Relative salience of chord-type and chord-voicing changes:
A two-oddball paradigm

Ivan Jimenez[1] Tuire Kuusi[1] & Juha Ojala[1]

[1] Sibelius Academy, University of the Arts Helsinki, Finland

Author Note

This paper is the *Final Draft* of the following article:
Jimenez, I., Kuusi, T., & Ojala, J. (2021). Relative salience of chord-type and chord-voicing changes: A two-oddball paradigm. *Psychology of Music*. DOI: 10.1177/03057356211055214

Correspondence concerning this article should be addressed to Ivan Jimenez, Ivan Jimenez, Sibelius Academy, University of the Arts Helsinki, Helsinki, P.O. Box 30, FI-00097 UNIARTS, Finland.

Email: ivan.jimenez.rodriguez@uniarts.fi

Abstract

Our research project investigated the effect of background and stimuli factors on the relative salience of chord-type and chord-voicing changes. Earlier studies have shown that surface features tend to be easier to perceive than deeper features and that musical training attenuates this general tendency. For further studying how deeper-level and surface-level musical features are perceived, we used a two-oddball paradigm. Each item consisted of a succession of five same-root chords: one chord-type oddball (deeper feature), one voicing oddball (surface feature), and three standards. Participants chose the chord that sounded most different to them. All chord-type pairings formed of major, minor, dominant seventh, major seventh, and minor seventh chords were tested. Chord-type oddball and voicing oddball were chosen equally often, together forming the majority of the responses. Musical training and conceptual knowledge of chords affected the chord-type oddball responses, but not the voicing-oddball responses. However, chord-type oddballs were chosen regardless of the musical training. Chord-type responses were easiest for pairs consisting of a major-based and a minor-based chord and for pairs involving two pitch-class changes. Our results suggest that musical training and conceptual knowledge about chords is not the only factor influencing the relative salience of chord- type changes over voicing changes.

Keywords: chord type, voicing, harmony, oddball paradigm, salience

Relative salience of chord-type and chord-voicing changes:
A two-oddball paradigm

In music listening, surface features such as dynamics or melodic contour tend to be easier to perceive than deeper features such as motivic or harmonic relationships. This tendency to perceive changes in surface over structure has been observed not only in studies using excerpts from real pieces of music with complex textures (e.g., Deliège, Mélen, Stammers, & Cross, 1996; Lamont & Dibben, 2001) but also in block-chord textures – the simplest and most common way harmony is instantiated in music theory textbooks, ear training activities, and cognitive research (e.g., Beal, 1985; Farbood, 2012). The diversity of contexts in which surface features have been found to be easier to perceive than structure suggests that this tendency might be generalizable to a wide variety of listening conditions (e.g., puzzle paradigm in Deliège et al., 1996; similarity ratings in Lamont & Dibben, 2001; same/different task in Beal, 1985; and matching musical stimuli to visual representations of tension in Farbood, 2012).

Musical training has sometimes been found to attenuate listeners' general tendency to focus on surface changes over harmonic changes (e.g., Deliège et al., 1996; Schubert & Stevens, 2006). This effect may be explained by the great emphasis that Western formal musical training places on the development of conceptual understanding and aural identification of tonal harmony (Snodgrass, 2016). This training usually involves specialized concepts, terms, and labels by which the chord type (that is, the harmonic identity of the chord, e.g., major chord, dominant seventh chord) or the progression type (e.g., I–iii–IV–vii^{o}6–I) can be described abstractly in isolation from other music elements. Everyday musical activities outside the classroom such as aural identification of chords, playing and improvising based on chord labels, as well as using the concepts in music analysis might still add to the effect. Together these activities strengthen the understanding of harmony which, in turn, is likely to intensify attention and sensitivity to harmony when listening to music.

Chord types and voicings

As stated, concepts, terms, and labels of harmony allow the description of harmonic elements in isolation from other music elements. Yet, chords and harmony cannot be played without pitch-related features (e.g., voicing in terms of inversion, spacing, and doublings; register; transposition; and texture); timbre (e.g., instrumentation); or dynamics and duration. Due to its complexity as a musical phenomenon and the general tendency of surface features to attract listeners' attention, harmony is often taught and studied by using as simple realisations as possible (e.g., block chords in middle register, same timbre, same duration, etc.; see, e.g., Christensen, 2006; Snodgrass, 2016). Investigating the complex connection between harmony and other musical features is difficult when real pieces of music are used (Deliège et al., 1996; Granot & Jacoby, 2011, 2012; Williams, 2005). Simpler stimuli, such as block chords, instantiate harmony in a relatively clear way by reducing the complexity of some surface features such as the rhythm and texture, while still allowing sufficient freedom and control for the manipulation of other musical features, such as chord voicings, voice-leading between chords, and melodic line formed of the highest pitch of each chord.

The oddball paradigm

One way to study the complex interaction between harmony and other musical features is to use the oddball paradigm. The oddball paradigm tests responses to deviant stimuli inserted in a succession of otherwise homogeneous or relatively homogeneous stimuli. Even though the oddball paradigm was originally created for studying brain responses (especially event-related potentials such as the mismatch negativity, MMN) for auditory stimuli, it has also been used outside the neuroscientific realm. For instance, Kuusi (2010) presented participants with successions of five different chords. Four of the chords shared their harmonic identity (set-class), while the fifth one deviated in it, but all chords, whose serial order was randomized, were different in terms of transposition and voicing. Successions were composed in a way that the harmonically deviant chord was "horizontally" competing for participants' attention against other chords having the closest voicing or the highest or lowest register. In this way, it was possible to test whether participants perceived the changes in harmonic identity regardless of the constant changes in pitch height or voicing.

During the past decade, the popularity of the oddball paradigm as a tool for studying chord perception has increased (e.g., Goldman, Jackson, & Sajda, 2020; Klein & Zatorre, 2011; Kuusi, 2010, 2015; Linnavalli, Ojala, Haveri, Putkinen, Kostilainen, Seppänen, & Tervaniemi, 2020; Putkinen, Tervaniemi, Saarikivi, Ojala, & Huotilainen, 2014; Sturm, Blankertz & Curio, 2017; Virtala, Huotilainen, Lilja, Ojala, & Tervaniemi, 2018). In order to minimize the effects of horizontal pitch patterns in this type of experiments, successive chords are often transposed in pseudo-randomized patterns that exhibit some of the following characteristics: (a) all 12 pitch classes of the chromatic scale are used as evenly as possible; (b) chords with shared tones are not played in immediate succession; (c) chords transposed to the same pitch level (e.g., same root) are not played in close succession or same roots are not used at all; (d) chord successions typical of Western tonal music are avoided. These conditions increase the likelihood that participants' performance reflects discrimination of chord types as opposed to mere detection of pitch changes (e.g., the change from E to Eb in the chord succession C–Cm) or tonally unexpected chord successions. Additionally, this type of pseudo-randomized transposition embeds chords in harmonic contexts that more closely resemble post-tonal than tonal musical styles and is therefore particularly well suited for testing post-tonal chord types (actually, set-classes; e.g., Kuusi, 2010). However, this type of transposition scheme provides a less naturalistic environment when the oddball paradigm is used to study the perception of chord types typical of tonal repertoires such as major and minor triads.

