Article:

Arthurs, Y, Beeston, AV orcid.org/0000-0003-2796-1947 and Timmers, R (2018) Perception of isolated chords: Examining frequency of occurrence, instrumental timbre, acoustic descriptors and musical training. Psychology of Music, 46 (5). pp. 662-681. ISSN 0305-7356

https://doi.org/10.1177/0305735617720834

Perception of isolated chords: Examining frequency of occurrence, instrumental timbre, acoustic descriptors and musical training

Journal Title
XX(X):2–24

Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Yuko Arthurs[1,2], Amy V Beeston[3] and Renee Timmers[2]

Abstract

This study investigated the perception of isolated chords using a combination of experimental manipulation and exploratory analysis. Twelve types of chord (five triads and seven tetrads) were presented in two instrumental timbres (piano and organ) to listeners who rated the chords for consonance, pleasantness, stability and relaxation. Listener ratings varied by chord, by timbre, and according to musical expertise, and revealed that musicians distinguish consonance from the other variables in a way that other listeners did not. To further explain the data, a principal component analysis and linear regression examined three potential predictors of the listener ratings. First, each chord's frequency of occurrence was obtained by counting its appearances in selected works of music. Second, listeners rated their familiarity with the instrumental timbre in which the chord was played. Third, chords were described using a set of acoustic features derived using the Timbre Toolbox and MIR Toolbox. Results of the study indicated that listeners' ratings of both consonance and stability were influenced by the degree of musical training and knowledge of tonal hierarchy. Listeners' ratings of pleasantness and relaxation, on the other hand, depended more on the instrumental timbre and other acoustic descriptions of the chord.

Keywords
Musical chords, consonance, frequency of occurrence, timbre familiarity, musical training

Introduction

A single chord has not only physical sonority, but also the power to refer the listener to an underlying tonal hierarchy. A single chord can play a significant role in listeners' experience of music, for example to create an expectation for an upcoming chord (Bharucha and Stoeckig, 1987), to generate a sense of key (Krumhansl, 1990), signify a variety of emotions (Lahdelma and Eerola, 2016), or to remind listeners of *un signifié* in the manner of a leitmotif, such as the Tristan chord (Nattiez, 1990). Repetition has been reported to reinforce a chord's importance within the structure of a musical work, and also in a listener's musical schema: a chord that appears frequently and at crucial musical moments becomes more important and stable within the underlying system of music (Krumhansl, 1990). Little is known at present, however, about whether factors such as frequency of occurrence and tonal hierarchy also play a role for chords presented in isolation, as they do for chords presented in more musical contexts. To address this point, the present study investigates the extent to which the perception of the consonance and dissonance of isolated chords relies on knowledge of tonal hierarchy (which governs a chord's frequency of occurrence), and the degree to which it depends on more direct acoustic features of the specific chord, or a listeners' familiarity with the particular instrument on which the chord is played.

Chord perception

According to Western classical music theory, chords are generally categorised as dissonant when they contain dissonant intervals (Piston, 1950), or have an added seventh (Sadai, 1980). For example, diminished and augmented triads are considered dissonant as

[1]School of Music and Performing Arts, Bath Spa University, UK
[2]Department of Music, The University of Sheffield, UK
[3]Department of Computer Science, The University of Sheffield, UK

Corresponding author:
Yuko Arthurs, School of Music and Performing Arts, Bath Spa University, Newton St Loe, Bath, BA2 9BN, UK.
Email: y.arthurs@bathspa.ac.uk

they contain a diminished and an augmented fifth, respectively. Numerous behavioural studies have shown that the most consonant triad is the major, followed by minor and diminished triads, while the augmented triad is the most dissonant (see e.g., Arthurs and Timmers, 2016; Cook and Fujisawa, 2006; Johnson-Laird et al., 2012; Roberts, 1986). The augmented major seventh is usually considered to be a particularly dissonant chord since it includes an augmented triad and a major seventh (Sadai, 1980). The dominant seventh tends to resolve onto the tonic, whereas other seventh chords such as the half-diminished seventh or diminished seventh are enharmonic and remain ambiguous in terms of their function and resolution (Sadai, 1980; Benward and Saker, 2003).

Psychoacoustic accounts have attempted to explain the consonance/dissonance (C/D) of individual intervals using terms such as roughness (e.g., Helmholtz, 1877/1954; Hutchinson and Knopoff, 1978, 1979) or harmonicity (e.g., Cousineau et al., 2012; McDermott et al., 2010). In comparison with the C/D of individual intervals, however, the C/D of chords is less well understood. Following Hutchinson and Knopoff (1978, 1979), roughness values have been computed to predict the C/D of intervals and chords (Sethares, 1998; Vassilakis, 2005; Vassilakis and Fitz, 2007), providing results which are, in the main, consistent with current music theory and existing behavioural data.

Chords can also induce perceptions of tension and instability in listeners. Roughness, in particular, has been reported to influence the perception of instability and tension (Lerdahl and Krumhansl, 2007): rough instrumental sounds triggered tension (Pressnitzer et al., 2000); and chords with high roughness values were perceived as more tense in a harmonic context (Bigand et al., 1996), and as more unstable even in atonal music (Dibben, 1999).

Further studies have examined the relationship between acoustic properties of sound and listeners' ratings of various attributes. For instance, Eerola et al. (2012) demonstrated that high energy arousal ratings correlated positively with large changes in spectral flux, and correlated negatively with spectral skewness. A study by Lahdelma and Eerola (2016) reported positive correlations between acoustic properties such as brightness, roughness, flux, and attack time and listeners' ratings for nostalgia and tenderness.

Frequency of occurrence

Listeners are thought to acquire an implicit knowledge of music through repeated exposure (Bharucha, 1994): listeners are able to learn and internalise regularities that organise and constrain that music. In this way, frequency of occurrence has been shown to be important for learning the structures and systems of music in traditional Indian music (Castellano et al., 1984), in Finnish Sami yoik (e.g., Krumhansl et al., 2000) and even in newly invented musical systems (Jonaitis and Saffran, 2009; Oram and Cuddy, 1995).

Several studies have examined the frequency of a chord's occurrence (Broze and Shanahan, 2013; Budge, 1943; Rohrmeier and Cross, 2008). Budge (1943) examined the frequency of the occurrence of diatonic chords in musical pieces from the eighteenth to the nineteenth century, and found that chords built on the tonic accounted for 41.79% of the total, whereas chords built on the mediant comprised a mere 1.35% of the total in major mode contexts. Data gathered by Krumhansl (1985, 1990) strongly correlates with Budge's diatonic chord distribution: chords that appeared more frequently tended to be given higher ratings of 'fittingness' so that more commonly occurring items become encoded as more important and stable (Bigand, 1997; Krumhansl, 1985, 1990), with familiar objects being perceived more favourably than less familiar ones (see e.g., Huron, 2006).

