constant across trials and participants). A brief practice session preceded the actual experiment. The stimuli were presented using E-Prime, which randomized the trials and collected participant responses.

Each participant heard the 72 trials in random order. On each trial, the participant’s task was to determine whether the two melodies were the same or different. Responses were made on a 6-point Likert scale so that the participants could express how sure they were of their response. The ‘1’ button was designated ‘absolutely sure same’; the ‘6’ button was designated ‘absolutely sure different’. Participants were encouraged to use the full range of the scale as they felt appropriate.

**Results and Discussion**

A mixed design 3 X 2 X 2 (Melody Type X Change X Musical Expertise) ANOVA was performed on the data, with Musical Expertise as the between subjects variable, and Melody Type and Change as the within-subjects variables. There was a significant effect of Melody Type, $F(2,48) = 17.78, p < .001$, with both Professional Musicians and Non-Musicians performing better on Stylistic than Non-Stylistic trials. Though Professional Musicians were more adept than Non-Musicians, both groups were able to perform the task, showing a significantly different response for Change than Same stimuli, $F(1,24) = 76.30, p < .001$.

Mean ratings were analyzed to determine the participants’ performance; better performance consisted of ratings closer to ‘6’ on Change trials and ratings closer to ‘1’ on Same trials. As hypothesized, due to their extensive training and performance of classical music, Professional Musicians consistently outperformed Non-Musicians, $F(1,24) = 4.85, p < .05$, with the exception of the Random Same trials, for which they were outperformed by Non-Musicians. The Professional Musicians’ poor performance on Random Same trials appeared to stem from a strong bias to report that these trials were changed. Consequently, the results were examined in terms of signal detection theory.

Figure 1 shows the results of the signal detection analysis with the criterion, $c$, values plotted in the graph above and the detectability, $d'$, values plotted in the graph below. Figure 1a shows a strong criterion shift for Random melodies for the Professional Musicians, so that they had a bias to judge Random Same melodies as different. In addition to the main effect of Melody Type ($F(2,48) = 11.17, p < .001$), there was a significant interaction between Melody Type and Musical Expertise ($F(2,48) = 5.29, p < .01$) reflecting this criterion shift.

Once this criterion shift is taken into account, Figure 1b, which shows $d'$ (the ability to distinguish between Same and Change trials), indicates that Professional Musicians were better than Non-Musicians in discriminating between Same and Change trials for all Melody Types ($F(1,24) = 6.99, p = .014$). There was a main effect of Melody Type ($F(2,48) = 4.64, p = .014$) with discriminability diminishing with decreased musical structure. Melody Type did not interact significantly with Musical Expertise.

The results confirmed that tonal structure has a large effect on listeners' ability to detect relatively large changes in melodies. All participants found the task more difficult when the musical structure was less conventional. This verifies the hypothesis that tonality is a strong factor in the global processing of the melodies, so that when the tonality was made unclear in the less well-structured melodies, performance on the task was impaired. Thus, it appears as though prior knowledge about musical style facilitates memory of the melody when it is conventional and hinders it when it is unconventional. Though these global characteristics of the music provide insight into how music remains in memory, the specific types of changes that elicit change deafness must be explored.
Experiment 2: Tonality and Rhythm

To gain a fuller understanding of what melodic properties affect memory for melodies, different musical characteristics must be thoroughly tested empirically. If the more salient musical elements are more likely to be encoded in memory, in a kind of musical 'gist', then these should be remembered more accurately (and changes to these elements should be detected more reliably).

To explore the specific musical properties that influence how musical elements are encoded in memory, this experiment was a complete factorial design with four factors: Tonality, the Interval of pitch change, the Position of pitch change, and Rhythm. Also, because musical expertise plays a role in how efficiently and effectively music is encoded in memory, two groups were tested: Professional Musicians and Non-Musicians. Timbre (the type of instrument playing) and dynamics (the loudness of the music) might also affect musical memory but these factors are not manipulated in this experiment.

To assess the role of Tonality in change detection, trials with scale and non-scale tones were used. On Change trials, a scale tone could be changed to a different scale tone or a non-scale tone, and a non-scale tone could be changed to a scale tone. We predicted that a non-scale tone in the first melody was not likely to be encoded in the gist and the change to a scale tone would be difficult to detect. In contrast, a change from a scale tone to a non-scale tone should be easy to detect, given the violation of the overall tonality in the second melody. Finally, changes from one scale tone to another may be very difficult to detect if the gist strongly encodes scale membership and less strongly encodes the particular tones in the melodies.

The Interval of pitch change was systematically varied in this experiment and ranged from one to four semitones. If the gist of a melody encodes tones, not only in terms of how musical elements are encoded in memory, in a kind of musical 'gist', then these should be remembered more accurately (and changes to these elements should be detected more reliably).