One of the potential problems with pseudo-randomized chromatic transposition is that some listeners are more accustomed than others to concentrating on the chord type and ignoring other musical features. Specifically, it is likely that formally trained musicians tend to find it easier to focus on the chord type regardless of the context since aural identification of tonal chords in randomized chromatic transpositions is extensively taught in formal musical training (e.g., Buonviri & Paney, 2020; Thomson & Blombach, 1988). This potential effect of training on listeners' ability to downplay horizontal relationships in chord perception is consistent with the research showing that, when chords are transposed chromatically in a pseudo-randomized way, formally trained musicians outperform non-trained participants in neural discrimination of tonal chords (e.g., Virtala, Huotilainen, Partanen, & Tervaniemi, 2014) but not of nontonal chords (e.g., Linnavalli et al., 2020).

Aim

The aim of the study was to investigate the relative salience of chord-type and chord-voicing changes. Chord type is a conceptually abstract feature, while chord voicing (hereafter, voicing) is a surface-level feature without which a chord of any type cannot be played nor heard. For studying the relative salience of the two features, we used a two-oddball paradigm.

We believe that a two-oddball paradigm may be better suited than previous oddball paradigms to study the relative salience of changes in two competing features simultaneously. This paradigm combines aspects of two different types of oddball experiments. First, as in early uses of the oddball paradigm to study neural chord discrimination, chords in immediate succession in our paradigm share their same root (e.g., Koelsch, Schröger, & Tervaniemi, 1999), providing a context that is more naturalistic than pseudo-random transpositions and in which deviant chords are easier to notice. Second, as it has become customary in oddball paradigms that study chord discrimination using neural or behavioural responses, each succession of chords includes more than one type of deviant chord (e.g., Virtala et al., 2014), or more than one acceptable response (e.g., Kuusi, 2010; 2015).

In this experiment, we used successions of five chords which included three “standard chords” that were identical to each other in terms of both pitch-class content and voicing, and two types of deviant chords. One deviant chord changed the voicing of the pitches (hereafter “voicing oddball”), indicating that the dispersion of chord tones across register differed from those of the other chords. The other deviant chord changed the pitch-class content, and, hence, the chord type (“chord-type oddball”), indicating a pitch-class change of one (e.g., C–Cm) or two notes (C–Cm7). An example of an item and more description of creating the items can be seen in Stimuli and Figure 1.

One of the limitations of earlier oddball paradigms that used fixed-root chord successions and only one type of chord change was that it was not possible to disentangle participants’ sensitivity to vertical intervallic structures from their sensitivity to pitch-class change (e.g., the change from E to Eb in the succession C–Cm in Tervaniemi, Sannemann, Nöyränen, Salonen, & Pihko, 2011). Including two different types of pitch changes, only one of which was theoretically understood to change the chord type, allowed us to assess the relative salience of changes of vertical harmonic identity in relation to other types of pitch changes. We also analyzed the effect the participants' conceptual understanding of chords as well as the type of chord-type changes on the responses.

Methods

Participants

Altogether 1096 participants visited the online experiment between June 4 and June 12, 2020. Of them, 715 were discarded based on the loudness pre-test and the headphone question (see Procedure, below). Another 122 participants decided to skip the test. The total number of participants who completed the experiment was 247. Of these 247 participants, we discarded 128 because they were likely to have completed some parts of the experiment without actually listening to the item or with the help of autofillers or bots (for more information about these participants, see Table 1; for more information of discarding participants, see Appendix Table 1). The total number of participants whose responses were included in our main analysis was 116 (78 male, 36 female, 2 other; age *M* = 36.86, *SD* = 10.64; for more information about the participants, see Table 2a and Table 2b). Although there were more male participants than female, there is no evidence to our knowledge of a gender effect on the perception of chord type or chord voicing.

Table 1. Description of completed attempts that did not pass our "seriousness" criteria.

Criteria	Cases	Percentage
1, 2, and 3	65	51%
1 and 2	31	24%
1	18	14%
1 and 3	4	3%
2 and 3	2	2%
2	4	3%
3	4	3%

Criterion 1: Respondents heard less than 5 chords in 50% or more of the 60 main trials. For more details about the extent to which participants could control the number of chords they heard please see Procedure.

Criterion 2: Responses too close to random distribution (chord-type oddballs in 10-30% of the trials, voicing oddballs in 10-30% of the trials, and standards in 50-70% of the trials). We considered this distribution to indicate non-serious participants because difficulty in identifying oddballs should have led to higher percentages of responses in which participants said "all chords sound identical to me" than responses in which participants choose a standard chord, one of three identical chords, as being the most different sounding chord.

Criterion 3: Respondents did not understand questions about musical background or provided careless or automatic responses (bots, autofillers, etc.). The most common type of response that we considered to be an indication of participants not fully understanding our question or providing a careless or automatic response were instances in which participants responded "Best," "GOOD," or "Yes" to the question "What musical instrument have you played best?"

We collected background information of the participants by a questionnaire which they filled in after completing the experiment, and it also included a chord-identification post-test. The information can be seen in Tables 2a and 2b and will be explained in Results.

Table 2. Participant background variables.

Table 2a. Participants' experience playing and practicing musical instruments.

Experience	Participants n	Participants %
5 years or more	31	26.7%
Less than 5 years	18	15.5%
Had never played and instrument	67	57.8%
Total	116	100%

Table 2b. Other information about participant background variables.

Variable name	**Explanation of the variable**	*M*	*SD*	Min.	Max.	% "never" or 0*
V1_inst_years	Years of playing main instrument	3.94	7.80	0	38	57.8%
V2_chord_ knowledge	Participants' self-report about their knowledge of major and minor chords**	3.80	1.54	1	6	39.7%
V3_aural_ chord_ID	Aural chord identification score in the post test	19.1%	21.4%	0%	90%	42.2%
V4_attention_ to_melody	Attention to melody when listening to music in everyday life	4.36	1.47	1	6	3.4%
V5_attention_ to_chords	Attention to chords when listening to music in everyday life	3.30	1.59	1	6	16.4%
V6_play_chords _by_ear	Total hours of having played chords by ear***	321.25	1375.47	0	10420	73.3%
V7_play_chords _from_notes	Total hours of having played chords from notation***	450.15	2041.61	0	15630	75.9%
V8_music_ theory_years	Years of studying music theory or analysis	0.44	1.72	0	12	87.9%
V9_aural_skills _years	Years of studying aural skills or ear training	0.16	0.74	0	5	92.2%
V10_chord-type_ID_years	Years of practicing chord-type identification by ear	0.23	1.14	0	10	90.5%

NOTE:
*Percentage of participants responding "never." In the case of V2, the value indicates the percentage of participants" who responded 5, 6, or 7 (see **). In the case of V3, the value indicates the percentage of participants who had zero correct responses in the aural chord ID test or who were not asked to take the test because their responded 4, 5, 6, or 7 to V2 (see**).
**Participants were asked to respond to the question "Can you identify major and minor chords just by listening to them?" by choosing one of the following options: (1) yes, (2) most of the time, (3) only sometimes, (4) no, (5) I know what those terms mean, but I have never tried to identify them by ear, (6) I have heard those terms before, but I do not know what they mean, (7) I have never heard those terms before.
***In order to obtain a more accurate estimate of total hours, we asked participants to estimate the approximate number of years and average hours per week.

Stimuli

Each item in our experiment constituted of a succession of five chords: three standards, one chord-type oddball, and one voicing oddball (see Figure 1).