Another approach is to consider the type of chord itself, such as whether a triad is major, minor, or diminished. Again, studies reveal an uneven distribution in the frequency of occurrence of different chord types. Rohrmeier and Cross (2008) reported that the most prevalent chord type in the chorales of J.S. Bach in both major and minor modes is the major chord (60.8% in major modes, 44.9% in minor), followed by the minor chord (17.1%, 33.8%), and the major seventh chord (7.1%, 6.1%), whereas the diminished triad accounts for only 2.3% in major modes and 3.3% in minor. Contrasting this, Broze and Shanahan (2013) reported that the dominant seventh is the most frequently occurring chord in Jazz pieces from 1924 to 1968, accounting for 40.3% of the total, followed by the minor seventh (26.3%) and the major triad (22.1%), while diminished and augmented chords represent only 2.2% and 0.1% of the total respectively. More recently, Mauch et al. (2015) reported that the major triad appeared most frequently in US pop music released between 1960 and 2010, and that use of the dominant seventh was declining.

Uncertainty remains, however, over whether and how listener perception of an isolated chord might be influenced by the frequency of that chord's occurrence in a given music tradition. We may hypothesis that listeners will perceive more frequently occurring chord types more favourably, and for this reason will judge them to be more consonant, pleasant, stable and relaxed. This hypothesis is consistent with the observation that some intervals and chords that were considered 'dissonant' up until the 13th century (e.g. the major third, the major sixth, major and minor triads) gradually became more familiar until, by the time of the Renaissance, their place within Western music had become firmly established and they came to be regarded as fused sonorities (Parncutt, 2011).

Familiarity with timbre

Aspects of listener perception also vary with the instrumental timbre in which stimuli are played. For instance, excerpts played on a piano are perceived as more tense than orchestral versions of the same music, which can be attributed to the piano's sharp attack at note onset (Paraskeva and McAdams, 1997). Chords played using string instruments trigger more reports of nostalgia, melancholy, and sadness than those using a piano timbre, and fewer reports of happiness or joyful emotions (Lahdelma and Eerola, 2016). Familiarity in general is an important and influential factor in music perception (Huron, 2006), but it still remains to be tested whether more familiar timbres tend to promote more positive listener ratings.

Musical training

Musical training seems to make listeners more sensitive not only to the sonority but also to the harmonic structure of Western tonal music. Schellenberg and Trehub (1994) established that listeners with more extensive musical training relied more on the prevalence of intervals in Western music than on the mere simplicity of the frequency ratios of two tones when judging the pleasantness of intervals. Listeners with more extensive musical training perceived intervals with simple frequency ratios, such as the major third or the major sixth, as more pleasant than intervals with even simpler frequency ratios, such as the octave or the perfect fifth (Malmberg, 1918; Van de Geer et al., 1962). According to Schellenberg and Trehub (1994), this may be explained by the fact that the major third and the major sixth are more common in Western music than either the octave or the perfect fifth, and that the judgements of musically trained listeners were influenced by their familiarity with these intervals.

Previous work has reported that musicians are generally more sensitive to C/D than non-musicians (see e.g., Roberts, 1986; Rogers, 2010), and that musicians seem to rely on an understanding of harmonicity when judging C/D (Kung et al., 2014; McDermott et al., 2010). Consonant and dissonant sounds were found to activate different regions of the brains of musicians and non-musicians (Minati et al., 2009), with musicians' brainstem responses to dissonant intervals being more robust and coherent than those of non-musicians (Lee et al., 2009). Bigand et al. (1996) also found that listeners' tension ratings correlated with the roughness of chords, pitch distance and tonal stability, and that these effects were more pronounced for musicians than non-musicians. Lahdelma and Eerola (2016) additionally reported that musicians gave higher tension ratings than non-musicians for augmented triads.

Some aspects of chord perception, however, seem to be less susceptible to the influence of musical training. In terms of judgements of the aesthetic qualities of chords such as pleasantness, musical training was found to have 'no reliable effect' on the pleasantness ratings of chords either in isolation (Johnson-Laird et al., 2012), or in context (Roberts, 1986).

We thus have two contradictory hypotheses regarding the difference between musicians' and non-musicians' chord perception. According to the first, musicians are able to make finer distinctions than non-musicians between different types of chord in terms of C/D and tension since they possess a greater sensitivity to the acoustic features of chords. If this hypothesis is correct, then musicians will tend to perceive consonant chords as more consonant and dissonant chords as more dissonant than their non-musical counterparts. The second hypothesis, however, predicts the opposite result: musicians will perceive dissonant chords to be less dissonant than non-musicians, since increased exposure to and familiarity with dissonant chords will make them seem more agreeable.

Consonance, pleasantness, stability and tension

An early study by Van de Geer et al. (1962) reported that consonant sounds tend to be evaluated as aesthetically pleasing, perhaps contributing to the now commonplace equivalence between a chord's consonance and pleasantness, or dissonance and unpleasantness (Tenney, 1988). Indeed, a few studies have recently used a 'pleasant/unpleasant' (P/U) metric in place of a 'consonant/dissonant' (C/D) metric to evaluate the C/D perception of chords (Cook and Fujisawa, 2006; Johnson-Laird et al., 2012; McDermott et al., 2010). Judgements of consonance and dissonance have also been shown to be coincident, respectively, with judgements of stability and instability (Bigand et al., 1996; Bigand and Parncutt, 1999).

Some evidence stands contrary to this position, however. Examining intervals, Guernsey (1928) reported that the intervals musically trained listeners judged as most fused and smooth were not those rated as most pleasant. Examining chord sequences, Arthurs and Timmers (2016) found that some listeners' C/D and P/U ratings were also uncorrelated. It remains to be tested, for isolated chords, whether these ratings describe one and the same quality of chords or not.

Present study

As identified above, further work is required to understand the various factors which influence listeners' perception of isolated chords. Based on the reviewed studies, we would expect chords that appear frequently to be judged as more consonant, pleasant, stable, and relaxed than chords that appear less often, especially when they are heard in a familiar instrumental timbre. Additionally, we would expect there to be differences between musicians' and non-musicians' perceptions of chords in terms of perceived consonance, stability, and tension.

