The effect of chord duration on the relative salience of chord-type and voicing changes

Ivan Jimenez,[1] Tuire Kuusi,[1] & Juha Ojala[1] & Peter Harrison[2]

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

[2] Centre for Music and Science, University of Cambridge, United Kingdom

Author Note

This paper is the *Final Draft* of the following article:

Jimenez, I., Kuusi, T., Ojala, J., & Harrison, P. (2024). The effect of chord duration on the relative salience of chord-type and voicing changes. *Psychology of Music*, (*Online First)*. DOI: 10.1177/03057356241247533

Correspondence concerning this article should be addressed to Ivan Jimenez, Sibelius Academy, University of the Arts Helsinki, Töölönlahdenkatu 16 C, 00100 Helsinki, Finland.

Email: ivan.jimenez.rodriguez@uniarts.fi

Abstract

This study investigates the effect of chord duration on the relative salience of chord-type and voicing changes. Participants (N = 111) with varying levels of musical training were presented with sequences of five block chords on the piano and asked to indicate which chord sounded most different. Each sequence consisted of three identical chords and two oddballs, one with a voicing change and one with a chord-type change. All possible chord-type pairings between standard and oddball formed of major, minor, dominant seventh, major seventh, and minor seventh chords were tested. Additionally, each sequence of five chords was tested using three chord duration conditions (500, 1000, and 1500 ms), and the durations were pseudo-randomized throughout the experiment. Chord-type changes became more salient with longer durations and this effect could be seen for all participants regardless of their levels of musical training. However, with higher level of musical training chord-type changes became more salient across all duration conditions. Leman's model of tonal contextuality suggests that the effect of duration in our experiment could be explained by sensory mechanisms related to echoic memory. The potential contribution of other factors to the effect of duration is discussed.

Keywords: chord type, voicing, salience, chord duration, echoic memory, tonal contextuality

The effect of chord duration on the relative salience of chord-type and voicing changes

Harmony is one of the fundamental elements of Western tonal music and it is often analyzed and discussed in terms of chords. The pitch content of chords can be described in two basic ways: chord type (e.g., major chord, minor seventh chord) and chord voicing (distribution of the chord tones across register). While both chord type and chord voicing (hereafter, voicing) refer to the intervallic relationships between the pitches of a chord, pitch register is only considered in the latter. For instance, the chord formed by the pitches C4, E4, and G4 and the chord formed by the pitches C3, G4, and E5 are identical to each other in terms of chord type (i.e., both chords are major) but not in terms of voicing (i.e., pitches are more widely spaced in register in the second chord). Traditional harmonic analysis often treats chord type as an integral part of the harmonic identity while voicing tends to be considered harmonically relevant only in exceptional cases (e.g., voicing a chord by placing its fifth in the bass can change its tonal function). In the realm of musical performance, simplified music notation (e.g., figured bass or lead sheets) specifies the chord type while leaving many aspects of the voicing open, for the performer to decide. This suggests that, at least in some contexts, voicing is usually treated as a surface-level feature.

A recent study showed that participants with experience in harmonic analysis and playing chords tend to hear chord-type changes as being more salient than voicing changes (Jimenez et al., 2022a). This finding is consistent with previous studies showing that musical training is associated with greater attention to deep structural features (e.g., Deliège et al., 1996; Schubert & Stevens, 2006). However, Jimenez et al. (2022a) also showed that, in sequences of chords with the same root, changes between common chord types can be experienced by both musicians and non-musicians.

In addition to chord type and voicing, chords have multiple other features such as duration, timbre, and texture that are inevitably involved when pitches are instantiated in the physical world. When compared to voicing, these other features are included even less in harmonic analysis or simplified chord notation, even though there is reason to believe that they can affect the relative salience of chord type and other aspects of harmonic perception. For instance, there is some evidence that differences in amplitude envelope, a feature concerning both loudness and timbre but not pitch chroma, can hinder the identification of major and minor chords played with sinusoidal tones (e.g., Skorik et al., 2018). Further, Tekman and Bharucha (1998) showed that even small changes in short-chord durations could determine whether sensory or cognitive mechanisms predominate during the mental processing of chords.

The role of chord duration on the perception of chord type

Some aspects of the perception of chord types act very rapidly. For instance, event-related potential (ERP) studies have shown that the brain can automatically detect the structural difference between major and minor chord types in ~150 ms (e.g., Virtala et al., 2011; Virtala et al., 2014). However, it usually takes longer than 800 ms to consciously distinguish between major and minor chords (Bharucha & Stoeckig, 1986; Virtala et al., 2014) and even longer to distinguish different types of seventh chords (Dunmire, 2019). Furthermore, when assessing whether major and minor chords separated by distraction tones are the same or not, musicians become more accurate when chord durations are increased from 300 to 1000 ms (Pechmann, 1998).

Most Western-enculturated listeners might also be accustomed to hearing relatively long chords. For instance, chord duration in Western popular music tends to be in the range of 1–3 seconds (de Clercq, 2022), which is well above the time used in the above-mentioned studies.[1]

Except for studies in which participants are asked to select the duration of the chords (e.g., Bigand et al., 1999), behavioral experiments that use chordal stimuli are inevitably forced to select a specific duration for their chords. Although chord durations in these experiments range from 50 ms (e.g., Tekman & Bharucha, 1992) to 4 seconds (e.g., Lerdahl & Krumhansl, 2007), most authors offer little or no explanation for their choice of chord duration, implicitly downplaying the role that duration has in harmonic perception. Early studies on the effect of chord duration on harmonic perception (Tekman & Bharucha, 1992, 1998; Bigand et al., 2003) found that cognitive mechanisms would prevail over sensory mechanisms for a wide range of chord durations except for extremely short chords but found no other effect of chord duration. However, these findings come from the specific context of harmonic priming paradigm which does not rule out the possibility that chord duration could affect chord perception in other contexts. A better understanding of this effect would lead to better-informed decisions regarding the construction of chordal stimuli, and it would also help clarify whether the large range of chord durations found in tonal music (e.g., de Clercq, 2022; Jimenez et al., 2023), has any implications for harmonic perception. More broadly, this study further develops a critical, integrated understanding of how the perception of harmony is tied to the perception of *extra-*

---

[1] In de Clercq (2022), chord duration was defined to be the interval of time between two successive non-identical chord symbols which were encoded by human annotators. These annotations followed the standard approach for describing chords in sheet music of popular music, which notates changes of root (e.g., C to F), quality (e.g., C to Cm), bass (e.g., C to C/E), or extension (e.g., C to C7).

harmonic features: other, inevitably involved musical features such as duration, timbre, and loudness.

The current study increases the understanding of the effect of chord duration on chord perception by using a two-oddball paradigm. It investigates the relative salience of two competing elements, chord-type and voicing changes, in three different durations. A recent study (Jimenez et al., 2022a) demonstrated the high sensitivity of the two-oddball paradigm as a measuring tool for harmonic perception, allowing the study on chord duration. This paradigm is also capable of measuring nuanced but consistent differences between participants in terms of their perception of the relative salience of chord-type and voicing changes. Although this paradigm showed that musical training increased the perceived salience of chord-type over voicing changes, the paradigm is adequate for testing non-musicians since it does not demand any knowledge of chord types (nor any other conceptual knowledge about music or harmony). Further, Jimenez et al. (2022a) revealed that the non-musicians' responses to this paradigm are individually consistent and demonstrate~~d~~ the same salience ranking of the different chord-type changes as the responses of musically trained participants.

Leman's model and tonal contextuality

Leman created a model to describe the contribution of echoic memory to the tonal perception of individual pitches (Leman, 2000). More recently, the model has been used for predicting aspects of chord perception (e.g., Koelsch et al., 2007; Bigand et al., 2014; Goldman et al., 2021). It considers not only the relationship between pitches but also their duration, and it simulates psychoacoustic processes that have been proposed to be involved in tonal perception. It takes raw audio files as its input and filters them through a simulation of the peripheral auditory system to produce auditory nerve images and periodicity-pitch images. It then simulates the accumulation of periodicity-pitch images in an echoic memory buffer. This is akin to playing a sound in a reverberant space; since every individual sound takes a while to dissipate, the listeners end up with a ‘smearing’ whereby at any given point in time they hear a superposition of the most recent sounds to be played.

Leman’s model computes two memory buffers and continuously updates them. The short-term memory buffer represents the current auditory event while the long-term memory buffer represents the preceding context. At each point in time, the model computes the correlation between the current auditory event (local periodicity-pitch image in the short-term memory buffer) and the preceding context (global auditory periodicity-pitch image in the long-term memory buffer). The correlation between local and global images is known as “tonal contextuality” (Marmel et al., 2010). High tonal contextuality (TC) means that the current auditory event is spectrally consistent with its preceding auditory context. TC is likely to be affected by chord-type and voicing changes, and we were interested to see how TC is affected by the different durations of our stimuli. If the level of TC changed with chord duration and if these changes were consistent with participants’ responses in our experiment, it would indicate that the accumulation of spectral information in echoic memory simulated by Leman’s model may explain the role of duration for the relative salience of chord changes.

Aims and hypotheses

Our current study used a two-oddball paradigm to test the effect of chord duration on the relative salience of chord-type changes and voicing changes (for chord-type and voicing oddballs, see stimuli). We expected to find that the salience of chord-type change increases as the duration increases, especially for participants with higher level of musical training and ability to identify and label chord types. This finding would underline the fact that longer durations facilitate the conscious mental labelling of chord type. However, we expected longer durations would increase chord-type responses even by participants who have no experience with labeling chord types by ear. This kind of effect would suggest that the longer duration of a chord gives time for richer experiences of chord qualities, chord type being one of those qualities even when listeners are not trying or able to consciously identify chord types.