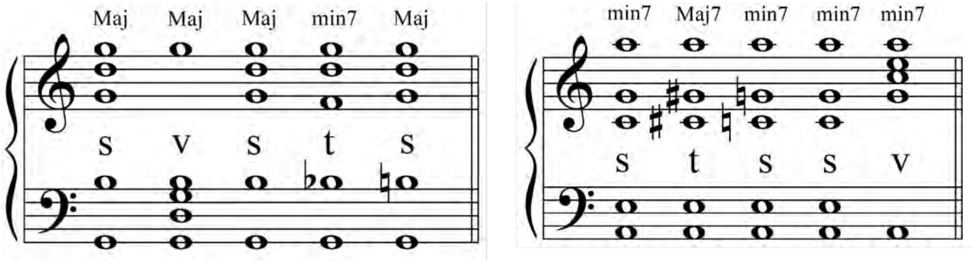


Figure 1. Examples of items used in the experiment. s = standard chords, t = chord-type oddball, and v = voicing oddball.

Voicings

All chords in our experiment were voiced using five pitches. For the highest and lowest pitch (outer voices), we always used the root of the chord, and these two pitches were always three octaves apart. We avoided pitch changes in the outer voices since they would have been too salient (as shown in pilot experiments). Furthermore, we avoided doubling any chord tone other than the root in order to prevent any idiosyncratic tonal effects related to chordal doublings (Huron, 1993). Finally, we avoided voicings containing harmonic intervals of a second to reduce the role of sensory dissonance in the task.

Each item in the experiment included two voicings. These two voicings corresponded to one of the three voicing pairs shown in Figure 2. These voicing pairs were characterized by two pitch changes. The results from our pilot experiments indicated that the voicing pairs shown in Figure 2 were less likely than other voicing pairs to have the pitch changes hidden by frequency masking.

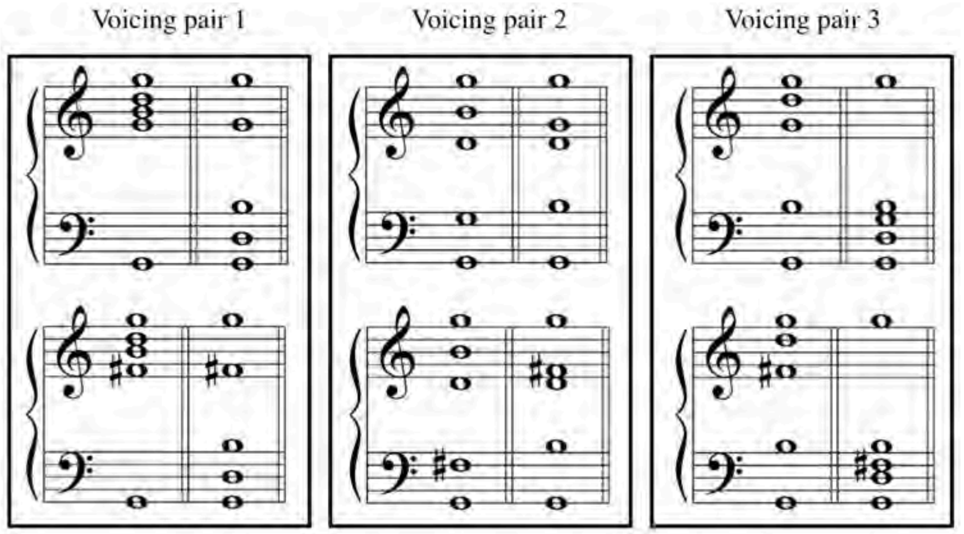


Figure 2. Voicing pairs used in the experiment. The upper and lower systems show examples of each voicing pair using triads and seventh chords respectively. Both voicings within each pair were used as both the standard and the voicing oddball when creating the items.

Although sometimes found in piano music, we did not include any voicings in which the third of the chord was three or four semitones above the bass because in the register we used for the bass notes (G2 to A2), these thirds fall within the critical band (Huron 2016).

Chord type

Each item in the experiment included two chord types. Items were created using all potential pairings of major, minor, dominant seventh, minor seventh, and major seventh chord types (see Table 3). Both orders (e.g., major standard, minor deviant and minor standard, major deviant) were used. In the context of our experiment in which roots did not change within items, these pairs could be divided in two categories according to whether the chords differed by one or two pitch classes.

Table 3. Chord-type pairs

One-pc change		**Two-pc change**	
maj	min	maj	min7
maj	dom7	min	maj7
maj	maj7	min	dom7
min	min7	min7	maj7
min7	dom7		
dom7	maj7		

Note: Both chord types within each pair were used as both the standard and the chord-type oddball when creating the items.

Roots and transpositions

In each item, all chords of the five-chord progression had the same root (see Figure 1). Even though music theory has traditionally paid more attention to chord successions in which the root changes, successions of chords with a fixed root also exist in Western tonal music (e.g., Doll, 2017; Scott, 2000). Although only some of the five-chord successions we used correspond to fixed-root chord successions found in tonal Western music (e.g., our stimuli C–C–C–C–C7 relates to common occurrences of V–V7 or I–V7/IV in real music), all our five-chord successions provide a more naturalistic harmonic context than pseudo-randomized chromatic transpositions because they imply one clear tonal centre or a movement between two closely related keys. Therefore, the fixed-root chord successions increase the ecological validity of the results. Further, since the chords of the fixed-root successions include many common pitches and common pitch-classes, it is also a much easier task for the participants to discriminate the changes in them than in the pseudo-randomized chromatic transpositions with constantly changing pitches and pitch-classes. Taken together, we believe that the experiment using fixed-root successions is a more suitable tool to study differences between non-musicians.

In order to reduce participants' fatigue and habituation, items were played on three different pitch levels: G, Ab, and A.

Timbre and loudness

Ninety individual chords (6 voicings, 5 chord types, 3 transpositions) were composed using sampled piano tones (Bösendorfer from Logic Pro X). In order to further downgrade the salience of voicing changes, notes in the outer voices (G2, Ab2, A2, G5, Ab5, A5) were intentionally made louder (MIDI key-velocity range = 61–77) than the inner voices (pitches D3–E5, MIDI key-velocity range = 39–46). We used a range of key-velocities (instead of a fixed value) for both outer and inner voices, because of the noticeable differences in subjective loudness across individual pitches in Logic Pro X Bösendorfer, despite their identical key-velocities. These differences were adjusted to the sets of key velocities used in the experiment, based on several pilot tests with similar types of participants and listening conditions.

Duration

The inter-onset-interval between chords was 1500 ms. In order to avoid any frequencies from acoustically lingering from one chord to the next, and in order to do so in a way that sounded natural, we took the following measures:

1) The delay, ambience, and reverb controls of the Bösendorfer Logic Pro X instrument were set to zero.
2) Each chord was created in Logic Pro X as an individual track.
3) The MIDI duration of each individual chord was set to 750 ms in order to allow for most of the acoustic resonance of the sampled piano tones to fade out in a natural way before 1500 ms have passed since the initial attack.
4) Each chord was exported from Logic Pro X as an individual audio WAV file.

5) A ramp was applied to the last 10 ms of each of the 1500-ms WAV files.

Serial position

The total number of items was sixty, consisting of each of the 10 chord-type pairs played in both orders (e.g., major standard, minor deviant and minor standard, major deviant) and each order played using the three voicing-pair types. The 60 items were composed making sure that:

1) Chord-type and voicing oddballs never followed each other in immediate succession within an item. That is, there was always at least one standard chord separating and serving as reference for both oddball chords.
2) Chord-type and voicing oddball were never separated by more than two standard chords. Pilots showed that the serial-position arrangements t-s-s-s-v and v-s-s-s-t (s = standard chords, t = chord-type oddball, and v = voicing oddball) created a very strong bias for participants to hear the last chord as being the most different sounding chord.
3) Chord-type and voicing oddball occurred equally often on each of the five serial positions.
4) The three items for each combination of chord-type pair and order differed in terms of the serial position of their chord-type and voicing oddballs.