To examine the extent to which these factors interact, we adopted a dyadic structure in the experimental work reported below. Firstly, a listening test was conducted to assess participants' chord perception for twelve different types of chord, each presented in isolation. Participants comprised three groups who differed by level of musical training (i.e., musicians, learners, and non-musicians), and chords were presented in two instrumentations thought likely to differ in familiarity (i.e., piano and organ). Ratings of consonance, pleasantness, stability, and relaxation were recorded for each chord stimulus. Secondly, to help understand participants' responses, three potential predictors were then examined: (i) the frequency of occurrence of the twelve chord types; (ii) familiarity with the two selected instrumental timbres; and (iii) acoustic descriptions of each stimulus computed using two signal processing toolboxes. This allowed the relationship between the various influences to be analysed, and their relative contribution to listeners' perception of consonance, pleasantness, stability, and relaxation to be assessed.

Listening experiment

A listening experiment was carried out to measure participants' perceptions of isolated chord stimuli in terms of their consonance, pleasantness, stability and relaxation. As described in this section, twelve chord types were presented using two instrumental timbres to participants who differed in their degree of musical training. Listener response data was analysed to assess any effects of chord type, instrumental timbre and musical background and, further, to assess any correlation between consonance, pleasantness, stability and relaxation ratings.

Method

Participants Thirty-three adults (13 male and 20 female, aged 19 to 69 years, with a median age of 24) were divided into three groups according to musical background.

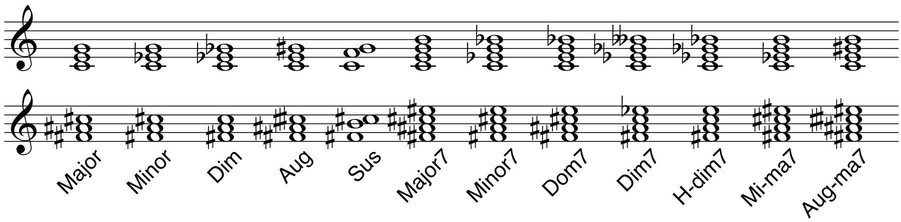

Figure 1. Musical stimuli comprised twelve types of chords: five triads (major, minor, diminished, augmented, and suspended), and seven tetrads (exclusively seventh chords: major, minor, dominant, diminished, half-diminished, minor-major, and augmented-major). These were built on two roots (C and F♯) and played in two instrumental timbres (organ or piano), yielding 48 isolated chord stimuli in total (12 chord types $\times$ 2 root notes $\times$ 2 instrumental timbres).

Eight participants who had completed a music degree were labelled ‘musicians’; 17 participants who had had music lessons (for 3 to 10 years, average 7.88 years) were labelled ‘learners’; and the remaining eight (who had had no musical training of any sort) were labelled ‘non-musicians’.

Materials The twelve types of chord are shown in Figure 1, and comprise five triads and seven tetrads: major, minor, augmented, diminished, suspended, major seventh, minor seventh, dominant seventh, diminished seventh, half-diminished seventh, minor-major seventh, and augmented-major seventh. Equally tempered chords were built on both C4 (261.63 Hz) and F♯4 (369.66Hz), in close root-position. Chords of one second duration and equal intensity were synthesised in Cubase with organ and piano timbres, using the Church Organ and YAMAHA S90ES Piano sounds, respectively. This resulted in 48 stimuli (12 chord types $\times$ 2 root notes $\times$ 2 instrumental timbres).

Procedure The experiment was subdivided into four blocks for each participant, with each block focused on just one of the four variables (consonance, pleasantness, stability, or tension). Each block presented the 48 stimuli in a randomised order, and block order was randomised among participants. Experiments were conducted individually, using PsyScope software to present sound stimuli binaurally through headphones at a comfortable listening level. Participants were permitted to adjust the volume as desired prior to the experiment beginning, but thereafter the level remained fixed. Participants rated stimuli for the selected variable using a horizontal axis whose extremities were labelled with the negative sense at the left (i.e., ‘dissonant’, ‘unpleasant’, ‘unstable’ or ‘tense’) and positive sense at the right (i.e., ‘consonant’, ‘pleasant’, ‘stable’ or ‘relaxed’)

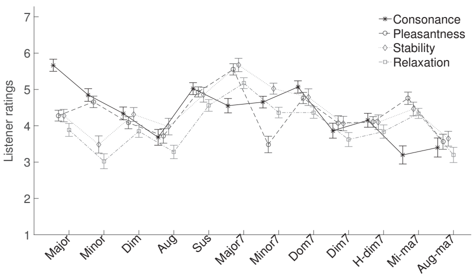

Figure 2. Mean and standard error of participants' ratings of consonance, pleasantness, stability, and relaxation for chords heard with two root pitches and two instrumental timbres.

according to the current experimental block. Seven response alternatives were available to participants (1: extremely; 2: very; 3: moderately; 4: neither; 5: moderately; 6: very; 7: extremely).

ANOVA results

Participants' evaluations of consonance, pleasantness, stability and relaxation were analysed in a mixed model ANOVA with chord type (12 levels) and instrumental timbre (2 levels) as within-subjects factors, and musical background (3 levels) as a between-subjects factor.

The effect of chord type Figure 2 shows the mean and standard error for listeners' ratings of consonance, pleasantness, stability, and relaxation, respectively. There were significant main effects of chord type on all four ratings: consonance ($F_{(5.59,167.70)} = 32.51$, $p < .001$, $\eta_p^2 = .52$), pleasantness ($F_{(3.92,117.80)} = 20.40$, $p < .001$, $\eta_p^2 = .40$), stability ($F_{(5.27,158.04)} = 18.41$, $p < .001$, $\eta_p^2 = .38$) and relaxation ($F_{(4.99,149.96)} = 24.83$, $p < .001$, $\eta_p^2 = .45$) using Greenhouse-Geisser corrections for violations of sphericity. As can be seen from Figure 2, major triads were the most consonant of the twelve chord types, whereas augmented triads, augmented major seventh, and minor-major seventh were judged to be least consonant.

The effect of instrumental timbre Significant main effects of instrumental timbre were found for ratings of consonance ($F_{(1,30)} = 17.46$, $p < .001$, $\eta_p^2 = .37$), pleasantness ($F_{(1,30)} = 32.82$, $p < .001$, $\eta_p^2 = .52$) and relaxation ($F_{(1,30)} = 25.11$, $p < .001$, $\eta_p^2 = .46$), but not for stability ($p = .091$). The chords with piano timbre were always judged to be more consonant, pleasant and relaxed than those with organ timbre.