The main analysis of participants' responses will be complemented by an exploratory analysis using Leman's model. This model has proved successful in predicting outcomes from various experiments on tonal perception (Bigand et al., 2014; c.f. Marmel et al., 2020). By analyzing the data using the Leman model, we expected to see that low TC between the standard chords and the oddball would make it easier for the participants to differentiate the oddball from the standards. Furthermore, we expected that chord-type oddballs have lower TC than the voicing oddballs, indicating a potential echoic-memory-related sensory component of the salience of chord-type change in our experiment.

Method

Participants

The online experiment was completed by 115 participants.[2] Since 4 of the participants had taken the experiment twice and we only accepted the first response from them, the total number of participants whose responses were included in our main analysis

---

[2] For information about the online experiment, pre-testing and criteria for rejecting non-serious participants or autofillers, see supplementary material 1.

was 111 (71 male, 39 female, 1 other; age *M* = 36.99, *SD* = 13.22). Although there were more male participants than females, there is no evidence to our knowledge of a gender effect on the perception of chord type or voicing. We collected background information of the participants through a questionnaire to which they responded at different points during the experiment.

The 111 participants were divided into three groups according to their responses to the questionnaire. We also analyzed the differences between the three groups in terms of background variables, and the analysis supported the grouping (see supplementary material 2).

- Group 1: *Concept-oriented* participants (*N* = 20, 17 males, Mean age 40.5 years, *SD* = 12.61 years) mentioned chord types or chord-type labels in the post-experiment questions. Abbreviated as "c-o" in tables and figures.
- Group 2: *Playing-oriented* participants (*N* = 41, 26 males, Mean age 39.4 years, *SD* = 14.73 years) did not mention chord types in the post-experiment questions but self-evaluated their instrument-playing or singing ability being beyond beginner's level. Abbreviated as "pl-o."
- Group 3: *Listening-oriented* participants (*N* = 50, 28 males, Mean age 33.62 years, *SD* = 11.50 years) neither mentioned chord types nor played any instrument or self-evaluated their instrument-playing or singing ability being at beginner's level. Abbreviated as "l-o."

Stimuli

The stimuli and procedure followed that described by Jimenez et al. (2022a). Each item consisted of a sequence of five chords: three standards, one chord-type oddball, and one voicing oddball. All ten possible chord-type pairings between standard and oddball formed of major, minor, dominant seventh, major seventh, and minor seventh chords were tested. We used five-pitch chords with the chord root in the outer pitches and all changes occurring in the inner pitches. The voicing oddballs differed from the standard chords by two pitch-register changes (one or two octaves; see chord v in Figure 1). The chord-type oddballs differed by either one or two pitch classes (and, consequently, pitches), and the pitches always moved to the neighboring tone, that is, there were no octave changes and register changes (see chord t in Figure 1).

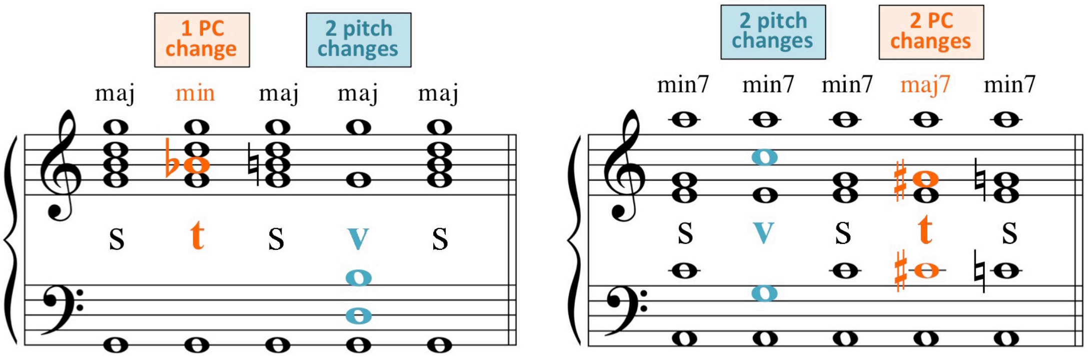


Figure 1. Examples of Items Used in the Experiment.
PC= Pitch class, s = standard chords, t = chord-type oddball, and v = voicing oddball.

We used fixed-root successions since their suitability for studying differences among non-musicians has already been demonstrated in a previous experiment (Jimenez et al., 2022a). To reduce participants' fatigue and habituation, the five-chord items were played on three different pitch levels: G, Ab, and A.

Timbre and loudness. We composed the sixty chords (4 voicings, 5 chord types, 3 transpositions) using the procedure described by Jimenez et al. 2022a. The Bösendorfer sampled piano sound from Logic Pro X was used, and the notes in the outer voices (G2, Ab2, A2, G5, Ab5, A5) were made louder (MIDI key-velocity range = 61–77) than the inner voices (MIDI key-velocity range = 39–46). Inner voices were made softer to decrease the chances of hearing the pitch changes as forming individual voices. For a balance between (a) how naturalistic the piano tones sounded and (b) how the spectral and spectro-temporal differences between the three duration conditions could be minimized, we took the measures described in supplementary material 3.

Duration. We defined the chord duration as the inter-onset interval (IOI), that is, the duration of each block-chord plus a silent inter-stimuli interval. Three different IOIs (500, 1000, and 1500 ms) were used and the choice of the IOIs was influenced by results from earlier research. First, an earlier study using the same two-oddball paradigm and 1500 ms IOI obtained results that were relatively balanced in terms of chord-type and voicing responses (Jimenez et al., 2022a). Second, our pilots indicated that 5-chord items with IOIs longer than 1500 ms were too demanding for non-musicians in terms of working memory. Third, 500 and 1000 ms IOIs have often been used in chord priming experiments (e.g., Tekman & Bharucha, 1992, 1998; Tillmann & Bigand, 2004) whereas 1500 and 1000 ms IOIs are within the typical range of chord duration in Western popular music (de Clercq, 2022). Finally, Jimenez et al. (2023) showed that piano block chords shorter than 500 ms can provide enough information for listeners to identify songs and that chord type modulates such identification. To minimize the potential rhythmic uncertainty derived from listening to three different IOIs, the duration of the block chords was set to 300, 800, and 1300 ms, which meant that the onset of chords 2, 3, 4, and 5 was always cued by a 200 ms silent interval.

Serial position. The total number of the main items was 120, consisting of the 10 chord-type pairs played in both orders (chord-type A of the pair being used to create four of the five chords within a stimulus, see Figure 1, and chord type B being the one chord-type oddball and vice versa) and each order played using the two voicing-pair types and the three duration conditions.

The 120 items were composed making sure that:

1) Chord-type and voicing oddballs never followed one another 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 oddballs never occurred in the last serial position. Unlike with the first four chords, the auditory image of the last chord can linger in echoic memory without any sound interference until the next item. This could increase the salience of the last chord and decrease the effect of duration conditions. This decision was balanced by additional items explained in supplementary material 4.
3) Chord-type and voicing oddball occurred roughly equally often in each of the remaining four serial positions.

Decay parameters in Leman's model

To calculate TC, the duration of the short-term memory buffer (local decay) and the long-term memory buffer (global decay) are first specified by half-life parameters that dictate how long a sound takes to decay to half of its original level. The *local decay* could be understood as determining the level of smoothing for the local auditory image. The main function of the smoothing is to reduce the potential impact of artifacts from the input signal. The *global decay* represents the duration of echoic memory which is the "lingering or resonance of [auditory] sensory experiences for a brief period of time" (Huron & Parncutt, 1993, p. 157). Although there is no consensus regarding the duration of echoic memory, with estimates ranging from 0.5 to 60 sec, most estimates are in the range of 2 ± 1.5 sec (Huron & Parncutt, 1993; Nees, 2016).

Considering the model's sensitivity to the decay parameters (Bigand et al., 2014) we decided to trial Leman's model with a wide range of decay parameters. The local decay varied from 100 to 400 ms by steps of 100 ms and the global decay, from 0.5 to 4 sec by steps of 0.25 sec. Altogether, there were 75 combinations of global and local decay, and for most analyses, we averaged them to get one TC value for each chord.[3]

Procedure

Participants were mainly recruited using Amazon Mechanical Turk (MTurk), a crowdsourcing platform that provides access to more than a hundred thousand potential participants (Difallah et al., 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 like MTurk. Because of the low number of musically trained participants in MTurk, we also recruited participants by sending the links to music institutions in Europe and the US.

We used the software PsyToolkit (Stoet, 2010, 2017) for data collection. In the two-oddball paradigm, the participants were asked to respond to each item by choosing one chord that sounded most different to them in comparison to the other four. The main items were presented in a pseudo-randomized order in which roots, chord duration, and standard chord type always changed from one trial to the next. To reduce participants' fatigue, the 150 trials (120 main items and 30 additional items) were presented in three blocks of 40 trials and one final block of 30 trials. Participants were asked the questions from all five subscales of Gold-MSI (Müllensiefen et al., 2014) between blocks. After finishing, participants were asked additional questions about their experience with music. Participants who self-reported some ability to identify major and minor chords by ear were also asked to complete a chord-type

---

[3] It should be noted that TC was calculated approximately every 10 ms throughout the full duration of each stimulus (see Appendix B). Consequently, we initially obtained multiple (approximately 50, 100, or 150) TC values per chord, and they were averaged to obtain a single averaged TC value per chord.

identification post-test. Most participants completed the entire session in less than 50 minutes.

Results

We analyzed the responses of chord-type oddball and voicing oddball separately. Because each participant had responded to all three duration conditions, we used *General Linear Model Repeated Measures* for the analyses. The three participant groups were compared (between-subject factor) and the three duration conditions formed the within-subjects factor. The responses from each participant group are shown in Figure 2. In general, the responses according to chord type increased with longer durations while the responses according to voicing decreased with longer durations.