Item repetition

For each trial, an audio file was created in which the item was repeated three times and items were separated by a 1500-ms silence.

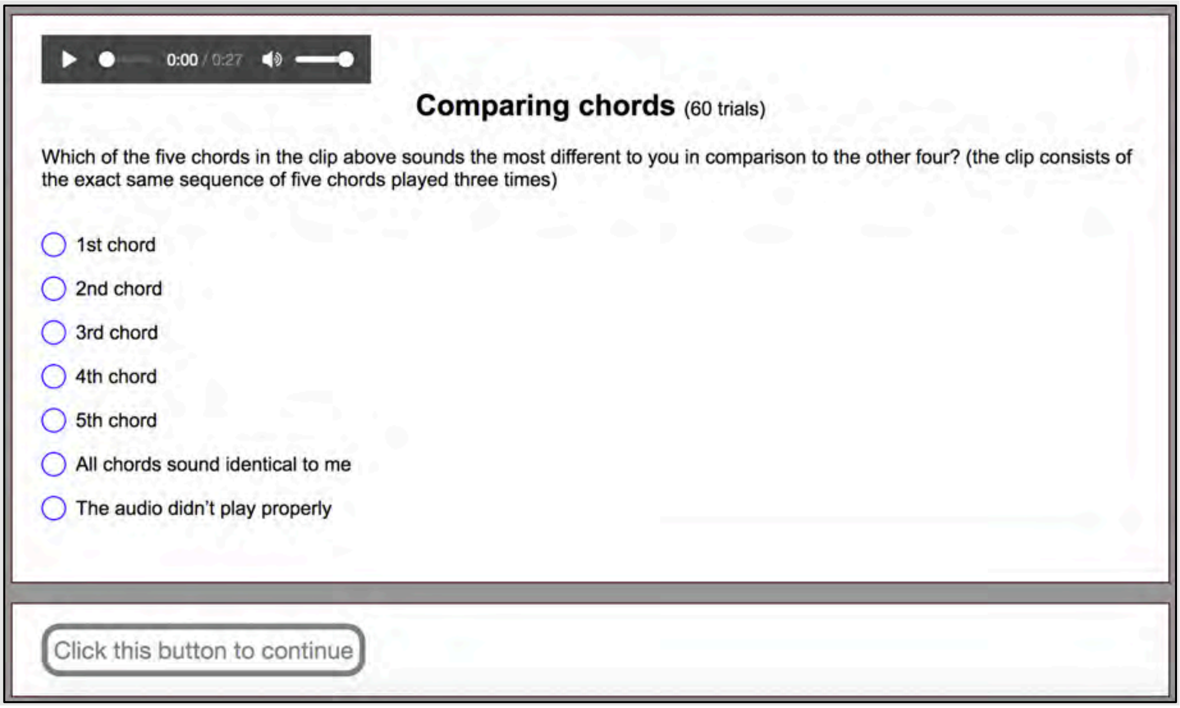
Procedure

Participants were recruited using Amazon Mechanical Turk (MTurk), a crowdsourcing platform that provides access to more than a hundred thousand potential participants (Difallah, Filatova, & Ipeirotis, 2018). Armitage and Eerola

(2020) have shown that the results of music cognition experiments on chord perception carried out in standard lab settings are comparable to those from online experiments that recruit participants using services similar to MTurk. However, study of crowdsourcing platforms has shown that the percentage of non-serious participants and survey-bots is relatively large (Ahler, Roush, & Sood, 2019; Dennis, Goodson, & Pearson, 2020). Hence, a pre-test and other ways to detect non-serious participants and survey-bots were included in our experiment (see Participants and below).

For online data collection, the software PsyToolkit was used (Stoet, 2010, 2017). Participants were given a general description of the experiment and were told that headphones or earphones were required. As stated, the online version of the experiment was visited a total of 1096 times. The loudness pre-test, in which the participants' task was to choose the loudest tone of a series of five otherwise identical piano tones, was taken 1083 times and successfully completed in 35% of those attempts. The difficulty of the loudness pre-test was set relatively high in order to minimize the influence of the quality of participants' headphones, the environmental noise, and participants hearing deficiencies like hearing loss. The loudness pre-test included three separate trials, and participants were allowed to listen to the series of five tones from each trial as many times as they wanted before moving on to the next trial. The experiment was completed in 68% of the attempts that passed the loudness pre-test. This completion rate is relatively high compared to completion rates from other online experiments (Bosnjak & Tuten, 2003; O'Neil & Penrod, 2001; O'Neil, Penrod, & Bornstein, 2003; Tuten, Galesic, & Bosnjak, 2004).

In the two-oddball paradigm, participants were asked to respond to each item by choosing one chord that sounded most different to them in comparison to the other four. Figure 3 is a screenshot of the task as presented to participants. The playback controls at the top of the page disappeared immediately after participants clicked on the playback button. This meant that participants heard all the three repetitions of the five-chord sequence included in the audio clip unless they responded and moved on to the next trial before the 27-sec audio clip had played in its entirety. Participants were allowed to choose only one of the seven response options provided in the screen and they were able to change the response until they pressed the continue button. The continue button at the end of the page was enabled (turning from grey to blue) only after participant had selected a response.



Comparing chords (60 trials)

Which of the five chords in the clip above sounds the most different to you in comparison to the other four? (the clip consists of the exact same sequence of five chords played three times)

○ 1st chord
○ 2nd chord
○ 3rd chord
○ 4th chord
○ 5th chord
○ All chords sound identical to me
○ The audio didn't play properly

Figure 3. Screenshot of the task as presented to participants.

Design and analyses

We used measures of central tendency and distribution (averages, ranges, standard deviations) as well as shares (as percentages) for analysing general response patterns, and Pearson correlations for analysing the general role of musical education on responses. For grouping the participant background variables, we used principal components analysis (PCA) with varimax rotation, and for examining how well the PCA components explained the chord-type-oddball and voicing-oddball responses, we used stepwise linear regression. Finally, we modified the chord-type oddball score into a similarity score and ran a multidimensional scaling analysis (MDS) to deepen our understanding about the chord types used in the experiment. All analyses were made using the SPSS.

Results

General Response Patterns

First, we analyzed if there were any general response patterns. We calculated four main scores over each individual participant's responses (as percentages). "Chord-type-oddball score" and "voicing-oddball score" corresponded to how often participants chose chord-type and voicing oddballs, respectively. In addition, we calculated a main score for participants' choices of any of the three standard chords ("standard score"), and a main score showing how often the participants responded "all chords sound identical to me."

The mean of individual participants' responses show that, on average, the participants (N = 116) chose the chord-type oddball in 35.62% of the items, and the voicing oddball approximately as often, in 40.39% of the items (Table 4). This indicates that, at the participant level, they differentiated at least one of the two oddball chords from the standard chords in more than 75% of the items. They also chose one of the standards, even though rarely (*M* = 10.34%), and they heard all chords being the same in some items (*M* = 13.65%). Finally, there were also a few cases in which participants responded "no sound" (*M* = 0.32%). Taken together, chord-type-oddball or voicing-oddball responses formed the majority of the responses. Yet, no participant responded according to the same response category in 100% of the test items (see the Maximums in Table 4).