Rhythm and Position were manipulated to test whether metrical emphasis plays a role in detecting changes. This experiment used two different rhythms, as follows, Rhythm 1: \( \text{\#\#\#\#\#} \), and Rhythm 2: \( \text{\#\#\#\#\#} \). Either the fourth, fifth, or sixth tone, called Position 1, 2, and 3, respectively, could be changed within these two rhythms.

Noting the indication of the relative stress of the different metrical positions (e.g. Lerdahl & Jackendoff, 1983), we predicted that metrically stressed tones draw attention and are more likely to be encoded in a gist, so that changes to the stressed positions of the measure would be easier to detect. In particular, Position 1 of Rhythm 1 (the third beat of measure 1) should be particularly strongly encoded, as this position has both the metrical emphasis of being on a “strong beat”, and has a long note-duration (a quarter note).

Method

Participants
A group of 20 Cornell undergraduates volunteered to participate in the experiment for extra credit in a psychology course. This non-musician group had little musical training (average training = 1.6 years, std = 1.9 yrs). In addition, a group of 16 Professional Musicians from the Indianapolis Symphony Orchestra were recruited and paid $20 to participate in the experiment. They comprised the Professional musician group, and had an average musical training / musical performance career of 43.9 years (std = 7.4 yrs).

Materials
Because the second experiment had a full-factorial design of the 4 within-subjects variables (Rhythm, Interval, Position, and Tonality), 288 stimuli were composed for the study. As in Experiment 1, each stimulus contained two melodies, which again were two measures long, and consisted of two quarter note (long) tones and four eighth note (short) tones per measure. The length of the melodies and white noise in between the melodies was the same as in Experiment 1. Stimuli were made from one of two Rhythms, as shown before: quarter-eighth-eighth (long-short-short) and eighth-eighth-quarter (short-short-long). Position refers to the serial position of the tone that was changed within the melody. The changed tone was always one of the last three positions in the first measure of the melody. Interval refers to the interval of change; the changed tone shifted up or down a m2, M2, m3, or M3 (one to four semitones). To avoid confounding melodic contour, this change always maintained the contour of the melody.

All of the melodies were composed in the musical key of C major. In Change stimuli, there were three types of changes from the first melody to the second: a scale tone was changed to a different scale tone, a scale tone was changed to a non-scale tone, or a non-scale tone was changed to a scale tone. Thus, there were three types of tonality conditions, Scale-Scale, Scale-NonScale, and NonScale-Scale. The NonScale-Scale trials were obtained by reversing the order of presentation of the two melodies in the Scale-NonScale trials. For example, one trial would present melody A and then melody B, and the reversed trial would present melody B and then melody A.

In addition, two stimuli were composed for each combination of the Rhythm/Position/Interval/Tonality factors, and there were twice as many Change as Same stimuli. In all, this totaled 96 Changed stimuli in both presentation orders, and 96 Same stimuli.

Procedure
Participants wore Bose noise-canceling headphones and completed a short practice session. Using E-Prime software, all of the stimuli were presented in random order to each participant, in three 15-17 minute blocks (which were counterbalanced across subjects). Participants were asked to judge whether the two melodies on each trial were either the
same or different. They responded on a 6-point Likert scale as in Experiment 1.

Results and Discussion
A mixed five-factor design 3 (Tonality) X 2 (Rhythm) X 3 (Position) X 4 (Interval) X 2 (Musical Expertise) ANOVA was performed to assess the impact of musical training on the detection of change for varying levels of rhythm and metrical emphasis. Tonality, Rhythm, Position, and Interval were within-subject factors, and Musical Expertise was a between-subjects factor. Because the specific role of different variables on change deafness was of primary interest, only results for Change trials are reported here.

Tonality had a very large effect on the ability to detect changes, $F(2, 68) = 459.8, p < .0001$, with changes from a scale to a non-scale tone easiest to detect. The role of musical training on the perception of Tonality (shown in Figure 2) was also of interest. Musical Expertise was found to be highly significant, $F(1, 34) = 25.1, p < .0001$, with Professional Musicians outperforming Non-Musicians. Tonality significantly interacted with Musical Expertise, $F(2, 68) = 69.5, p < .0001$, which was expected due to the extensive musical training and experience with tonality that Professional Musicians acquire (see Figure 2).

![Figure 2: Mean ratings (where 6 means “sure of change”) for Professional and Non-Musicians across levels of Tonality.](image)

The interaction between Tonality and Musical Expertise stems from the relatively low mean for Professional Musicians in the condition in which a scale tone is changed to another scale tone. In fact, they performed just slightly above the chance level of 3.5, the mid-point of the response range. Recall that some of these changes are as large as 4 semitones. This striking finding provides additional evidence for the idea that tonality affects what kind of information is encoded in the remembered gist. Apparently, even highly trained musicians encode tones largely in terms of whether or not they belong to the scale, so when another scale tone is substituted, it goes undetected (as long as the contour of the melody is unchanged).