The multivariate test (Table 1) showed that the three participant groups differed from each other ($F(4, 216) = 19.066, p < .001$)[4] and the difference can especially be seen with chord-type responses in Figure 2. The *concept-oriented* participants responded according to chord-type in 49.7% of the 500 ms duration trials, while the *playing-oriented* participants responded lower than this even in the 1500 ms trials (37.2%). Further, the response rate of the *playing-oriented* participants for the 500 ms condition was approximately the same (25.6%) as the response rate for the *listening-oriented* participants for the 1500 ms condition (23.9%). The post-hoc test confirmed the finding: All three groups of participants differed from each other on the $p > .001$ level (see the statistics in Table 2).[5]

---

[4] The need to use Pillai's Trace criterion and correction of degrees of freedom was caused by the violation of the assumption of sphericity (measured by Box's M).

[5] We used Tamhane post-hoc test for the analysis of chord types because the Levene's test showed that the assumption of equal variances was violated for those responses. However, we used Bonferroni post-hoc test for the analysis of voicing responses because for those responses the variances were equal.

The changes in the voicing responses, on the other hand, were smaller (Figure 2). The *concept-oriented* group seemed to differ from the other two with response rates varying between 35.8% (500 ms chords) and 28.8% (1500 ms chords), but the Bonferroni post-hoc test (Table 2) showed that the *concept-oriented* group differed from the *listening-oriented* group on a statistically significant level (p = .035) but did not differ from the *playing-oriented* group (p = .128). Further, the response rates of the *playing-oriented* and *listening-oriented* group were very similar to one another, particularly for 500 ms chords where the rates for both groups were practically the same (41.9% for *playing-oriented* and 41.8% for *listening-oriented*). For 1500 ms chords, the ratings differed a little (37.1% for *playing-oriented*, and 41.3% for *listening-oriented*). The result was confirmed by the post-hoc test showing no difference in voicing responses from groups *playing-oriented* and *listening-oriented* (p = 1.000).

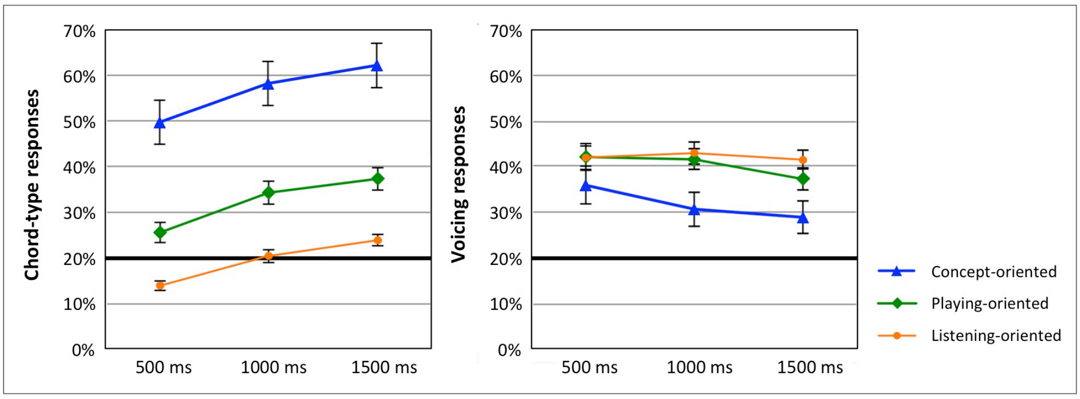


Figure 2. Percentage of Chord-type Responses (Left) and Voicing Responses (Right) for Three Groups of Participants and Three Duration Conditions.
Error bars indicate standard errors. The thick black line indicates the 20% chance level.

Table 1. Multivariate Tests for Three Participant Groups and Three Duration Conditions.

Multivariate Tests^a

Effect			Value	F	Hypothesis df	Error df	Sig.
Between Subjects	Intercept	Pillai's Trace	.954	1097.646^b	2	107	<.001

	Group	Pillai's Trace	.522	19.066	4	216	<.001
Within Subjects	Duration	Pillai's Trace	.584	36.924^b	4	105	<.001
	Duration * Group	Pillai's Trace	.090	1.249	8	212	.272

a. Design: Intercept + Group
Within Subjects Design: duration
b. Exact statistic
c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Table 2. Post-hoc Tests for the Three Participant Groups.

Post Hoc Tests
Multiple Comparisons

Measure		(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Chord type	Tamhane	*c-o*	*pl-o*	24.339%	5.178%	<.001	11.194%	37.483%
			l-o	37.257%	4.781%	<.001	24.855%	49.660%
		pl-o	*l-o*	12.919%	2.478%	<.001	6.822%	19.015%
Voicing	Bonferroni	*c-o*	*pl-o*	-8.409%	4.098%	.128	-18.374%	1.555%
			l-o	-10.189%	3.975%	.035	-19.855%	-0.523%
		pl-o	*l-o*	-1.779%	3.1652%	1.000	-9.477%	5.918%

C-o = concept-oriented, *pl-o* = playing-oriented, *l-o* = listening oriented.
Based on observed means.
The error term is Mean Square(Error) = 225.696.

The multivariate test (Table 1) also showed that the three duration conditions differed ($F(4, 105) = 36.924$, $p < .001$). Further, the univariate tests (Table 3) showed that the duration had a statistically significant effect on both the chord-type responses ($F = 86.172$, $p < .001$) and voicing responses ($F = 6.101$, $p = .004$).[6] The percentage of chord-type responses increased more than 10 percentage points for *concept-oriented* participants (from 49.7% to 62.1%) and *playing-oriented* participants (from 25.3% to 37.2%), and a little less than 10 percentage points for the *listening-oriented* participants (from 14.0% to 23.9%). The changes of duration were statistically

---

[6] The Maulchly's test of sphericity showed that the sphericity-assumed assumption was violated for both the chord-type variable and the voicing variable. Hence, we corrected the degrees of freedom by using the Huynh-Feldt correction in the Univariate tests.

significant for the chord-type responses, both when the duration changed from 500 to 1000 ms (F = 71.569, p < .001) and when the duration changed from 1000 to 1500 ms (F = 104.103, p < .001; see Table 4 for the statistics). On the other hand, the total decrease of voicing responses was 7 percentage points for *concept-oriented* participants (from 35.8% to 28.8%), and less than 5 percentage points for *playing-oriented* participants (from 41.9% to 37.1%) and *listening-oriented* participants (from 41.8% to 41.3%). The statistics (Table 4) further confirm that the change of duration had no effect in voicing responses between 500 and 1000 ms. Only between durations 1000 and 1500 ms can we find a statistically significant difference (F = 11.193, p = .001).

Table 3. Univariate Tests for Chord-type and Voicing Responses.

Univariate Tests

Source	Measure		Type III Sum of Squares	*df*	Mean Square	*F*	Sig.
Duration	Chord type	Huynh-Feldt	6425.321	1.938	3315.099	86.172	<.001
	Voicing	Huynh-Feldt	798.430	1.822	438.273	6.101	.004
Duration * Group	Chord type	Huynh-Feldt	85.280	3.876	22.000	.572	.678
	Voicing	Huynh-Feldt	507.622	3.644	139.322	1.939	.112
Error(duration)	Chord type	Huynh-Feldt	8052.877	209.326	38.471		
	Voicing	Huynh-Feldt	14134.050	196.751	71.837		

Table 4. Statistics for the Effect of Duration on Chord-type and Voicing Responses.

Tests of Within-Subjects Contrasts

Source	Measure	Duration	Type III Sum of Squares	*df*	Mean Square	*F*	Sig.
Duration	Chord type	500 vs 1000 ms	5882.239	1	5882.239	71.569	<.001
		1000 vs 1500 ms	5226.303	1	5226.303	104.103	<.001
	Voicing	500 vs 1000 ms	236.621	1	236.621	1.688	.197
		1000 vs 1500 ms	1020.179	1	1020.179	11.193	.001
Duration * Group	Chord type	500 vs 1000 ms	129.096	2	64.548	.785	.459
		1000 vs 1500 ms	31.098	2	15.549	.310	.734
	Voicing	500 vs 1000 ms	553.851	2	276.925	1.975	.144

		1000 vs 1500 ms	346.046	2	173.023	1.898	.155
Error (duration)	Chord type	500 vs 1000 ms	8876.466	108	82.190		
		1000 vs 1500 ms	5421.966	108	50.203		
	Voicing	500 vs 1000 ms	15143.144	108	140.214		
		1000 vs 1500 ms	9843.717	108	91.146		

Tonal contextuality

To analyze the relationship between TC, our stimuli, and participant responses, we first calculated the averaged TC values (see *Method and footnote 3*) for the chord-type oddballs and the voicing oddballs using the 75 combinations of local and global decays. We then calculated correlations between participant responses (percentages of chord-type and voicing responses) and the averaged Leman's TC for each oddball. The correlations (-0.74 for chord-type oddballs and -0.62 for voicing oddballs) indicated that lower TC increased the discrimination of the oddballs. It is possible that the correlation between Leman's model and participants' responses was high not only because this model includes both the pitch and duration of the individual chords but also because it includes the accumulated duration from the immediate repetition of the exact same chord or same chord-type in the item, as well as fine-grained spectral information of the audio stimuli which includes the register changes of the voicing oddballs. The correlation was higher for the chord-type oddballs than the voicing oddballs, suggesting that TC can predict the different degrees of salience of various chord-type and voicing changes and that echoic memory may play a role in the effect of duration on harmonic perception.

Our analyses also showed that the averaged TC values for the chord-type oddballs were lower than for voicing oddballs for all durations (Figure 3), indicating that changes in pitch-class content were more salient in terms of TC than the register changes we used. The differences of the TC in the same-duration voicing and chord-type oddballs were small, but the paired-samples t-test showed that the differences were statistically significant (Table 5).