Table 4. General share of responses. Four main scores and “no sound” responses.

	M	**Min**	**Max**	***SD***
Chord-type oddball	35.62%	5.00%	91.67%	18.99%
Voicing oddball	40.39%	1.67%	75.00%	15.87%
Standard	10.34%	0.00%	51.67%	10.56%
“All chords sound identical to me”	13.65%	0.00%	58.33%	14.81%
“No sound”	0.32%	0.00%	5.00%	0.93%

There were six participants who chose the chord-type oddball in a very systematic way (in more than 75% of their responses). None of the participants chose the voicing oddball as systematically, indicating that whereas some participants focused on chord types and used them as a guiding principle in the experiment, no participant focused on voicings to that same extent.

During the entire experiment, each chord-type pair was presented using three voicings in both orders. We analyzed the consistency of each participant in their chord-type responses by comparing the odd and even responses for each chord-type pair (indicating one kind of split-half reliability analysis). The analysis showed high consistency (see Figure 4), and the correlation between the odd and even responses was high ($r_{116} = .89$, $p < .001$). A similar analysis showed high consistency for voicing responses as well ($r_{116} = .80$, $p < .001$).

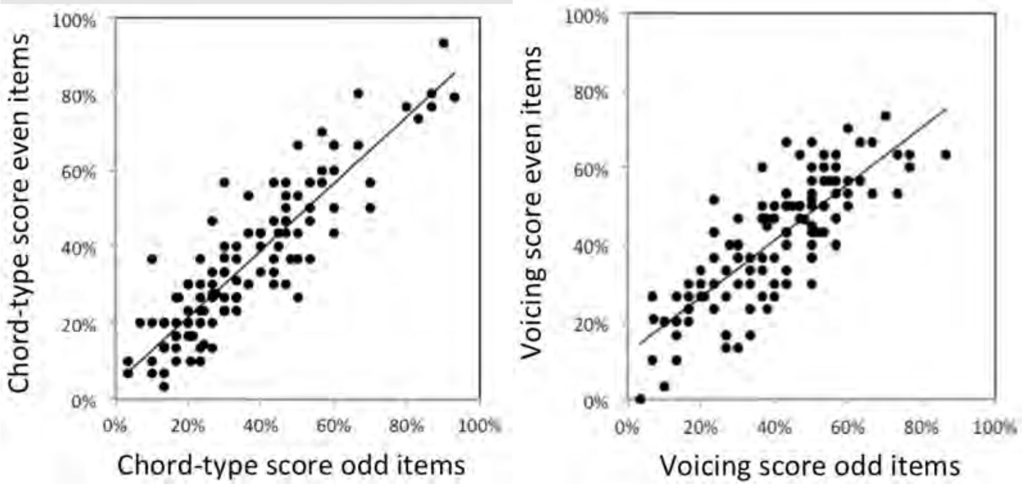


Figure 4. Individual participants' responses according to chord type and voicing for odd and even items.

Analysis of the participant background variables

Musical training, chord identification, and main scores

Since we were interested in identifying potential connections between musical training and responses, we calculated correlations between the four main scores defined above and an average score of main instrument years, music theory years, and ear training years (defined as a combined musical training variable; see Table 2 for the variables). These correlations are shown in Table 5, and as can be seen, musical training seems to be connected to chord-type oddball responses ($r_{116} = .42, p < .001$) but not to voicing oddball responses ($r_{116} = -.09, p = .337$). Additionally, musical training is negatively connected to "all chords sound identical to me" responses ($r_{116} = -.31, p < .001$).

As stated, the participants responded to a chord-identification post-test in which they were asked to identify single block chords by choosing one response from a list of chord types. The correct response was scored as 1 and incorrect response as 0, and the post-test score was a sum score of the 10 trials (see Table 2, variable V3). The correlations between the post-test score and the four main scores (see Table 5) show a relatively strong relationship ($r_{116} = .48, p < .001$) between participants' performance in the chord-ID post-test and chord-type oddball score. They also show a negative connection between chord-ID post-test and the voicing oddball score ($r_{116} = -.30, p < .001$), indicating that the better the participants identified chord-types by ear, the less they selected a voicing oddball. As expected, the participants' ability to identify chords was connected to their years of musical training ($r_{116} = .45, p < .001$).

Table 5. Correlations between four main scores, musical training, and chord-identification post-test score. N = 116.

	Chord-type oddball	**Voicing oddball**	**Standard**	**"All sound identical to me"**
Musical training (combined)	.422 (p < .001)	-.090 (*p* = .342)	-.189 (*p* = .042)	-.310 (*p* = .001)
Chord-ID post-test	.484 (p < .001)	-.303 (*p* = .001)	-.089 (*p* = .342)	-.232 (*p* = .012)

NOTE: Musical training (combined) is the average of years of main instrument, years of music theory, and years of aural skills.

Principal components analysis and regression analysis of participant variables

In addition to the music-training variables mentioned above, we also collected other participant background variables which were named and explained in Table 2. We ran a principal components analysis (PCA) with varimax rotation to group the 10 variables. Both the Kaiser-Meyer-Olkin test (.790) and the Bartlett's test ($C^2 = 612.395$; $df = 45$; $p < .001$) indicated that the data was suitable for a PCA analysis. The 10 variables loaded on 3 components, and the model explained 71.13 % of the variance (see Table 6). We interpreted and named the components according to the variables. Since music theory and analysis training as well as work with chords loaded on component C1, it was named as "Experience with chords". Variables that loaded on C2 were connected with the listeners' attention to either melody or chords while listening to music and with their experience with instrument playing. Since playing probably increases conscious attention pitches and by that to pitch structures, the component was named as "Attention to pitch structure". Component C3 was named as "Chord naming" since the variables that loaded on it were the self-reported knowledge of chords and the chord-type identification score.

Table 6. The Principal Component Analysis and the Rotated Component Matrix. with the highest loadings on each component marked with bold print.

Total Variance Explained

	Initial Eigenvalues			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.440	44.404	44.404	3.339	33.394	33.394
2	1.545	15.448	59.852	1.906	19.062	52.457
3	1.128	11.279	71.131	1.867	18.674	**71.131**

Rotated Component Matrix

	Components		
	C1 Experience with chords	**C2 Attention to pitch structure**	**C3 Chord naming**
V1_inst_years	.280	**.532**	.375
V2_chord_knowledge	.087	.198	**.880**
V3_aural_chord_ID	.301	.065	**.848**
V4_attention_to_melody	.133	**.868**	-.017
V5_attention_to_chords	.107	**.840**	.211
V6_play_chords_by_ear	**.683**	.206	.220
V7_play_chords_from_notes	**.504**	.212	.313
V8_music_theory_years	**.900**	.080	.138
V9_aural_skills_years	**.899**	.117	.101
V10_chord-type_ID_years	**.892**	.115	.111

Only the first three components with eigenvalues higher than 1.00 are shown.Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 5 iterations.

Thereafter we ran two separate stepwise regression analyses to see if the three components explained the choice of the chord-type oddball and the choice of the voicing oddball. The Durbin-Watson (1.473 for chord-type oddball and 1.857 for voicing oddball) as well as the residuals showed that the data could be analyzed using linear regression. Since the components were extracted using varimax rotation, there was no multicollinearity between them. Altogether the three components explained 16.1 % of the variance of the chord-type oddball score, but only 4.8 % of the voicing oddball score. This indicates that the participant background variables had a medium and statistically significant effect on the selection of chord-type oddballs ($p < .001$; for the effect size, see Cohen 1988), but they did not explain the selection of voicing oddballs (see Table 7 for chord-type oddball; no data provided for the voicing oddball).