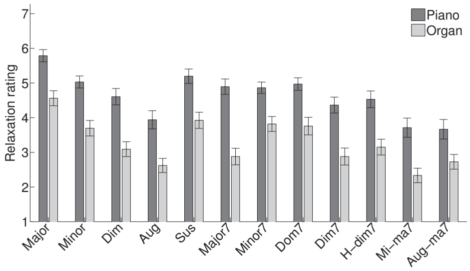

Figure 3. Mean and standard error of participants' relaxation ratings for chords heard with piano and organ timbres.

The interaction between chord type and instrumental timbre There was a significant interaction between the effects of chord type and instrumental timbre for relaxation ratings ($F_{(11,330)} = 2.14$, $p = .017$, $\eta_p^2 = .07$), but not for any of the other ratings (consonance, $p = .766$; pleasantness, $p = .094$; stability, $p = .333$). Follow-up analyses were carried out to understand this interaction, comprising one-way ANOVAs for each timbre separately, with chord type as an independent variable. The effect of chord type was significant for both timbres, and showed a slightly larger effect size for organ ($F_{(5.56,178.17)} = 21.35$, $p < .001$, $\eta_p^2 = .40$) than for piano ($F_{(5.05,161.72)} = 15.66$, $p < .001$, $\eta_p^2 = .33$). Figure 3 reveals that piano chords were generally rated to be more relaxed than organ chords, and that the difference between the two instruments was particularly large for the major seventh chord.

The interaction between musical background and chord type Three of the four ratings revealed an interaction between the effects of chord type and musical background: consonance ($F_{(11.18,167.70)} = 4.13$, $p < .001$, $\eta_p^2 = .22$), stability ($F_{(10.53,158.04)} = 2.13$, $p = .023$, $\eta_p^2 = .12$), and relaxation ($F_{(9.99,158.04)} = 2.54$, $p = .007$, $\eta_p^2 = .15$), but not pleasantness ($p = .086$).

In order to examine the effect of listener group on ratings for these three variables in more detail, post-hoc one-way ANOVAs were conducted for each chord separately with two timbres averaged, with a Bonferroni adjusted alpha level of .004 (.05/12). Although there was no significant effect of listener group on consonance ratings, the ratings for the augmented-major seventh and minor-major seventh were close to being significant ($p = .006$, and $p = .005$, respectively). A significant effect of listener group was found on stability ratings for the major seventh ($F_{(2,30)} = 6.56$, $p = .004$); and for minor-major seventh ($F_{(2,30)} = 6.61$, $p = .004$). The pairwise comparisons between the

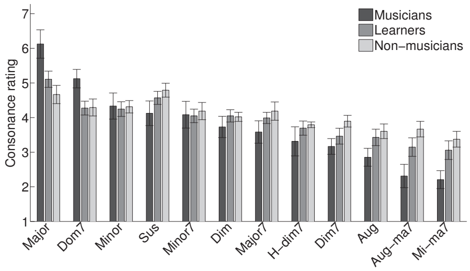

Figure 4. Mean and standard error of participants' consonance ratings, revealing differences between the three listener groups (musicians, learners and non-musicians). Chords are ordered according to their perceived consonance by musicians.

musicians group and the non-musicians were significant for the major seventh ($p = .003$) and the minor-major seventh ($p = .003$), as the musicians group rated these chords as more unstable than the non-musicians did. As for relaxation ratings, the effect of listener group was strongest for the augmented triad and for the major seventh ($p = .032$, and $p = .034$, respectively), although post-hoc analysis using a Bonferroni adjusted alpha level of .004 revealed no significant effect of listener group on ratings for relaxation.

Figure 4 shows the interaction between listener group and consonance ratings of particular chord. The general trend is for musicians to show stronger differences than non-musicians between consonant and dissonant chords, rating the dissonant chords as more dissonant and the consonant chords as more consonant. This suggests that musicians are better able to conceptualise the consonance of different chords. Note that this general interaction was significant, although effects of musicianship on the ratings of particular chords were not, after correction of the alpha value for multiple testing.

The mixed model ANOVA revealed no other significant F-ratios.

Correlation of consonance, pleasantness, stability and relaxation Pearson's product-moment coefficients were computed to assess the relationship between the ratings of consonance, pleasantness, stability and relaxation for each stimulus. For the participant group as a whole, all pairs of ratings were positively correlated. In particular, there was a strong correlation between ratings of pleasantness and relaxation ($r = .977$, $df = 48$, $p < .001$). Consonance ratings were strongly correlated with ratings for the other three dimensions (with pleasantness: $r = .929$, $df = 48$, $p < .001$, with stability: $r = .923$, $df = 48$, $p < .001$; with relaxation: $r = .923$, $df = 48$, $p < .001$), such that more consonant chords were also typically judged to be more pleasant, stable and relaxed.

Table 1. Pearson's product-moment coefficients were computed to assess the relationship between ratings of consonance (C/D), pleasantness (P/U), stability (S/I) and relaxation (R/T) for each stimulus. All variables were correlated for the learners (L) and non-musicians (N), however musicians' (M) ratings of consonance did not correlate with the other three variables.

		C/D			P/U			S/I		
		M	L	N	M	L	N	M	L	N
P/U	r	.168	.607	.741						
	p	.254	$<$ .001	$<$ .001						
S/I	r	.204	.423	.651	.848	.806	.710			
	p	.165	.003	$<$ .001	$<$ .001	$<$ .001	$<$ .001			
R/T	r	.245	.578	.686	.921	.940	.870	.856	.763	.761
	p	.093	$<$ .001	$<$ .001	$<$ .001	$<$ .001	$<$ .001	$<$ .001	$<$ .001	$<$ .001

However, when listener groups were analysed separately, as Table 1 shows, musicians' ratings of consonance did not significantly correlate with the other three dimensions. The remaining pairs (pleasantness and stability, pleasantness and relaxation, stability and relaxation) were positively correlated for all groups.

Predictors of chord perception

This section initially describes three potential predictors which were computed to better explain listeners' perception of isolated chords: (i) the frequency of occurrence of each chord type; (ii) listeners' familiarity with the instrumental timbre; (iii) selected acoustic descriptors for each stimulus. Subsequently, a principle component analysis reduced the numbers of variables involved, and a linear regression assessed the degree to which listeners' ratings of consonance, pleasantness, stability and tension could be predicted using the three predictors proposed.

Predictor 1: Frequency of occurrence

To assess participants' likely familiarity with the twelve chord types, we counted the number of times each chord type appeared in one of J.S. Bach's keyboard works, and in a set of songs by The Beatles. For Bach, we analysed all three movements of the *Italianishes Konzert* BWV 971, counting all and only those chords that formed part of the fundamental harmonic progression of the three movements (n.b., ornamental tones were not considered constituent parts of the chords). In the case of The Beatles, we selected 30 UK best-selling songs, and counted the guitar chord symbols indicated in *The Beatles: complete scores* (1993).