The Interval of pitch change also played a significant role in change detection, $F(1, 136) = 5.2, p < .01$. As predicted, performance depended on the category of the change in terms of seconds and thirds. Larger intervals of change (Minor and Major thirds) were detected more frequently than smaller intervals of change (Minor and Major Seconds). A linear contrast comparing the means of 2nds to 3rds was highly significant, $F(1, 136) = 11.1, p < .001$. Interestingly, the Interval that elicited the most change deafness was that of a Major 2nd, perhaps due to its frequency in Western classical music (as there are five Major 2nds in a musical scale, and only two minor 2nds, for example).

In addition, there was also a large effect of metrical and durational emphasis. As expected, there was no main effect of Rhythm on change detection, $F(1, 34) = 0.32, p = .57$, but there was a highly significant interaction between Rhythm and Position, $F(2, 68) = 11.88, p < .0001$ which was driven by the good performance for Position 1 of Rhythm 1. In Rhythm 1, the first Position was always a long note, while Positions 2 and 3 were short notes. In Rhythm 2, Positions 1 and 2 were short notes, and though Position 3 is a long note, it does not occur on a strong beat. A linear contrast yielded significantly better performance (a higher mean response) for Position 1 of Rhythm 1 as compared to all other positions, $F(1, 168) = 19.1, p < .0001$. Thus, the combination of metrical and durational emphasis appears to make it more likely that a changed tone in that position would be noticed more easily.

General Discussion
A large number of interacting parameters can play a role in musical change detection. People do not encode detailed information about all of the characteristics of music; rather, they form a gist of the salient properties of music. Tonality is one of the most fundamental properties of a piece of music, and is therefore encoded into a gist. When a gist is formed of an initial melody that falls within a certain tonal category, a comparison melody containing a non-scale tone is very obvious to listeners. Rhythm and metrical structure can also give emphasis to a group of notes or a passage of music that are then encoded in the gist. When a lack of musical structure or style is present, listeners are worse at encoding features of the music. A tonal and metrical structure seems to give listeners a template on which to build their gist. When the melodies presented in Experiment 1 were Random or Non-stylistic (lacking in tonal structure), both Professional Musicians and Non-Musicians could not reliably encode features of the music. The listeners’ experience and familiarity with Western music was essential for these tasks, and is what enabled better performance for Stylistic melodies.

The failure to detect change may be driven by similar perceptual processes across modalities. In vision, alternative theoretical explanations suggest that either the input is not richly represented, not retained in working memory, or is not explicitly compared with new input (Simons, 2000;
Simons & Ambinder, 2005; Rensink, O’Regan, & Clark, 1997). Further, it has been shown that objects consistent with a schematic representation are better retained (Brewer & Treyens, 1981).

The two studies presented here also found that schematically inconsistent elements were not encoded reliably. Non-scale tones, as well as tones not emphasized by meter and duration, were not consistently retained in short term memory. This suggests that the reason relatively large changes in the melodies go undetected is because some tones are not retained in working memory.

Research on change deafness has been critiqued as a poor analog of the visual phenomenon because the auditory stimuli used were not static, like the images often used in change blindness experiments (Eramudugolla et al., 2005; Demany, Trost, Serman, & Semal, 2008). However, there have been examples of dynamic stimuli in change blindness studies (Simons & Levin, 1997), such as movie clips in which an actor or set change goes unnoticed after a change in the camera angle. Therefore, we believe that melodic stimuli that take place in the temporal domain are analogous to change blindness for dynamic motion picture scenes. It should also be noted that detecting the pitch changes in Experiments 1 and 2 should be well within the realm of auditory short-term memory (Sams, Rif, & Knuutila, 1993): The time from the to-be changed note in the initial melody to the changed note in the comparison melody is less than five seconds.

Future goals of this work include using different populations of listeners, styles of music, instrumentation, and rhythms differing in metrical stability. In line with the recent interest to explore crossmodal change blindness (see Gallace, Auvray, Tan, & Spence, 2006), we also plan to examine the effects of combining visual change blindness with auditory change deafness, and test whether there is an interaction between the modalities in change detection. Another primary goal of future research is to further examine the musical features that are encoded in a musical gist.

Acknowledgements
We extend our thanks to Dr. James Booth and the Cornell Statistical Consulting Unit for statistical advice.

References
Friedman, A. (1979). Framing Pictures: The role of knowledge in automatized encoding and memory for gist.