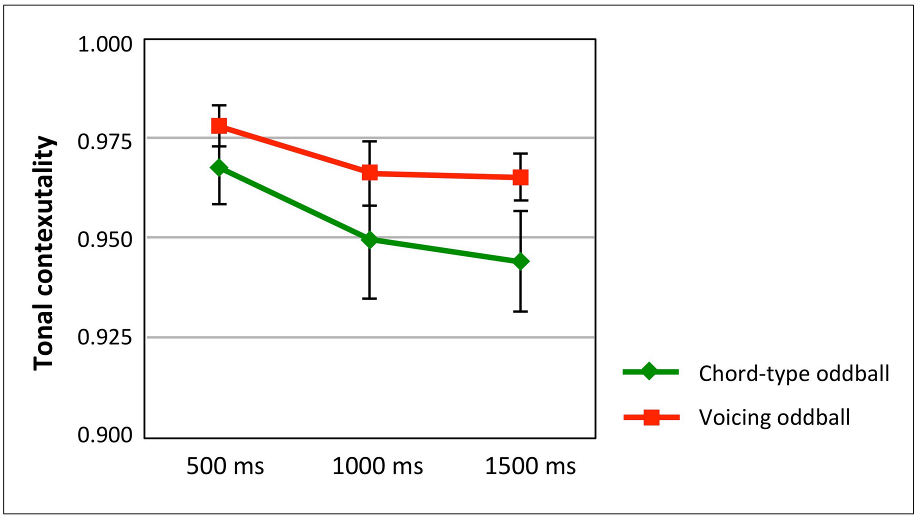


Figure 3. Average Tonal Contextuality for Chord-type and Voicing Oddballs in the Three Duration Conditions.
Tonal contextuality for each oddball was calculated using seventy-five different combinations of global and local decay.

Table 5. Paired-samples t-test for Same-duration Chord-type (t) and Voicing (v) Oddballs.

Pair	mean	st.dev	st. error mean	*t*	*df*	*p*
t_500 v_500	-.010495	.026076	.004123	-2.545	39	.015
t_1000 v_1000	-.016531	.042374	.006700	2.467	39	.018
t_1500 v_1500	-.020973	.033431	.005286	-3.968	39	< .001

Since we were interested in the effect of chord duration for the Leman TC and the possible difference of the effect on chord-type oddballs vs. voicing oddballs, we created a subtraction value (chord-type TC minus voicing TC; hereafter, t_TC_minus_v_TC) for each of the global-decay and local-decay combinations.

We assumed that duration had a greater effect on chord-type oddballs TC than voicing oddballs TC and we wanted to see whether such an effect was equally predicted by the different local and global decays we chose. To compare the duration conditions, we calculated three sets of values by subtracting the shorter-duration t_TC_minus_v_TC from the longer-duration ones. For instance, if for a given combination of local and global decay from the Leman's model, t_TC and v_TC were .90 and .91 for the 500 ms condition and .85 and .88 for the 1500 ms condition, then t_TC_minus_v_TC would be -.01 for the 500 ms condition and -.03 for the 1500 ms condition, and the difference between the 1500 and 500 ms conditions would be -.03 minus -.01. This example resembles participants' responses, since duration has a stronger effect on t than v (the difference between t_TC and v_TC increases with duration), predicting that the salience of t over v will increase when chords are longer.

Figure 4 shows by colors the differences between longer and shorter duration conditions in terms of t_TC_minus_v_TC for each of the 75 combinations of Leman's local and global decay parameters (the exact values are not shown). Green colors indicate results that support our assumption (duration has stronger effect on chord-type oddball than on voicing oddball), and red color shows the opposite effect, darker shade indicating stronger effect. The effect is strongest (darkest green shading in Figure 4) with the biggest duration difference (1500 minus 500 ms), and weakest (or opposite) for the difference between the two longest durations (1500 minus 1000 ms). The pattern is consistent with the effect of duration on participant responses, since the effect of duration on chord-type responses was the largest for the longest duration. Results that contradict our assumption (red color) are relatively few and only generated for combinations of short local and global decays.

The prevalence of a larger effect of duration on chord-type than voicing changes within Leman's model (green cells in figure 4), can be explained by the combination of two factors. First, TC for our voicing oddballs tends to be higher (less harmonic contrast) than our chord-type oddballs even for the shortest duration. Second, a given change in duration tends to affect TC in different ways depending on the starting value of TC. TC values that start very high only become slightly lower with increased duration, whereas TC values that start lower experience a larger drop.

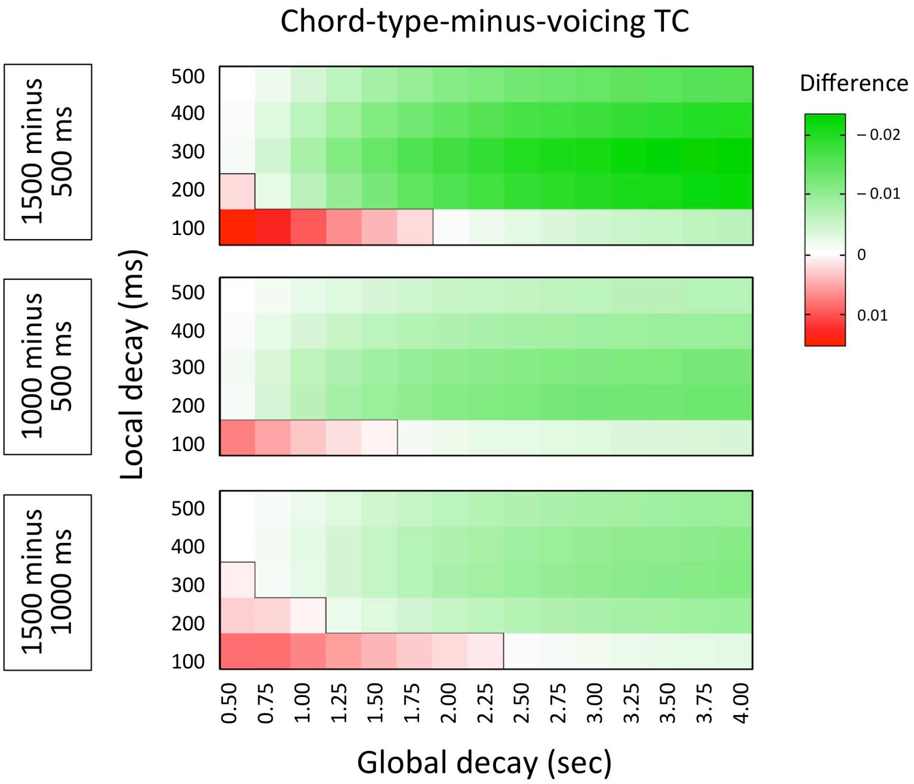


Figure 4. Differences between Duration Conditions in Terms of Chord-type-minus-voicing Tonal Contextuality.
Green indicates results that are in line with our participants responses, while red indicates the opposite results, the line separating the areas. The darker the shade of the color, the stronger the effect.

Discussion

The current study used a recently developed two-oddball paradigm to investigate the effect of duration on the relative salience of chord-type and voicing changes. Three duration conditions (500, 1000, and 1500 ms) were tested. As hypothesized, the percentage of chord-type responses increased with duration while the percentage of voicing responses did not. The three participant groups did not overlap: The percentage of chord-type responses was highest among *concept-oriented* participants and lowest among *listening-oriented* participants, the playing-oriented participants being in-between. However, the relative effect of duration on the percentage of chord-type responses was the same (approximately 10 percentage points from the duration 500 ms to the duration 1500 ms) in all participant groups. This finding suggests that chord-type changes become perceptually clearer with longer durations even when listeners are not able or are not trying to consciously identify chord type.

Concept-oriented participants used chord-types as a guiding factor

Even though our instructions did not use any chord labels nor chord-type terminology, the *concept-oriented* participants used words related to chord-types to indicate that they had focused on chord-type, which is in line with their music-education background (they also scored high in the chord-type post-test). Being able to easily identify chord types increased the likelihood of becoming consciously aware of the fact that one chord-type oddball was included in each stimulus. This was probably the case even when participants were trying not to analyze the stimuli. Additionally, conscious awareness of the specific chord types involved in the chord-type changes and the structure of the items could have increased the likelihood of focusing on the chord-type oddballs instead of any other changes. The awareness of the structure of the items can even have changed the bottom-up process of listening into a top-down process guided by the chord-types.

The *concept-oriented* participants also had more years of music theory training on average, and they had probably learned that a chord-type change (but not a voicing change) constitutes a “chord change”. The average of chord-type responses was 56.7% in the *concept-oriented* group, but the range was wide (from 15.0% to 100%). The wide range of chord-type responses in the main test could indicate that they had used two different strategies: While some of them may have used chord types as the only criterion and occasionally mistakenly chosen a voicing oddball, others may have compared the two changes and chosen the more salient one without assuming that the voicing oddballs were just distractors. Taken together, the distribution of responses means that responses from the *concept-oriented* participants cannot be fully explained by top-down processes.

The effect of duration

Tonal contextuality. We selected Leman's model because, to our knowledge, it is currently the only model of pitch perception that takes duration into account and can be used to analyze voicing changes. Because Leman's model uses raw audio as input, it can consider very fine-grained spectral information including the changes of register involved in our voicing changes. Our analysis showed a clear tendency for participants to choose chords that had lower TC more often, suggesting a strong link between TC and perceived chord (dis-)similarity. Further, although both chord-type and voicing oddballs had lower TC (indicating lower spectral similarity) for longer durations, the effect of duration was usually greater for chord-type than voicing oddballs, mirroring the effect of duration on participant responses. The success of Leman's model in predicting the effect of duration on participants responses suggests that sensory factors in general, and echoic memory in particular (represented in Leman's model by the global decay), play a role in chord perception.