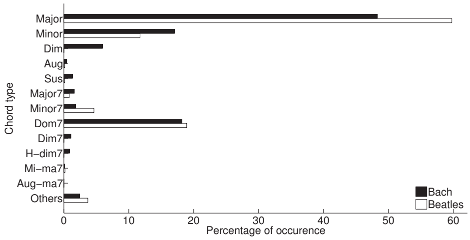

Figure 5. The percentage occurrence of each chord type counted in Bach's *Italianishes Konzert* BWV 971 and in The Beatles' 30 best-selling UK hits.

The percentage occurrence of each chord type is shown in Figure 5, and reveals that major triads were the most prevalent chord type in both oeuvres, followed by dominant sevenths and minor triads. In general Bach's music used a greater variety of chords than featured in The Beatles' guitar chord symbols; however, both oeuvres shared a similar trend in terms of the frequency of chord occurrence according to Spearman's rank-order correlation ($r_s(12) = .692$, $p = .013$). The percentage occurrence of each chord type derived from The Beatles' data was used further in the principle component analysis described below.

Predictor 2: Familiarity with instrumental timbre

To assess a possible influence of timbre familiarity, we gathered additional response data from our 33 participants, detailing their familiarity with the organ and piano instrumental timbres used in the listening experiment described above.

Stimuli for this comprised another 6 chords in root and close position (major chords on C♯4 (277.18 Hz) and G4 (392.00 Hz), and a minor chord on G4). These were created in each of the two instrumental timbres selected (organ and piano), following the synthesis procedure detailed above. Following the procedures described earlier, stimuli were presented to participants in a randomised order, at the same comfortable listening level. Participants rated their familiarity with the instrumental timbre of each chord using a 7-point scale (with 1 being extremely unfamiliar to 7 being extremely familiar).

A mixed model ANOVA with two factors for instrumental timbre and musical background revealed a significant main effect of timbre ($F_{(1,30)} = 14.59$, $p = .001$, $\eta_p^2 = .33$) with the piano timbre typically being reported as more familiar than the organ timbre. Additionally, there was a significant main effect of musical background ($F_{(2,30)} = 14.25$, $p < .001$, $\eta_p^2 = .47$) but no interaction between the effects ($p = .170$).

Table 2. Audio descriptors were computed for each of the experimental stimuli in spectral (S), temporal (T), spectro-temporal (ST) and harmonic (H) domains. Ten values resulted from Timbre Toolbox (TT) analyses, comprising the median and interquartile range (iqr) of the temporal energy envelope values, and of spectral centroid, skewness, flatness and crest. A further four values resulted from MIR Toolbox (MT) analyses, which computed the mean value of spectral regularity, flux and roughness in addition to the key-strength probability relative to the corresponding major key (C or F♯). Further details are described in the Appendix.

Domain	Toolbox	Descriptor	Explanation
T	TT	RMS-energy envelope (median, iqr)	Temporal energy envelope values.
S	TT	Spectral centroid (median, iqr)	Geometric centre of the spectrum.
S	TT	Spectral skewness (median, iqr)	Asymmetry of the spectrum about its mean.
S	TT	Spectral flatness (median, iqr)	Flatness or peakiness of the signal.
S	TT	Spectral crest (median, iqr)	Comparison of maximum and mean spectral values.
S	MT	Spectral regularity (mean)	Degree of uniformity in neighbouring peaks of the spectrum.
ST	MT	Spectral flux (mean)	Degree of change in the temporal evolution of the spectrum.
ST	MT	Spectral roughness (mean)	Estimation of sensory dissonance.
H	MT	Key strength (probability)	Key-strength relative to corresponding major key (C or F♯).

A pairwise comparison showed that the differences between pairs of listener groups were all significant ($p < .05$), and that familiarity ratings increased with musical expertise: musicians reported greater familiarity with both instrumental timbres than learners, and learners in turn reported greater familiarity than non-musicians.

Familiarity ratings were subsequently averaged across chord types to produce a single familiarity value for each instrument, used further in the principle component analysis described below.

Predictor 3: Acoustic feature description

To assess the possible influence of acoustic properties on listeners' chord ratings, each stimulus was represented with an array of numerical values termed 'audio descriptors' that are taken to characterize certain properties of the audio signal. Two widely-used MATLAB-compatible signal processing libraries were selected, the Timbre Toolbox (Peeters et al., 2011) and the MIR Toolbox (Lartillot and Toiviainen, 2007). After initial pre-processing stages, sound analyses were computed in spectral, temporal, spectro-temporal and harmonic domains as shown in Table 2, and detailed further in the Appendix, to produce fourteen acoustic descriptors in total.

Combining the three approaches to explain listeners' perception of chords

Before relating these predictors to listeners' evaluations of consonance, pleasantness, stability and tension, a principal component analysis was undertaken to reduce the

Table 3. Using the frequency of each chord's occurrence, listeners' timbre familiarity ratings, and the 14 acoustic descriptors computed for each stimulus, a principal component analysis extracted three principle components (PCs) with an eigenvalue greater than 1. These relate to instrumentation (PC1), roughness (PC2), and chord strength (PC3). Varimax-rotated values are reported.

Variance explained	PC1 62.24%	PC2 12.94%	PC3 8.47%
Familiarity with timbre	−.976	.159	−.018
Spectral flux	.972	−.200	.026
Spectral flatness median	.957	−.251	.032
Spectral centroid median	.954	−.204	.050
Spectral skewness iqr	−.939	.231	−.077
Spectral centroid iqr	.894	.005	−.027
Spectral flatness iqr	.885	−.328	.070
Spectral skewness median	−.883	.042	.049
RMS energy envelope iqr	−.844	.277	−.182
RMS energy envelope median	.690	−.106	−.226
Spectral crest median	−.688	.680	.072
Spectral regularity	.614	.638	.100
Spectral roughness	.489	−.779	−.053
Spectral crest iqr	−.205	.735	.101
Key strength probability	.080	.257	.801
Frequency of occurrence	−.036	−.057	.899

number of variables observed – comprising the frequency of each chord's occurrence in the Beatles' pieces, the listener ratings of timbre familiarity and the 14 extracted acoustic descriptors – and to investigate correlations among these variables. The analysis extracted three principle components from these variables with an eigenvalue larger than 1, as shown in Table 3. The first component (PC1) correlated strongly with listener-rated timbre familiarity and spectral acoustic features, particularly flux, flatness and centroid. The second component (PC2) related to descriptors for spectral roughness, regularity and crest. The third component (PC3) related to knowledge-based variables such as the frequency of chord's occurrence and computed key strength value. These three components will be referred to as instrumentation (PC1), roughness (PC2), and chord strength (PC3).