Table 7. Model summary and ANOVA for the chord-type oddball score and coefficients for Model 3 with three predictors.

Model Summary[a]

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.220[b]	.048	.040	18.60426%	.048	5.808	1	114	.018	
2	.288[c]	.083	.067	18.34357%	.035	4.263	1	113	.041	
3	.427[d]	.183	.161	17.39509%	.100	13.659	1	112	.000	1.473

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2010.356	1	2010.356	5.808	.018[b]
	Residual	39457.517	114	346.119		
	Total	41467.873	115			
2	Regression	3444.889	2	1722.444	5.119	.007[c]
	Residual	38022.984	113	336.487		
	Total	41467.873	115			
3	Regression	7577.884	3	2525.961	8.348	.000[d]
	Residual	33889.989	112	302.589		
	Total	41467.873	115			

NOTE:
a. Dependent Variable: a1_chord_type_oddball_was_chosen
b. Predictors: (Constant), C1 (Experience with chords)
c. Predictors: (Constant), C1 (Experience with chords) and C2 (Attention to pitch structure)
d. Predictors: (Constant), C1 (Experience with chords), C2 (Attention to pitch structure), and C3 (Chord naming)

Coefficients

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
3	(Constant)	35.619	1.615		22.054	.000
	C1 (Experience with chords)	4.181	1.622	.220	2.578	.011
	C2 (Attention to pitch structure)	3.532	1.622	.186	2.177	.032
	C3 (Chord naming)	5.995	1.622	.316	3.696	.000

For chord-type oddballs, each of the three components added the explanatory power of the model, and each F change was statistically significant. The result that the participant background variables explained only the chord-type score is in line with our initial analysis revealing that musical training and the chord-ID post-test correlated positively with chord-type oddball responses but negatively with voicing-oddball responses and with “all chords sound identical to me” responses.

Participants’ conceptual knowledge of chord types

After participants had completed the main experiment, we asked them what aspect of the chords they paid attention to when selecting the oddball chords. When we analyzed the responses, we found three groups of participants. In the first group the participants had engaged conscious knowledge of concepts related to chord type (*Concepts group N* = 27), since in their free responses they included words that traditional music theory uses to refer to chord-type (e.g., major, minor, seventh, dissonant). In the second group the participants did not use those types of words, and later reported either not knowing what the terms "major chord" or "minor chord" mean or not being able to identify those chords by ear (*No-concepts group*; $N = 37$). The third group ($N = 52$) did not use chord-type related words either, but unlike the second group later reported both conceptual knowledge of and aural skills for the distinction between major and minor chords. In order to further analyze the effect of conceptual knowledge of chord types we counted the percentages of the chord-type oddball responses for each chord-type pair for the first and second groups of participants (see Figure 5). The two groups differed in a very systematic way in their responses since the percentages for the *Concepts* group were 10–29 units higher than the percentages for the *No-concepts* group. This addition to the percentages for the *Concepts* group can be called the effect of conceptual knowledge. At the same time the responses are highly correlated ($r_{10} = 0.98$; $p < .001$) suggesting that the relative gradation of perceptual salience of chord-type contrast is not dependent on conscious knowledge of chord types alone. The figure also shows that chord-type responses were most frequent for chord-type pairs involving two pitch-class changes and for pairs where the basic triad changes (major-based versus minor-based).

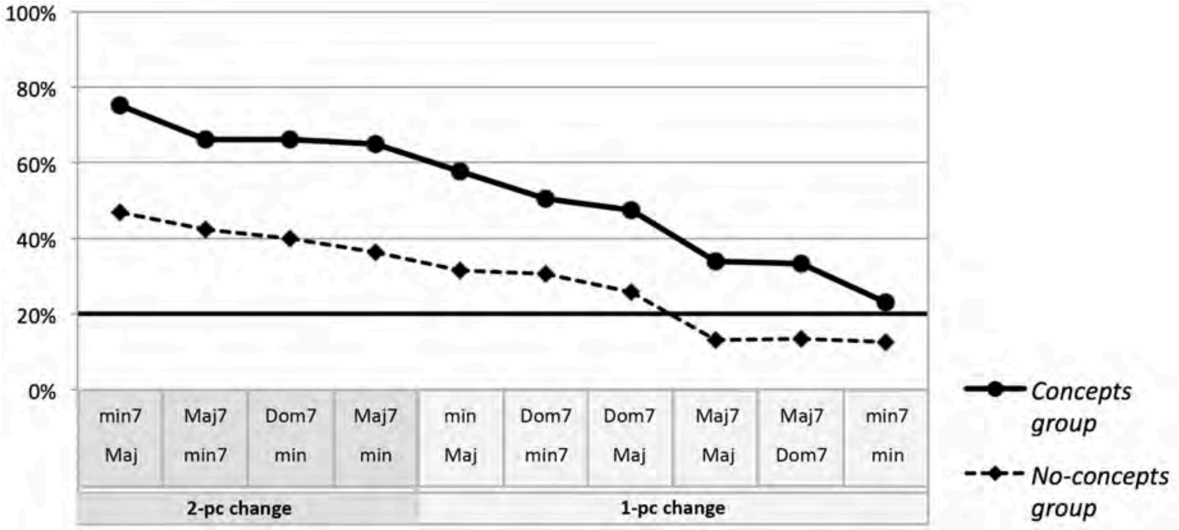


Figure 5. Chord-type score for all chord-type pairs and two groups of participants. The thicker black horizontal line indicates the 20% chance level.

MDS from chord-type data

Since the analysis of the conceptual knowledge of chord types showed that the responses were also affected by the number of changing pitch-classes and the type of the triad of the chord, we further analyzed these patterns. We ran a multidimensional scaling analysis using the chord-type oddball score for each chord-type pair in both orders of presentation. The analysis revealed a two-dimensional structure (Stress = .162; *RSQ* = .917) where dimension 1 can be interpreted as major versus minor (including both triads and tetrads) and dimension 2 as triads versus tetrads (Figure 6).

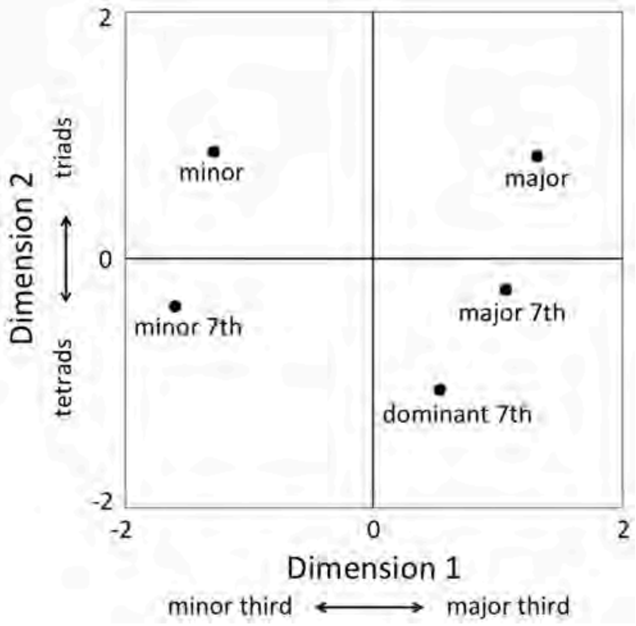


Figure 6. Two-dimensional solution of the similarities between chord types.