Previous research had found that the predictions of Leman's model could be affected by the selection of local and global decay (Bigand et al., 2014). For this reason, we decided to use a wide array of combinations of local and global decays. In a few instances, combinations of the shortest local and global decays generated TC values that were inconsistent with the effect of duration on participant responses (Figure 4). Appendix A shows that the combinations of shortest local and global decays also have the weakest correlations between participant responses and TC for the difference between chord-type and voicing oddballs.

The relatively weak performance of a local decay of 100 ms combined with a global decay of 1.5 sec is particularly surprising since these decays performed better than other decays in Lemans' original simulations of probe tone experiments (2001) and have been the decays most often used in subsequent implementations of Leman's model (e.g., Koelsch et al., 2007; Goldman et al., 2021; Marmel et al., 2010). To gain some insight into this somewhat surprising result, we took a closer look at the effects of local and global decay on TC and the fluctuations of TC in two 1500 ms stimuli. Appendix B shows the fine-grained fluctuations (approximately 150 time steps per chord). Fluctuation of TC is greatest when the local decay of 100 ms is used. With the shortest global and local decays (100 ms and 0.5 sec respectively), the fluctuation of TC closely resembles the inverted amplitude envelope of the noise component in the piano tone (i.e., component other than the fundamental frequency and partial components; e.g., Lehtonen, 2010, p. 33). As such, it suggests a strong (negative) correlation between TC for shortest decays and the timbrally complex noise component but shows little differentiation between the chords within the item. However, when longer local and global decays are used, the differentiation between oddballs and standard chords becomes clearer. We observed a similar pattern in the rest of the stimuli, suggesting that in longer local and global decays can capture a type of TC in our stimuli that better correlates with the differentiation between oddball and standards. Our results further suggest that temporal-spectral characteristics of the tones may affect the extent to which global and local decays generate predictions consistent with behavioral tasks. The average duration of echoic memory and the extent to which local auditory images are smoothed by the peripheral auditory system may explain the role of temporal-spectral characteristics in the underperformance of short decays in our analyses. However, the current uncertainty about the exact duration and functioning of echoic memory and our lack of knowledge regarding the potential smoothening of auditory images in the peripheral auditory system prevents us from further speculations. Despite the open-ended nature of this type of analysis, we believe that the current observations could provide a particularly interesting point of reference for the use of the model in future research.

Other potential factors contributing to the effect of duration. In addition to the above-discussed factors, there are other potential factors that could have contributed to this effect. As previously stated, conscious labelling of chords takes at least 800 ms which explains why chord-type responses increased with duration for participants in the *concept-oriented* group. Yet, it cannot explain the increase of chord-type responses for the *playing-oriented* and *listening-oriented* participants. It is possible that even though the brain can detect the structural difference between major and minor chords very quickly, chord-type changes become perceptually clearer with longer durations even for listeners who are not able or are not trying to consciously identify chord types. Since the total share of chord-type responses was far less than 40% even when the duration was 1500 ms, it is clear that the *playing-oriented* and *listening-oriented* participants did not focus on the chord-type quality alone. When asked what aspects of the chords they had paid attention to during the experiment, *playing-oriented* and *listening-oriented* participants provided responses suggesting a focus on individual pitches (e.g., responses that included the terms “note,” “sharp,” “accidental”) or voices (e.g., responses that included the terms “melody,” “tune”) as opposed to a focus on the chord-type quality of each chord. Although some *playing-oriented* and *listening-oriented* participants may have focused on the chord-type quality and simply lacked the technical terms to describe such focus, clearly not all participants had such focus. In what follows, we offer two tentative hypotheses that could explain the effect of duration in cases in which participants' focus was not the chord-type quality of the individual sonorities.

> *1. Individual pitches:* There is evidence that the perception of pitch chroma requires longer pitches than the perception of pitch register (Robinson & Patterson, 1995). Since chord-type oddballs involved changes of pitch chroma and voicing oddballs involved changes of pitch register in our experiment, short chords may have been enough for the participants to fully perceive the pitch changes involved in voicing oddball, but not to perceive pitch-chroma changes in chord-type oddballs. This phenomenon could explain the effect of duration not only when participants focused on individual pitches but also when they tried to focus on the chord-type quality. However, the plausibility of the effect of duration on the perception of pitch chroma in our experiment is hard to assess since Robison and Patterson (1995) only tested durations that were shorter than the pitches in our 500 ms condition and did not test the perception of pitches in the context of chords.
>
> *2. Focus on melodic changes:* In our experiments, the highest pitch was kept constant in each item to diminish the effect of melody for the evaluations. However, the participants could have heard some pitch changes in the inner pitches of the items as creating melodic gestures, and they could have heard the melodic gestures involved in voicing oddballs as more salient than the melodic gestures involved in the chord-type oddballs because the former always created

larger melodic intervals than the latter.[7] If participants sometimes heard the pitch changes in the sequences of chords in terms of melodic gestures, the effect of duration on chord-type responses could be explained as “melodic” hearing that is facilitated by shorter IOIs. Although there is some evidence that shorter IOIs facilitate the perception of successive tones as a continuous auditory stream, this phenomenon is greatly influenced by interval size (Van Noorden, 1975) and has not been tested in the context of block chords. Additionally, only two participants mentioned the term “tune,” only one participant mentioned the term “melody,” and no participant reported shifting their attention from melodies or tunes for shorter chords to attending to the vertical sonorities for longer chords. Further research is needed to study the role of melodic changes in chord progressions.

Even though the above-mentioned ways of listening to the chord stimuli were possible, participants’ own accounts of their focus during the experiment did not suggest that they focused on melody or scale membership. The above-listed reports of alternative ways of listening to the stimuli were few, and we had also designed the experiment taking measures to decrease the likelihood of participants adopting such modes of listening. We made the highest and lowest notes on each chord louder than the other chord tones to partially mask the pitch changes with the goal of promoting the hearing of chords as unified events as opposed to bundles of frequencies or vertical slices of a multi-voice texture. Although the 200 ms silences that separated chords within each item are not long enough to completely prevent the formation of auditory streams (Van Noorden, 1975), these silences are likely to have further

---

[7] All the pitch changes in the voicing oddballs are melodic leaps. Although these pitch changes could be understood as registral shifts (e.g., D5 to D3 in the first example of Figure 1), they are more likely to be heard as belonging to one of the three voices formed by the second, third, and fourth highest pitches from each chord respectively (e.g., D5 to G4 in the first example of Figure 1).

facilitated the hearing of each chord as separated events. Additionally, our instructions and visual aids only refer to individual chords as single unified events.

For cases in which *playing-oriented* and *listening-oriented* listeners seemed to have focused on chord-type changes, the effect of duration on chord-type responses could relate to two potentially complementary phenomena pertaining to the perception of pitch and timbre. In our experiment, variations of voicing and chord type meant that each chord had a unique set of timbral features. Timbral brightness has been shown to be the most perceptually salient feature of timbre in general (Caclin et al., 2005; Grey & Gordon, 1978; Iverson & Krumhansl, 1993; Lakatos, 2000; McAdams et al., 1995). This has also been shown in the specific case of piano block chords (Jimenez et al., 2023). One of the most common correlates of perceived timbral brightness is the spectral centroid. In our stimuli set, the average change of the spectral centroid in the voicing-oddball items was 7.58%, and it was above the just-noticeable threshold (4% for musicians and 5% for nonmusicians) for the perception of the spectral centroid changes (Allen & Oxenham, 2014). The average change of the spectral centroid in chord-type oddball items was 1.08%, far below the just-noticeable threshold. From the perspective of timbral brightness, we argue that voicing changes affected timbre more than chord-type changes in our stimuli, especially since voicing oddballs included register changes. Timbral features tend to be perceived faster than pitch (Robison & Patterson, 1995) which could explain why voicing oddballs seemed to have been more easily perceived as different than chord-type oddballs in the 500 ms condition by the *playing-oriented* and *listening-oriented* participants.

The second perceptual phenomenon relates to a time-dependent shift in the way tones are mentally represented. There is some evidence that mental representation of tones tends to focus on spectra for sounds nearby in time but on pitch chroma for sounds separated by silence (McPherson & McDermott, 2020). Although chords in our experiment were separated by a 200 ms silent interval in all three duration conditions, the time separation between the amplitude peaks of the chords increased with IOIs and these amplitude peaks were immediately followed by a sharp drop of amplitude (see Figure S1 in supplementary material 3). A shift from spectra (e.g., the spectral centroid) to pitch chroma may explain the increase of chord-type responses with longer durations. Yet, once more, the utility of previous research to support this hypothesis is limited (McPherson & McDermott, 2020 did not test the perception of tones in the context of chords).

Relationship to previous research: An effect of chord duration on harmonic perception has been found in previous research. Tonal priming experiments by Tekman and Bharucha (1998) and Bigand et al. (2003) suggest that cognitive mechanisms predominate over sensory ones in listening to chords except when the chords change extremely fast (e.g., chord durations of 75 ms). Two decades later, our study suggests that sensory mechanisms, represented by the output of Leman's model can affect chord perception for durations well above the 75 ms, thereby better corresponding to the duration of chords in Western popular music which tends to be in the range of 1–3 seconds (de Clercq, 2022).

The types of paradigms in earlier research and the current one are very different. Tonal priming participants are asked to tell whether the last chord in a pair or longer sequence is out of tune or dissonant, and the priming occurs as the participants' implicit or explicit expectations affect the accuracy and response times. Further, Tekman and Bharucha (1998) compared two pairs of chords, C - D vs. C - E while Bigand et al. (2003) used chord progressions that followed the general conventions of Western tonal music with successive chords always having different roots and basses. This was not the case in our study, in which neither the nature of the experimental task nor the type of stimuli encouraged the participants to use schematic knowledge about conventional chord progressions.