A linear regression was conducted in order to assess the ability of these three principle components – instrumentation, roughness, and chord strength – to predict chord perception ratings. As summarized in Table 4, all four ratings were well explained by the instrumentation (PC1) and chord strength (PC3) components, with more than 50% of the variance explained. Instrumentation (PC1) indicated an increase in listeners' ratings for all four variables (consonance, pleasantness, stability and relaxation) when familiar timbres were used, and when chords decreased in spectral flux, flatness, and centroid. Chord strength (PC3) indicated that chords which appeared frequently and which had

Table 4. Results of the linear regression analysis which assessed the ability of the three principal components (PCs, as described in Table 3 caption) to predict chord perception ratings for consonance (C/D), pleasantness (P/U), stability (S/I) and relaxation (R/T).

Variables	C/D $R^2_{adj} = .529$			P/U $R^2_{adj} = .649$			S/I $R^2_{adj} = .533$			R/T $R^2_{adj} = .645$		
	β	p	Partial r	β	p	Partial r	β	p	Partial r	β	p	Partial r
PC1	$-.482$	$< .001$	$-.581$	$-.656$	$< .001$	$-.749$	$-.368$	.001	$-.482$	$-.697$	$< .001$	$-.767$
PC2	.116	.251	.643	.161	.062	.277	.072	.424	.108	.160	.065	.274
PC3	.563	$< .001$	.643	.484	$< .001$	.641	.646	$< .001$	.695	.416	$< .001$	.581

high key strength probability values were more often rated as consonant, pleasant, stable, and relaxed. The roughness component (PC2) did not significantly contribute to the explanation of the listener ratings, however.

General discussion

This study provides empirical support for the notion that listeners' chord perception uses both acoustic properties and knowledge schemata, with an emphasis on the latter. Our results show that more commonly occurring chords tend to be perceived as more consonant: the major triad is the most frequently occurring and most consonant of all the chord types, followed by the dominant seventh and the minor triad. Previous research had shown that, for chords heard in a musical context, the rank order of frequently occurring chords generally corresponds to their relative consonance (e.g., see earlier discussion of Krumhansl (1990) who used the frequency of occurrence data from Budge (1943) to assess the function of major, minor and diminished triads in the context of the existing tonal hierarchy). The present study, however, demonstrates that the relationship between the frequency of occurrence and chord perception can additionally be observed when chords are heard in isolation. The question thus arises of whether a chord sounds consonant because of its frequent occurrence in musical works, or whether a chord occurs frequently because it sounds consonant. However, as our data is correlational, we can only observe that the two are closely associated.

Our regression analysis showed that the principal component for instrumentation was a good predictor of listener ratings of consonance, pleasantness, stability, and relaxation. This component was most strongly correlated with listener ratings of timbre familiarity, so that a more familiar timbre made a chord seem more consonant, pleasant, stable, and relaxed. Several audio descriptors also correlated with this component, suggesting the influence of particular psychoacoustic properties. More positive chord ratings typically occurred when chords had a higher degree of symmetry around the spectral mean value (reduced spectral skewness), and a smaller difference between a spectrum's maximum value and its mean (the spectral crest). This appears consistent with the findings of Eerola et al. (2012), who reported a correlation between spectral skewness and energy levels such that asymmetry leads to higher ratings for tiredness. Conversely, negative ratings occurred when the moment-by-moment variation in successive spectra increased (increased spectral flux), which again supports the findings of Eerola et al. (2012). Negative ratings also occurred when the tonal (vs. noise) content decreased (increased spectral flatness), and when the spectral centre of gravity rose (increased spectral centroid). However, our analysis revealed that the principal component related to roughness was not a significant predictor of listener ratings in our data. This stands contradictory to the general view obtained when chords are presented in context (Lahdelma and Eerola, 2016; Helmholtz, 1877/1954; Johnson-Laird et al., 2012; Pressnitzer et al., 2000; Bigand et al., 1996).

The difference between consonance ratings of listeners with and without formal musical training shows that the consonance of chords is not a purely acoustic phenomenon, but that it varies according to the listener's musical knowledge and experience. The idea that musical training makes listeners more discerning in regard to the harmonic relations between tones (McDermott et al., 2010; Kung et al., 2014) may account for the wider range of ratings of consonance, stability and relaxation provided by musicians in the present study. As such, our findings challenge the view that greater exposure to dissonance would make listeners rate chords less dissonantly due to their comparative familiarity. (Indeed, as Parncutt and Hair (2011) assert, it appears that familiarity need not make a dissonant chord consonant.) Rather, musical training appears to strengthen listeners' knowledge about the frequency (or otherwise) of a given chord's occurrence. Listeners with extensive musical training know that the minor-major seventh and the augmented-major seventh are rare chords to encounter, and that they are thus less important in the tonal hierarchy. This knowledge, acquired as it is through musical training, might make these chords appear more dissonant to musically trained listeners.

In contrast to the other ratings, there was no difference between the pleasantness ratings of the three groups, however. This result is in line with previous studies (Johnson-Laird et al., 2012; Roberts, 1986), and that indicates that musical training is not necessary for the listener to make an aesthetic judgement about a chord.

This study also demonstrated that, for some listeners, the perception of consonance does not always correlate with associated variables. When the participants are considered en masse, ratings for consonance, pleasantness, stability, and relaxation were positively correlated, indicating that chords perceived as consonant tended to be judged as pleasant, stable, and relaxed. However, for some listeners, and for musicians in particular, the perception of consonance did not correlate with perceptions of the other three variables. This is consistent with the findings of Arthurs and Timmers (2016) which found that musicians' consonance and pleasantness ratings were uncorrelated. For musicians at least, it would appear that consonance is an independent variable, and is not interchangeable with the other variables. Since 'consonance' and 'dissonance' are by origin musical terms – in contrast to the other terms used in this study, they are rarely encountered in non-musical contexts – and since musical training enhances one's sensitivity towards the factors which contribute to consonance, perhaps consonance is to some extent a concept for which perception is acquired, or at least strengthened, through extensive musical training.