Discussion

In this study, the participants were asked to choose which of the five same-root block chords played in immediate succession sounded most different to them. The five-chord series included three chords that were identical to each other (standard chords), one chord that differed in chord-type (chord-type oddball), and one chord that differed in voicing of inner voices (voicing oddball). By this type of two-oddball paradigm we examined the relative salience of chord-type and chord-voicing changes. We found that the responses were affected by the participants' experience of music, the musically trained participants choosing the chord-type oddball more often than the non-trained ones. We also noticed that 86 participants (74.1 % of all) responded according to chord-type oddball above the 20% chance level, indicating that chord-type changes were detected even by participants without formal aural training of the chords or conceptual knowledge about chord types. The difference between major-based chords (major, major 7th, and dominant 7th) from minor-based chords (minor and minor 7th) seemed particularly salient, and the difference was more salient in pairs with two than one changing pitch-classes.

This experiment using the two-oddball paradigm with constant root has revealed several positive aspects of this paradigm. We have been able to calibrate the specific conditions of the paradigm to obtain evenly and widely spread scores suggesting that the paradigm captures differences between listeners. Importantly, these differences tend to be stable, that is, participants' response patterns show internal consistency as indicated by the high split-half reliability reported in Results. Additionally, the combination of the loudness pre-test and the two-oddball paradigm is particularly suitable for online testing since it facilitates the identification of non-serious respondents and survey bots.

The effect of musical training and knowledge of chords

We found effects of several participant variables on their performance, e.g., the years participants have regularly sung or played instruments and their score on chord-type aural identification test. We found that these variables had an effect on the chord-type responses but not on the voicing responses. However, as stated, chord-type oddballs were also chosen by participants without musical training, and the effect of the specific chord-type pair on these responses was similar for those participants who were aware of chord-type concepts and those who did not know the concepts, despite the general tendency for the former group to choose the voicing oddball more often (see Figure 5). These results strongly indicate that the responses cannot be attributed only to top-down processes. In other words, conceptual knowledge of chord-types may have influenced some responses but can by no means fully explain our results. Instead, it is likely that the extent to which participants' chose chord-type instead of voicing was also influenced by different degrees of perceptual sensitivity to (in our case, tonal) harmony. It seems that perceptual sensitivity to harmony can, to some extent, be learned implicitly, by listening to music – like many other musical abstractions (for a review, see, Rohrmeier & Rebuschat, 2012) while musical training advances conceptual knowledge about harmony. Future experiments using brain responses to oddball paradigms under "attend" and "ignore" conditions could deepen our understanding of the relative salience of chord-type and chord-voicing changes and help tease apart conceptual and surface-level factors involved in perception of harmony.

The effect of chord-type pair

The fact that chord-type responses were more frequent for chord-type pairs in which the basic triad changes (major-based versus minor-based; see Figures 5 and 6) than in pairs with no such change suggests that participants were more sensitive to the change of the third than to the other pitch-class changes used in our experiment. The finding that changes between major and minor chord quality were more salient than changes involving modification and addition or omission of the seventh (from a triad to a seventh chord or vice versa) might have a neurocognitive basis (for a review, see Virtala & Tervaniemi, 2017). It has also been shown that errors in automatic chord recognition algorithms more often involve sevenths than thirds (Nadar, Abeßer, & Grollmisch, 2019) and that expert human transcribers are a lot more likely to agree about the third of chords than about the presence or quality of sevenths (Koops, de Haas, Burgoyne, Bransen, Kent-Muller, & Volk, 2019). Our results are also consistent with traditional tonal theories that consider the root, third, and fifth of a chord to be more structural than sevenths and other extensions (e.g., Aldwell & Schachter, 1989).

The two-oddball paradigm could be considered as an indirect way to test similarity between chord types in a way that attenuates the influence of conceptual top-down process. The multidimensional nature of chords means that thresholds of "sameness" and "difference" can be difficult to establish without, e.g., concept-driven weighting of some chordal dimensions over others. The two-oddball paradigm does not require participants to establish a threshold of "sameness" and "difference"; instead, the task is to select the chord that sounds most contrasting. Hence, the two-oddball paradigm is likely to be more suitable for measuring differences between listeners' perceptual processes, especially if participants are not musically trained. Future experiments can further investigate the viability of a two-oddball paradigm as an alternative to direct similarity ratings.

References

Ahler, D. J., Roush, C. E., & Sood, G. (2019, April 4–7). The micro-task market for lemons: Data quality on Amazon's Mechanical Turk. Paper presented at the 77th *Annual Conference of the Midwest Political Science Association.* Chicago, IL, USA.

Aldwell, E., & Schachter, C. (1989). *Harmony and voice leading*. New York, NY: Harcourt Brace Jovanovich.

Armitage, J., & Eerola, T. (2020). Reaction time data in music cognition: Comparison of pilot data from lab, crowdsourced, and convenience Web samples. *Frontiers in psychology*, *10*, 2883. https://doi.org/10.3389/fpsyg.2019.02883

Beal, A. L. (1985). The skill of recognizing musical structures. *Memory & cognition*, *13*(5), 405–412.

Bosnjak, M., & Tuten, T. L. (2003). Prepaid and promised incentives in web surveys: An experiment. *Social Science Computer Review*, *21*(2), 208–217. https://doi.org/10.1177/0894439303021002006

Buonviri, N. O., & Paney, A. S. (2020). Technology use in high school aural skills instruction. *International Journal of Music Education, 38*(3) 431–440. https://doi.org/10.1177/0255761420909917

Christensen T. (2006). *The Cambridge History of Western Music Theory*. Cambridge Univ. Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Deliège, I., Mélen, M., Stammers, D., & Cross, I. (1996). Musical schemata in real-time listening to a piece of music. *Music Perception: An Interdisciplinary Journal*, *14*(2), 117–159. https://doi.org/10.2307/40285715

Dennis, S. A., Goodson, B. M., & Pearson, C. A. (2020). Online worker fraud and evolving threats to the integrity of MTurk data: A discussion of virtual private servers and the limitations of IP-based screening procedures. *Behavioral Research in Accounting*, *32*(1), 119–134. https://doi.org/10.2308/bria-18-044

Difallah, D., Filatova E, & Ipeirotis, P. (2018). Demographics and dynamics of Mechanical Turk Workers. In Yu, Q & Jianhui Chen (Eds.), *Proceedings of WSDM 2018: The Eleventh ACM International Conference on Web Search and Data Mining* (pp. 135–143). Association for Computer Machinery. Marina Del Rey, CA, USA. https://doi.org/10.1145/3159652.3159661

Doll, C. (2017). *Hearing harmony: Toward a tonal theory for the rock era*. University of Michigan Press.

Farbood, M. M. (2012). A parametric, temporal model of musical tension. *Music Perception*, *29*(4), 387–428. https://doi.org/10.1525/mp.2012.29.4.387

Goldman, A., Jackson, T., & Sajda, P. (2020). Improvisation experience predicts how musicians categorize musical structures. *Psychology of Music*, *48*(1), 18–34. https://doi.org/10.1177/0305735618779444

Granot, R. Y., & Jacoby, N. (2011). Musically puzzling I: Sensitivity to overall structure in the sonata form?. *Musicae Scientiae*, *15*(3), 365–386. https://doi.org/10.1177/1029864911409508

Granot, R. Y., & Jacoby, N. (2012). Musically puzzling II: Sensitivity to overall structure in a Haydn E-minor sonata. *Musicae Scientiae*, *16*(1), 67–80. https://doi.org/10.1177/1029864911423146

Huron, D. (1993). Chordal-tone doubling and the enhancement of key perception. *Psychomusicology: A Journal of Research in Music Cognition*, *12*(1), 73. https://doi.org/10.1037/h0094115

Huron , D. (2016). *Voice leading: The science behind a musical art*. Cambridge, MA: MIT Press.