We believe that our experiment was able to detect an effect of duration on chord perception that has not earlier been detected. This improved understanding of the effect of duration on chord perception can contribute to making better informed decisions when creating chordal stimuli in harmonic studies and it also suggests that the large range of chord durations found in tonal music (e.g., de Clercq, 2022; Jimenez et al., 2023), has implications for harmonic perception. This, in turn, contributes to the critical, integrated understanding of perception of harmony vis-à-vis extra-harmonic features of music. Additionally, our analysis using Leman's model suggested that the effect of duration in our experiment was influenced by the chords' immediate context, but that this influence was predominantly sensory in nature and did not depend on harmonic schemata.

Future research could explore whether adapting our oddball paradigm to more conventional chord progressions could be used to investigate schematic expectations and their potential interaction with TC. Another option would be to use Leman's model and the current insight for developing stimuli (and hypotheses) for analyzing how sensory (bottom-up) vs. cognitive (top-down) components affect veridical and schematic memory for harmony e.g., in melody-and-accompaniment textures (as in Jimenez et al., 2022b).

References

Ahler, D. J., Roush, C. E., & Sood, G. (2019). The micro-task market for lemons: Data quality on Amazon's Mechanical Turk. *Political Science Research and Methods*, 1–20. https://doi.org/10.1017/psrm.2021.57

Allen, E. J., & Oxenham, A. J. (2014). Symmetric interactions and interference between pitch and timbre. *The Journal of the Acoustical Society of America, 135*(3), 1371–1379. https://doi.org/10.1121/1.4863269

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, 1–10. https://doi.org/10.3389/fpsyg.2019.02883

Bharucha, J. J., & Stoeckig, K. (1986). Reaction time and musical expectancy: priming of chords. *Journal of Experimental Psychology: Human Perception and Performance, 12*(4), 403–410.

Bigand, E., Delbé, C., Poulin-Charronnat, B., Leman, M., & Tillmann, B. (2014). Empirical evidence for musical syntax processing? Computer simulations reveal the contribution of auditory short-term memory. *Frontiers in Systems Neuroscience*, *8*, 94, 1–27. https://doi.org/10.3389/fnsys.2014.00094

Bigand, E., Madurell, F., Tillmann, B., & Pineau, M. (1999). Effect of global structure and temporal organization on chord processing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 184–197. https://doi.org/10.1037/0096-1523.25.1.184

Bigand, E., Poulin, B., Tillmann, B., Madurell, F., & D'Adamo, D. A. (2003). Sensory versus cognitive components in harmonic priming. *Journal of*

Experimental Psychology: Human Perception and Performance, *29*(1), 159–171. https://doi.org/10.1037/0096-1523.29.1.159

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/10.1177/0894439303021002

Burgoyne, J. A., Wild, J., & Fujinaga, I. (2011, October). An Expert Ground Truth Set for Audio Chord Recognition and Music Analysis. In *ISMIR* (Vol. 11, pp. 633–638).

Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *The Journal of the Acoustical Society of America, 118*(1), 471–482. https://doi.org/10.1121/1.1929229

de Clercq, T. (2022). Tempo versus average rates of harmonic rhythm in popular music: A study of three corpora. *Musicae Scientiae*, *27*(3), 521–540. https://doi.org/10.1177/10298649221091483

Deliège, I., Mélen, M., Stammers, D., & Cross, I. (1996). Musical schemata in real-time listening to a piece of music. *Music Perception 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, February). Demographics and dynamics of mechanical turk workers. In *Proceedings of the eleventh ACM* *international conference on web search and data mining* (pp. 135–143). https://doi.org/10.1145/3159652.3159661

Dunmire, C. E. (2019). *Audiating and vocalizing voice-leading in seventh chords: Harmonic recognition in freshman ear-training courses*. Liberty University.

Goldman, A., Harrison, P. M., Jackson, T., & Pearce, M. T. (2021). Reassessing syntax-related ERP components using popular music chord sequences: A model-based approach. *Music Perception 39*(2), 118–144. https://doi.org/10.1525/mp.2021.39.2.118

Grey, J. M., & Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. The *Journal of the Acoustical Society of America, 63*(5), 1493–1500. https://doi.org/10.1121/1.381843

Huron, D., & Parncutt, R. (1993). An improved model of tonality perception incorporating pitch salience and echoic memory. *Psychomusicology: A Journal of Research in Music Cognition*, *12*(2), 154–171. https://doi.org/10.1037/h0094110

Iverson, P., & Krumhansl, C. L. (1993). Isolating the dynamic attributes of musical timbre. *The Journal of the Acoustical Society of America, 94*(5), 2595–2603. https://doi.org/10.1121/1.407371

Jimenez, I., Kuusi, T., & Ojala, J. (2022a). Relative salience of chord-type and chord-voicing changes: A two-oddball paradigm. *Psychology of Music*, *50*(5), 1566–1585. https://doi.org/10.1177/03057356211055214.

Jimenez, I., Kuusi, T., & Ojala, J. (2022b). Veridical and schematic memory for harmony in melody-and-accompaniment textures. *Music Perception, 40*(2), 89–111. https://doi.org/10.1525/mp.2022.40.2.89

Jimenez, I., Kuusi, T., Czedik-Eysenberg, I., & Reuter, C. (2023). Identifying songs from their piano-driven opening chords. *Musicae Scientiae*, 27(1), 27–52 https://doi.org/ 10.1177/10298649211003631

Koelsch, S., Jentschke, S., Sammler, D., & Mietchen, D. (2007). Untangling syntactic and sensory processing: An ERP study of music perception. *Psychophysiology*, *44*(3), 476–490. https://doi.org/10.1111/j.1469-8986.2007.00517.x

Lakatos, S. (2000). A common perceptual space for harmonic and percussive timbres. *Perception & Psychophysics, 62*(7), 1426–1439. https://doi.org/10.3758/BF03212144

Lehtonen, H. M. (2010). Analysis, perception, and synthesis of the piano sound. *PhD thesis, Aalto University*.

Lerdahl, F., & Krumhansl, C. L. (2007). Modeling tonal tension. *Music perception*, *24*(4), 329–366. https://doi.org/10.1525/mp.2007.24.4.329

Leman, M. (2000). An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception*, *17*(4), 481–509. https://doi.org/10.2307/40285830

Marmel, F., Tillmann, B., & Delbé, C. (2010). Priming in melody perception: tracking down the strength of cognitive expectations. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(4), 1016–1028. https://doi.org/10.1037/a0018735

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research, 58*(3), 177–192. https://doi.org/10.1007/BF00419633

McPherson, M. J., & McDermott, J. H. (2020). Time-dependent discrimination advantages for harmonic sounds suggest efficient coding for memory.

Proceedings of the National Academy of Sciences, *117*(50), 32169–32180. https://doi.org/10.1073/pnas.2008956117

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). Measuring the facets of musicality: The Goldsmiths Musical Sophistication Index (Gold-MSI). *Personality and Individual Differences*, *60*, S35. https://doi.org/10.1016/j.paid.2013.07.081

Nees, M. A. (2016). Have we forgotten auditory sensory memory? Retention intervals in studies of nonverbal auditory working memory. *Frontiers in Psychology, 7*, 1892, 1–6. https://doi.org/10.3389/fpsyg.2016.01892

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*(2), 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*(2), 217–226. https://doi.org/10.3758/BF03202544

Pearce, M.T. (2018). Statistical learning and probabilistic prediction in music cognition: mechanisms of stylistic enculturation. Annals of the New York Academy of Sciences. https://doi.org/10.1111/nyas.13654.

Pechmann, T. (1998). Memory for chords: The retention of pitch and mode. *Music Perception*, *16*(1), 43–54. https://doi.org/10.2307/40285776

Robinson, K., & Patterson, R. D. (1995). The duration required to identify the instrument, the octave, or the pitch chroma of a musical note. *Music Perception*, *13*(1), 1–15.

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

Skorik, S. O., Almayev, N. A., & Bessonova, Y. V. (2018). Towards psychophysiological mechanisms of major and minor chords perception. *International Journal of Psychophysiology, 131*, S163. https://doi.org/10.1016/j.ijpsycho.2018.07.432

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior research methods*, *42*(4), 1096–1104. https://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

Taguti, T., Ohtsuki, K., Yamasaki, T., Kuwano, S., & Namba, S. (2002). Quality of piano tones under different tone stoppings. *Acoustical Science and Technology*, *23*(5), 244–251. https://doi.org/10.1250/ast.23.244

Tekman, H. G., & Bharucha, J. J. (1992). Time course of chord priming. *Perception & Psychophysics, 51*(1), 33–39. https://doi.org/10.3758/BF03205071

Tekman, H. G., & Bharucha, J. J. (1998). Implicit knowledge versus psychoacoustic similarity in priming of chords. *Journal of Experimental Psychology: Human Perception and Performance, 24*(1), 252–260. https://doi.org/10.1037/0096-1523.24.1.252

Tillmann, B., & Bigand, E. (2004). Further investigation of harmonic priming in long contexts using musical timbre as surface marker to control for temporal effects.

Perceptual and Motor Skills, 98(2), 450–458. https://doi.org/10.2466/pms.98.2.450-458

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

Van Noorden, L. S. (1975). Temporal coherence in the perception of tone sequences. *PhD thesis, Eindhoven University of Technology*.