Further support that the four variables are not interchangeable arose in the regression analysis. Here, a standard coefficient values of chord strength (which is strongly related to both the frequency of a chord's occurrence and its computed key strength) were higher for consonance and stability than for pleasantness and relaxation. On the other hand, the standard coefficient values of instrumentation (which is related primarily to timbre familiarity and to spectral flux, flatness and centroid) were higher for pleasantness and relaxation than for consonance and stability. Together, these results imply that – for isolated chord perception – consonance and stability ratings were influenced more strongly by participants' knowledge of tonal hierarchy, whereas pleasantness and relaxation were more influenced by the acoustic features of chords themselves and the familiarity of the instrument in which the chords were heard.

This conclusion opposes some existing reports about an influence of musical training (and hence knowledge of tonal hierarchy), however: previous studies reported that perception of stability does not differ with musical training (Bharucha and Krumhansl, 1983; Bigand, 1997), while relaxation ratings do depend on musical exposure (Bigand et al., 1996; Lahdelma and Eerola, 2016). A likely explanation for this discrepancy is the absence of musical context in the isolated chord stimuli of the present study, in contrast to the contextually richer stimuli of the earlier studies. Meyer (1956, p. 46) wrote that even a single chord in isolation, without a musical context, is incorporated into and heard within 'the prevalent style of Western music'. Still, there may be a difference of degree: chord sequences may more readily prompt listeners to activate their knowledge of tonal hierarchy than chords without any context. As such, it may have been more difficult, especially for those listeners with little or no musical training, to decode the tonal hierarchy information contained within a single chord due to the absence of context. This difficulty in understanding the underlying tonal hierarchy of chords may have led to clearer differences between musicians' and non-musicians' perceptions of consonance and stability, and to a relative increase in the influence of low-level acoustic features on the perception of relaxation ratings.

Conclusions and future work

Alongside tones and intervals, chords are one of the most fundamental elements in Western music. This study has shown that isolated chord perception is influenced both by physical acoustic features and by listeners' musical schemata. Which of these factors is considered the more influential depends on which attribute or feature of a given chord one chooses to focus on. It appears that judgements relating to tonal hierarchy (e.g. consonance and stability) vary greatly according to the listener's musical schema, whereas judgements relating to aesthetic quality or emotion (e.g. pleasantness and relaxation) depend more on a chord's acoustic features and thus yield results that are more consistent across listeners with differing levels of musical training. Additionally, our work revealed variation across listeners in the relationship between consonance and other associated variables. In particular, the conventional notion that 'consonance accompanies pleasantness' was not found to be true for listeners with musical training.

The present study highlights the need for holistic approaches to consonance perception in future research. In the current study, chord perception could be described to differing degrees by three potential predictors of listener behaviour, regarding the frequency of occurrence of a chord in the prevailing tonal hierarchy, familiarity with the instrumental timbre heard, and descriptive acoustic features of the chord stimuli themselves. Here, the three potential predictors of the chord-rating data were analysed after the listening experiment, meaning that their influence on listener perception of chords was not tested directly. Future studies could instead test the causal relationship between these factors and the perception of chords using a systematic methodology. While this work shows that chord perception can to some extent be explained by frequency of occurrence, future studies should employ a larger corpus of music to consolidate these findings. Moreover, a greater variety of instrumental timbres could be used to examine the influence of timbre familiarity, or acoustic variables could be varied systematically in synthetic stimuli to further develop and confirm our findings.

References

Arthurs Y and Timmers R (2016) On the fluidity of consonance and dissonance: The influence of musical context. *Psychomusicology: Music, Mind, and Brain* 26(1): 1–14.

Bach JS (2009) *Italianische Konzert BWI 971*. Vienna: Wiener Urtext Edition.

Beatles (1993) *The Beatles: Complete scores*. London: Wise Productions.

Benward B and Saker M (2003) *Music in Theory and Practice*. New York: McGraw-Hill.

Bharucha JJ (1994) Tonality and expectation. In: Aiello R and Sloboda JA (eds.) *Musical perceptions*. New York, NY, US: Oxford University Press, pp. 213–239.

Bharucha JJ and Krumhansl CL (1983) The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition* 13(1): 63–102.

Bharucha JJ and Stoeckig K (1987) Priming of chords: spreading activation or overlapping frequency spectra? *Percept Psychophys* 41(6): 519–524.

Bigand E (1997) Perceiving musical stability: the effect of tonal structure, rhythm, and musical expertise. *J Exp Psychol Human* 23(3): 808–822.

Bigand E and Parncutt R (1999) Perceiving musical tension in long chord sequences. *Psychol Res* 62(4): 237–254.

Bigand E, Parncutt R and Lerdahl F (1996) Perception of musical tension in short chord sequences: the influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Percept Psychophys* 58(1): 125–141.

Broze Y and Shanahan D (2013) Diachronic changes in jazz harmony: A cognitive perspective. *Music Percept* 31(1): 32–45.

Budge H (1943) *A study of chord frequencies based on the music of representative composers of the eighteenth and nineteenth centuries*. PhD Thesis, Columbia University, New York.

Castellano MA, Bharucha JJ and Krumhansl CL (1984) Tonal hierarchies in the music of north India. *J Exp Psychol Gen* 113(3): 394–412.

Cook ND and Fujisawa TX (2006) The psychophysics of harmony perception: Harmony is a three-tone phenomenon. *Empirical Musicology Review* 1(2): 106–126.

Cousineau M, McDermott JH and Peretz I (2012) The basis of musical consonance as revealed by congenital amusia. *P Natl Acad Sci Usa* 109(48): 19858–19863.

Dibben N (1999) The perception of structural stability in atonal music: The influence of salience, stability, horizontal motion, pitch commonality, and dissonance. *Music Percept* 16(3): 265–294.

Eerola T, Ferrer R and Alluri V (2012) Timbre and affect dimensions: Evidence from affect and similarity ratings and acoustic correlates of isolated instrument sounds. *Music Percept* 30(1):

49–70.

Guernsey M (1928) The role of consonance and dissonance in music. *Am J Psychol* 40(2): 173.

Helmholtz Hv (1877/1954) *On the sensations of tone, 1877*. New York: Dover.

Huron DB (2006) *Sweet anticipation: Music and the psychology of expectation*. MIT press.

Hutchinson W and Knopoff L (1978) The acoustic component of western consonance. *Interface-J New Mus* 7(1): 1–29.

Hutchinson W and Knopoff L (1979) The significance of the acoustic component of consonance in Western triads. *J Musicol Res* 3(1-2): 5–22.

Johnson-Laird PN, Kang OE and Leong YC (2012) On musical dissonance. *Music Percept* 30(1): 19–35.

Jonaitis EM and Saffran JR (2009) Learning Harmony: The Role of Serial Statistics. *Cognitive Sci* 33(5): 951–968.