Klein, M. E., & Zatorre, R. J. (2011). A role for the right superior temporal sulcus in categorical perception of musical chords. *Neuropsychologia, 49*(5), 878–887. https://doi.org/10.1016/j.neuropsychologia.2011.01.008

Koelsch, S., Schröger, E., & Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *Neuroreport*, *10*(6), 1309–1313.

Koops, H. V., de Haas, W. B., Burgoyne, J. A., Bransen, J., Kent-Muller, A., & Volk, A. (2019). Annotator subjectivity in harmony annotations of popular music. *Journal of New Music Research*, *48*(3), 232–252. https://doi.org/10.1080/09298215.2019.1613436

Kuusi, T. (2010). Comparing nontraditional tetrachords and pentachords: Both set-class and chord voicing guide evaluations. *Journal of New Music Research 39*(3)*,* 215–225. https://doi.org/10.1080/09298215.2010.502237

Kuusi, T. (2015). Musical training and musical ability: Effects on chord discrimination. *Psychology of Music*, *43*(2), 291–301. https://doi.org/10.1177/0305735613511504

Lamont, A., & Dibben, N. (2001). Motivic structure and the perception of similarity. *Music Perception: An Interdisciplinary Journal*, *18*(3), 245–274. https://doi.org/10.1525/mp.2001.18.3.245

Linnavalli, T., Ojala, J., Haveri, L., Putkinen, V., Kostilainen, K., Seppänen, S., & Tervaniemi, M. (2020). Musical expertise facilitates dissonance detection on behavioral, not on early sensory level. *Music Perception: An Interdisciplinary Journal*, *38*(1), 78–98. https://doi.org/10.1525/mp.2020.38.1.78

Nadar, C. R., Abeßer, J., & Grollmisch, S. (2019). Towards CNN-based acoustic modeling of seventh chords for automatic chord recognition. In *Proceedings* *for the 16th Sound and Music Computing Conference* (pp. 551–557). *Málaga, Spain*.

O'Neil, K. M., & Penrod, S. D. (2001). Methodological variables in web-based research that may affect results: Sample type, monetary incentives, and personal information. *Behavior Research Methods, Instruments, & Computers, 33*, 226–233. https://doi.org/10.3758/BF03195369

O'Neil, K. M., Penrod, S. D., & Bornstein, B. H. (2003). Web-based Research: Methodological variables' effects on dropout and sample characteristics. *Behavior Research Methods, Instruments, & Computers, 35*, 217–236. https://doi.org/10.3758/BF03202544

Putkinen, V., Tervaniemi, M., Saarikivi, K., Ojala, P., & Huotilainen, M. (2014). Enhanced development of auditory change detection in musically trained school-aged children: a longitudinal event-related potential study. *Developmental science*, *17*(2), 282–297. https://doi.org/10.1111/desc.12109

Rohrmeier, M. & Rebuschat, P. (2012). Implicit Learning and Acquisition of Music. *Topics in Cognitive Science 4*, 525–553. https://doi.org/10.1111/j.1756-8765.2012.01223.x

Schubert, E., & Stevens, C. (2006). The effect of implied harmony, contour and musical expertise on judgments of similarity of familiar melodies. *Journal of New Music Research*, *35*(2), 161–174. https://doi.org/10.1080/09298210600835000

Scott, R. J. (2000). *Money chords: A songwriter's sourcebook of popular chord progressions*. Bloomington, IN: iUniverse.

Snodgrass JS. (2016). Current status of music theory teaching. *College Music Symposium*. https://www.jstor.org/stable/26574444

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42*(4), 1096–1104. http://www.doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time *experiments. Teaching of Psychology, 44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Sturm, I., Blankertz, B., & Curio, G. (2017). Multivariate EEG analysis reveals neural correlates for the differential perception of chord progressions. *Psychomusicology: Music, Mind, and Brain*, *27*(4), 281–296. https://doi.org/10.1037/pmu0000196

Tervaniemi, M., Sannemann, C., Nöyränen, M., Salonen, J., & Pihko, E. (2011). Importance of the left auditory areas in chord discrimination in music experts as demonstrated by MEG. *European Journal of Neuroscience*, *34*(3), 517–523. https://doi.org/10.1111/j.1460-9568.2011.07765.x

Thomson, W. & Blombach, A. (1988). Reader's Response: What is an interval? *Journal of Music Theory Pedagogy*, *2*(2), 321–325.

Tuten, T. L., Galesic, M., & Bosnjak, M. (2004). Effects of immediate versus delayed notification of prize draw results on response behavior in Web surveys – An experiment. *Social Science Computer Review, 22* (3), 377–384. https://doi.org/10.1177/0894439304265640

Virtala, P., Huotilainen, M., Partanen, E., & Tervaniemi, M. (2014). Musicianship facilitates the processing of Western music chords—An ERP and behavioral study. *Neuropsychologia*, *61*, 247–258. https://doi.org/10.1016/j.neuropsychologia.2014.06.028

Virtala, P., Huotilainen, M., Lilja, E., Ojala, J., & Tervaniemi, M. (2018). Distortion and Western music chord processing: An ERP study of musicians and nonmusicians. *Music Perception: An Interdisciplinary Journal*, *35*(3), 315–331. https://doi.org/10.1525/mp.2018.35.3.315

Virtala, P. & Tervaniemi., M. 2017. Neurocognition of Major-Minor and Consonance-Dissonance. *Music Perception: An Interdisciplinary Journal, 34*(4), 387–404. https://doi.org/10.1525/MP.2017.34.4.387

Williams, L. R. (2005). Effect of music training and musical complexity on focus of attention to melody or harmony. *Journal of Research in Music Education*, *53*(3), 210–221. https://doi.org/10.1177/002242940505300303

Appendix

Appendix Table 1. *Total attempts*

	Count	**% from subtotal**	**% from total**
Completed attempts			
Participants whose data were included in the main analysis	116	45%	11%
Completed attempts discarded because responses did not pass our "seriousness" criteria	128	49%	12%
Completed attempts discarded because participants responded "all chords sound identical to me" more than 85% of trials and the rate of that type of response increased throughout their taking the experiment	3	1%	0%
Completed attempts discarded because participants had already taken the experiment	12	5%	1%
Subtotal	259	100%	24%
Incomplete attempts			
Attempts in which respondents passed the loudness pre-test but did not complete the entire experiment	72	59%	7%
Attempts abandoned by respondents during the loudness pre-test	4	3%	0%
Attempts abandoned by respondents before loudness pre-test	24	20%	2%
Attempts abandoned by respondents before responding to the first question about headphones	22	18%	2%
Subtotal	122	100%	11%
Attempts not wearing headphones or not passing the loudness test			
Attempts not allowed to take the experiment because respondents failed to answer the pre-test correctly	702	98%	64%
Attempts not allowed to continue because respondents reported not wearing headphones	13	2%	1%
Subtotal	715	100%	65%
Total attempts	1096		100%