Virtala, P., Berg, V., Kivioja, M., Purhonen, J., Salmenkivi, M., Paavilainen, P., & Tervaniemi, M. (2011). The preattentive processing of major vs. minor chords in the human brain: an event-related potential study. *Neuroscience Letters, 487*(3), 406–410. https://doi.org/10.1016/j.neulet.2010.10.066

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

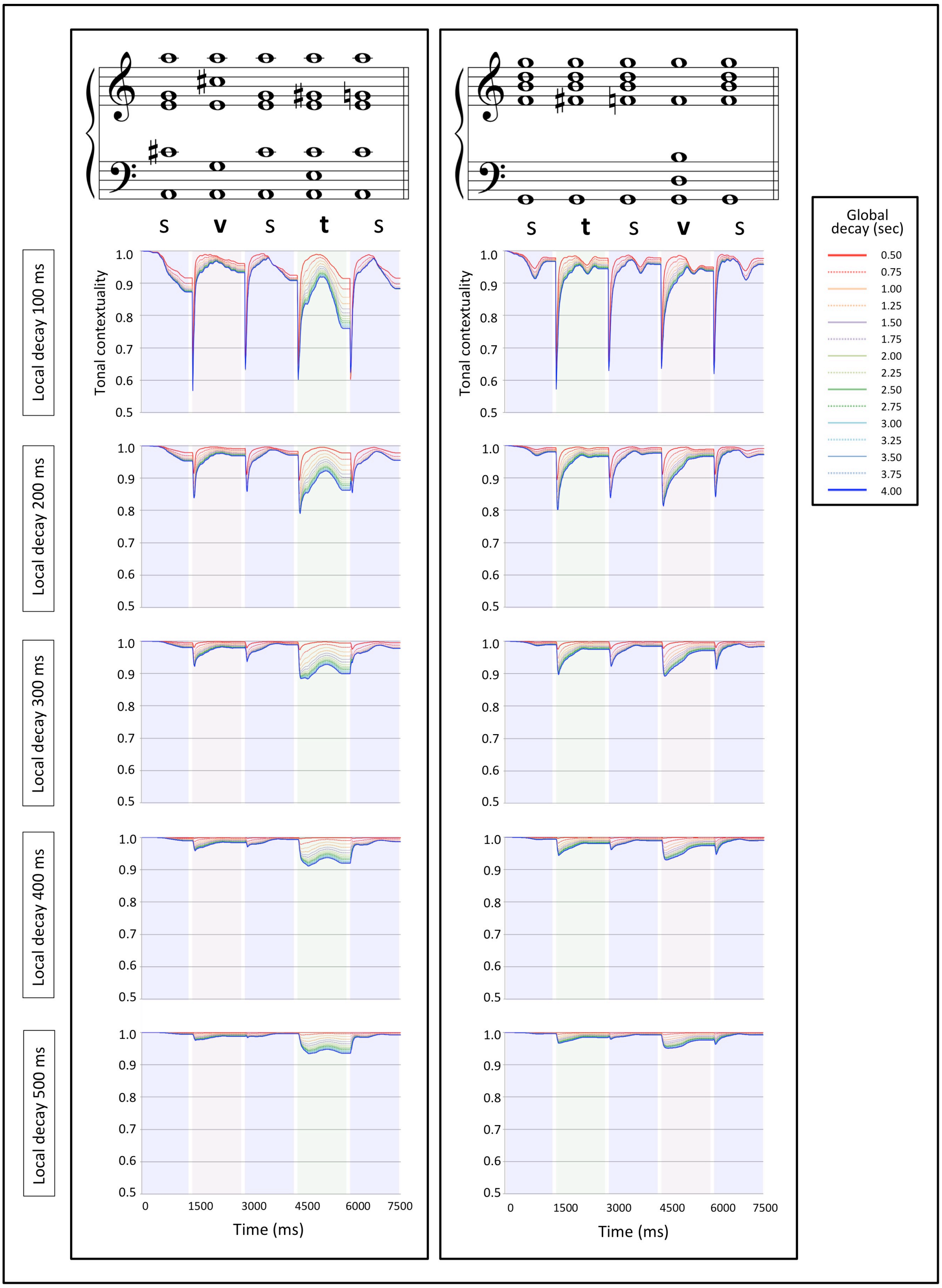
Appendix A

Local decay (ms)															
500		-0.81	-0.80	-0.80	-0.80	-0.79	-0.79	-0.79	-0.79	-0.79	-0.79	-0.79	-0.79	-0.79	-0.79
400	-0.80	-0.81	-0.81	-0.81	-0.81	-0.81	-0.80	-0.80	-0.80	-0.80	-0.80	-0.80	-0.80	-0.80	-0.80
300	-0.78	-0.80	-0.81	-0.81	-0.82	-0.82	-0.82	-0.81	-0.81	-0.81	-0.81	-0.81	-0.81	-0.81	-0.81
200	-0.74	-0.78	-0.80	-0.81	-0.82	-0.82	-0.82	-0.82	-0.82	-0.82	-0.82	-0.82	-0.82	-0.82	-0.82
100	-0.67	-0.72	-0.75	-0.77	-0.78	-0.79	-0.80	-0.80	-0.80	-0.80	-0.81	-0.81	-0.81	-0.81	-0.81
Global decay (sec)	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50	3.75	4.00

Appendix A. Correlations between Averaged Participants Responses and Averaged Tonal Contextuality for the Difference between Chord-type and Voicing oddballs.

Tonal contextuality for each oddball chord was calculated separately for every combination of local and global decays.

Appendix B



Appendix B. Tonal Contextuality for Two of the Items Used in the 1500-ms Condition.
s = standard chords (lavender shading), t = chord-type oddball (green shading), and v = voicing oddball (pink shading).

The Effect of Chord Duration on the Relative Salience of Chord-type and Voicing changes

Supplementary materials

Supplementary material 1. Online experiment

Supplementary material 2. Participant background variables

Supplementary material 3. Timbre and amplitude envelope.

Supplementary material 4. Additional items, training sessions and details of the experiment.

Supplementary material 1.

Online experiment

The online experiment was visited 1,016 times between February 26 and May 5, 2022. Since we knew that the amount of non-serious visitors and survey bots is large in crowdsourcing platforms (Ahler et al., 2019; Dennis et al., 2020), we used a pre-test to screen participants. In the three-trial pre-test, the visitors were to choose the loudest tone in a series of five piano tones. The difficulty of the loudness pre-test was set relatively high to ensure that all successful participants could hear the stimuli very well, hence minimizing the influence of the quality of participants' headphones, the environmental noise, and participants' hearing deficiencies like hearing loss. Altogether 124 visitors abandoned the survey before taking the pre-test, 2 visitors abandoned the pre-test while taking it, and 624 visitors were not allowed to take the survey because they failed to answer the pre-test correctly. Further, we used five criteria for recognizing and rejecting 91 visitors who completed or were likely to have completed some parts of the experiment without actually listening to the stimuli or with the help of autofillers or bots.[8] Of the 175 who remained, 115 (65.7%) completed the experiment. This completion rate is approximately the same as in online experiments using participants with high internal motivation (Bosnjak & Tuten, 2003; Tuten et al., 2004) and clearly higher than in some other studies (O'Neil & Penrod, 2001; O'Neil et al., 2003). Further, it should be noted that the completion

---

[8] The participants were rejected if they a) responded before listening to the whole stimulus in more than 30% of the main trials; b) provided incorrect responses for half or more of the 6 control stimuli; c) provided likely automatic responses to the open-ended questions (e.g., nonsensical or extremely repetitive responses); d) responded very differently in the last 75 trials than in the first 75 trials; e) provided responses too close to random distribution (approximately 20% of each of the five chords).

rate does not include those who completed the experiment but were rejected because of our inclusion criteria.

Supplementary material 2.

Participant background variables

We ran a one-way ANOVA to analyse ~~if~~ whether the three participant groups differed in terms of the participant background variables that were collected by the questionnaire. The variables are listed in Table S1 and organised according to how many participant groups differed in terms of the variable.[9] There is a clear connection between some background variables and the grouping decisions; for example, since instrument playing differentiated the *listening-oriented* participants from the other participants, it is understandable that *listening-oriented* group differs from the other groups in terms of all variables that are related to playing an instrument (V4, V5, V11, V12, V14). Further, the active and conscious use of chord types and chord-type labels during the experiment differentiated the *concept-oriented* from the *playing-oriented* group. It should, however, be noted that not using the chord-type concepts during the experiment does not necessarily indicate the lack of knowledge of such concepts. Actually, we can see that the *concept-oriented* and *playing-oriented* group differ in terms of V1 and V2 (chord-type identification) and V7 (formal music-theory studies), but not in terms of other variables related to chord-types V3, V4, and V6. Concepts and labels related to chord-types are often taught in formal music education, but they can also be learned in action of playing music from notation (especially lead sheets). It is possible that the ability to actively use the concepts in a quickly moving experiment needs thorough and multifaceted knowledge gained both by formal education and active playing.

---

[9] We used the Bonferrroni post-hoc test for variables that had equal variances and Tamhane post-hoc test for variables that had unequal variances.

Table S1. Results of the one-way ANOVA analysis. Participant Background Variables and Differences Between Groups. The p-values Indicating Statistically Significant Differences are in Bold Print.

Variable code	Explanation	Post-hoc test	*c-o* vs *pl-o* [a]	*c-o* vs *l-o*	*pl-o* vs *l-o*
A. Differences between all three participant groups					
V1	self-reported ability to identify major and minor chords by ear [b]	Tamhane	**< .001**	**.001**	**< .001**
V2	chord-type identification post-test	Bonferroni	**< .001**	**.001**	**< .001**
V11	years of playing main instrument	Tamhane	**.020**	**.001**	**< .001**
V14	GoldMSI Factor 3 (Musical training)	Bonferroni	**.001**	**.001**	**< .001**
V15	GoldMSI Factor 4 (Singing abilities)	Bonferroni	**.011**	**.001**	**.007**
B. Differences between c-o group and the other two groups					
V7	years of music-theory studies	Tamhane	**.040**	**.021**	.052
C. Differences between c-o group and l-o group					
V8	years of aural-skills studies	Tamhane	.062	**. 034**	.362
V9	number of composed pieces	Tamhane	.216	**. 019**	.198
V10	number of arranged pieces	Tamhane	.058	**. 007**	.239
D. Differences between l-o group and the other two groups					
V4	total hours of playing chords by ear [c]	Tamhane	.144	**.040**	**.005**
V5	total hours of playing music from notation [c]	Tamhane	.631	**< .001**	**.014**
V6	attention to chords when listening to music in everyday life	Bonferroni	1.000	**.010**	**.001**
V12	GoldMSI Factor 1 (Active engagement)	Bonferroni	.499	**< .001**	**.011**
V13	GoldMSI Factor 2 (Perceptual abilities)	Bonferroni	.094	**< .001**	**< .001**
V16	GoldMSI Factor 5 (Emotions)	Bonferroni	.499	**< .001**	**.035**
E. No difference between participant groups					
V3	years of ear-training of chords and chord-progressions	Tamhane	.238	.094	.494
V17	age	Bonferroni	1.000	.143	.111

[a] *C-o* = concept-oriented, *pl-o* = playing-oriented, *l-o* = listening oriented.

[b] 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, and (7) I have never heard those terms before.

[c] To obtain an estimate of total hours, we asked participants to estimate the approximate number of years and average hours per week.

Supplementary material 3.

Timbre and amplitude envelope.

For a balance between (a) how naturalistic the piano tones sounded and (b) how the spectral and spectro-temporal differences between the three duration conditions could be minimized, 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 chord was set to 2000 ms.

4) Each 2000-ms chord was exported from Logic Pro X as an individual audio WAV file.

5) A ramp was applied starting at 120 ms and ending at 300, 800, or 1300 ms (see Figure S1). This meant that the first 120 ms of chords from all three chord durations were identical. The amplitude of the first 120 ms was not modified because all the spectral components of the original 2000-ms piano tones reached their peak amplitude within the first 120 ms.

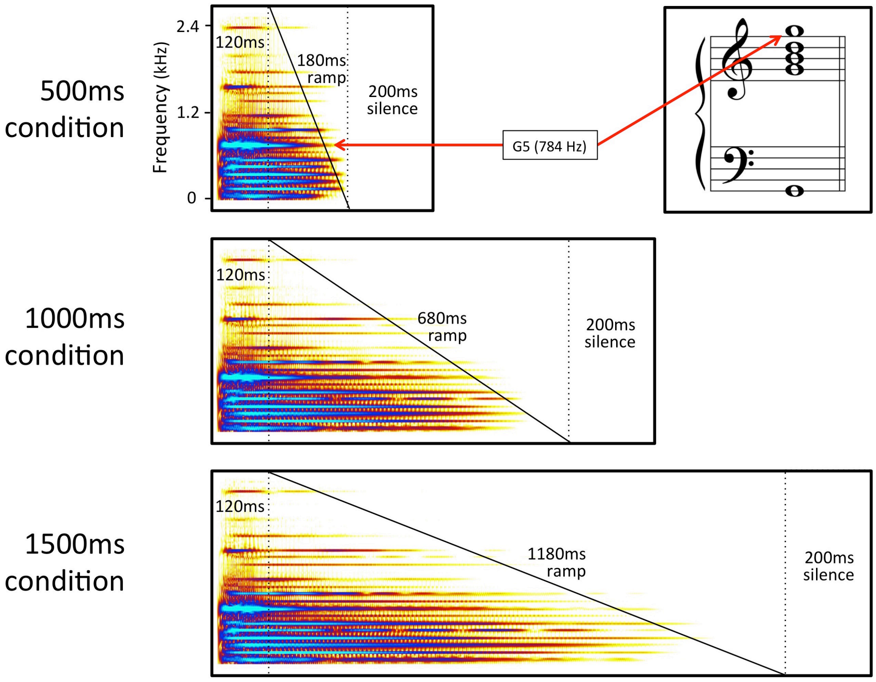


Figure S1. Ramps for the Three Chord-duration Conditions.
Time on the x-axis and frequency on the y-axis. The sonograms illustrate the typical time-variance of spectral energy in the chords used in our experiment. Light blue indicates the highest levels of spectral energy. Peak amplitudes for all the spectral components are reached within the first 120 ms.

Our assessment that the resulting block piano chords sounded naturalistic is consistent with Taguti et al. (2020). They asked participants to rate sequences of 250-ms piano tones that had different types of digitally-edited amplitude envelopes. Ratings for tone reverberation, beauty, and luster were highest when the amplitude plateau was – as in our block chords – shorter than the ramp.

Supplementary material 4.

Additional items, training sessions and details of the experiment.

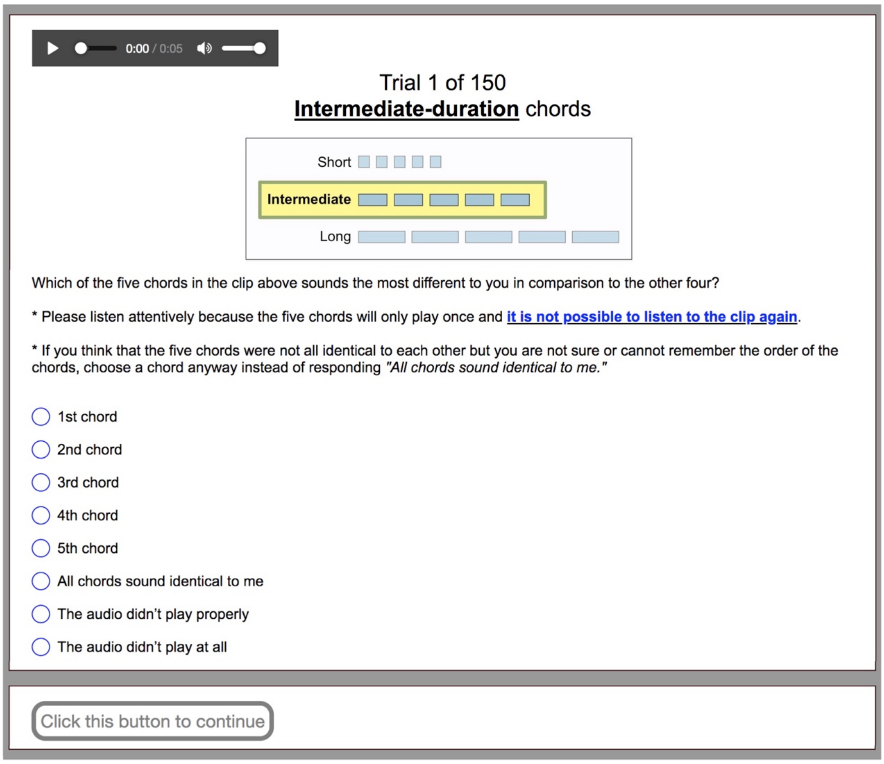
In addition to the main 120 items, we prepared 9 training items, 24 items with either oddball in serial position 5, and 6 control items, but none of these items were included in the analysis. In each of these three additional groups of items, there were an equal number of items corresponding to the three duration conditions (3 training items, 8 items with an oddball in serial position 5, and 2 control items). The goal of having the items with an oddball falling in serial position 5 was to minimize the likelihood that participants will assume that oddballs in our experiment never occurred in that serial position. Control items were used to verify that participants kept the same level of seriousness throughout the experiment. Oddballs in the control items were made easier to detect than in the rest of trials by having four standard chords and only one oddball chord which always occurred in serial position 5. Additionally, the chord-type pairs and voicings used in the control items were those in which the oddballs were easiest to detect (Jimenez et al., 2022).

There were 6 compulsory training trials and 3 additional voluntary training trials. Before starting the training trials, participants were given the following general instructions:

> "*This experiment has a total of 150 trials. In these trials we will ask you to respond based on your first impression. You do not need to analyze the chords. Try to avoid spending too much time on each trial; otherwise the experiment will become unnecessarily long for you. Please click the button below to start.*"

The main purpose of the training trials was to familiarize participants with the three duration conditions and the way participants are informed about duration condition. For all trials, including the training trials, participants were informed whether the chords will be "short", "long" or "intermediate" by the heading of the screen and also visually by a diagram that represents the proportions of the three chord durations and highlights the duration of the current item by means of a yellow box (see Figure S2).

Figure S2 is an example of how each item was presented on the screen to the participants. The playback controls at the top of the page disappeared immediately after participants clicked on the playback button. 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.



Trial 1 of 150

Intermediate-duration chords

Which of the five chords in the clip above sounds the most different to you in comparison to the other four?

* Please listen attentively because the five chords will only play once and **it is not possible to listen to the clip again**.

* If you think that the five chords were not all identical to each other but you are not sure or cannot remember the order of the chords, choose a chord anyway instead of responding *"All chords sound identical to me."*

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

Figure S2. Screenshot of the Task as Presented to Participants.

References for supplementary materials

Ahler, D. J., Roush, C. E., & Sood, G. (2019). The micro-task market for lemons: Data quality on Amazon's Mechanical Turk. *Political Science Research and Methods*, 1–20. https://doi.org/10.1017/psrm.2021.57

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/10.1177/08944393030210020

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

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*(2), 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*(2), 217–226. https://doi.org/10.3758/BF03202544

Taguti, T., Ohtsuki, K., Yamasaki, T., Kuwano, S., & Namba, S. (2002). Quality of piano tones under different tone stoppings. *Acoustical Science and Technology*, *23*(5), 244–251. https://doi.org/10.1250/ast.23.244

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