Krumhansl CL (1985) Perceiving Tonal Structure in Music. *Am Sci* 73(4): 371–378.

Krumhansl CL (1990) *Cognitive foundations of musical pitch*. Oxford: Oxford University Press.

Krumhansl CL, Toivanen P, Eerola T, Toiviainen P, Järvinen T and Louhivuori J (2000) Cross-cultural music cognition: cognitive methodology applied to North Sami yoiks. *Cognition* 76(1): 13–58.

Kung CC, Hsieh TH, Liou JY, Lin KJ, Shaw FZ and Liang SF (2014) Musicians and non-musicians' different reliance of features in consonance perception: A behavioral and ERP study. *Clin Neurophysiol* 125(5): 971–978.

Lahdelma I and Eerola T (2016) Single chords convey distinct emotional qualities to both naïve and expert listeners. *Psychology of Music* 44(1): 37–54.

Lartillot O and Toiviainen P (2007) A Matlab toolbox for musical feature extraction from audio. In: *DAFx-07*. Bordeaux, France.

Lee KM, Skoe E, Kraus N and Ashley R (2009) Selective subcortical enhancement of musical intervals in musicians. *J Neurosci* 29(18): 5832–5840.

Lerdahl F and Krumhansl CL (2007) Modeling tonal tension. *Music Percept* 24(4): 329–366.

Malmberg CF (1918) The perception of consonance and dissonance. *Psychol Monogr* 25(2): 93.

Mauch M, MacCallum RM, Levy M and Leroi AM (2015) The evolution of popular music: USA 1960-2010. *Royal Society Open Science* 2(5): 150081–150081.

McDermott JH, Lehr AJ and Oxenham AJ (2010) Individual differences reveal the basis of consonance. *Curr Biol* 20(11): 1035–1041.

Meyer LB (1956) *Music and meaning in music*. Chicago: The University of Chicago Press.

Minati L, Rosazza C, D'Incerti L, Pietrocini E, Valentini L, Scaioli V, Loveday C and Bruzzone MG (2009) Functional MRI/Event-related potential study of sensory consonance and dissonance in musicians and nonmusicians. *NeuroReport* 20(1): 87–92.

Nattiez JJ (1990) *Music and discourse: Toward a semiology of music*. Princeton, NJ: Princeton University Press.

Oram N and Cuddy LL (1995) Responsiveness of Western adults to pitch-distributional information in melodic sequences. *Psychol Res* 57(2): 103–118.

Paraskeva S and McAdams S (1997) Influence of timbre, presence/absence of tonal hierarchy and musical training on the perception of musical tension and relaxation schemas. In: *ICMC*. Thessaloniki, Greece.

Parncutt R (2011) The tonic as triad: Key profiles as pitch salience profiles of tonic triads. *Music Percept* 28(4): 333–366.

Parncutt R and Hair G (2011) Consonance and dissonance in music theory and psychology: Disentangling dissonant dichotomies. *Journal of Interdisciplinary Music Studies* 5(2): 119–166.

Peeters G, Giordano BL, Susini P, Misdariis N and McAdams S (2011) The Timbre Toolbox: Extracting audio descriptors from musical signals. *J Acoust Soc Am* 130(5): 2902–2916.

Piston W (1950) *Harmony*. London: Gollancz.

Pressnitzer D, McAdams S, Winsberg S and Fineberg J (2000) Perception of musical tension for nontonal orchestral timbres and its relation to psychoacoustic roughness. *Percept Psychophys* 62(1): 66–80.

Roberts LA (1986) Consonance judgements of musical chords by musicians and untrained listeners. *Acta Acust United Ac* 62(2): 163–171.

Rogers SE (2010) *The influence of sensory and cognitive consonance/dissonance on musical signal processing*. PhD Thesis, McGill University, Montréal.

Rohrmeier M and Cross I (2008) Statistical properties of tonal harmony in Bach's chorales. In: *ICMPC*. Sapporo, Japan.

Sadai Y (1980) *Harmony in its systemic and phenomenological aspects*. Jerusalem: Yanetz.

Schellenberg EG and Trehub S (1994) Frequency ratios and the perception of tone patterns. *Psychon Bull Rev* 1(2): 191–201.

Sethares WA (1998) *Tuning, timbre, spectrum, scale*. London: Springer-Verlag.

Tenney J (1988) *A history of 'consonance' and 'dissonance'*. New York: Excelsior Music Publishing Company.

Van de Geer JP, Levelt W and Plomp R (1962) The connotation of musical consonance. *Acta Psychologica* 20: 308–319.

Vassilakis PN (2005) Auditory roughness as a means of musical expression. *Selected Reports in Ethnomusicology* 12 (Perspectives in systematic musicology): 119–144.

Vassilakis PN and Fitz K (2007) SRA: A web-based research tool for spectral and roughness analysis of sound signals. In: *SMC07*. Lefkada, Greece.

Appendix: Audio descriptor computation

Pre-processing Since stereo data was presented to human listeners in this study, the two selected signal-processing libraries were firstly checked for compatibility with two-channel audio files. Secondly, experimental stimuli were batch-normalised by equalising the root-mean-square (RMS) energy level of each sound-file. This step minimised the impact of individual variability in any default normalisation procedures (documented or otherwise) of the various audio descriptor algorithms. Each toolbox then processed the sound files individually, typically by fragmenting the audio sample into a series of windowed input frames, and computing the required value repeatedly for each of these short duration segments. A measure of the central tendency of these segments was provided at output of the computational analysis: the median value in the case of the Timbre Toolbox, and the mean value for the MIR Toolbox algorithms. The temporal variation was additionally stated in terms of the inter-quartile range in Timbre Toolbox analyses.

Time-domain analysis The 48 stimuli comprise conceptually similar items: a single chord, of fixed duration, synthesised at a pre-determined signal level on one of two virtual instruments. Nonetheless, a degree of variation might arise in the temporal envelope of the signal due to the particular instrumental timbres selected since piano sounds typically have a strong onset and decay rapidly, whereas organ tones are sustained throughout the duration of the chord. The audio signals were thus segmented into 60 ms frames and the energetic content of the signal was calculated within each frame, producing an *RMS-energy envelope* that traces the signal strength through time.

Spectral-domain analyses Spectral descriptors disregard the temporal evolution of the sound, and instead consider the distribution of energy across frequency. Audio signals were again segmented into short frames, and a spectral representation of the sound power within each frame was computed using the squared-amplitude values of the short-term

Fourier transform of the signal. Five statistical processes then summarised the signal content, as follows. The *spectral centroid* (the first statistical moment of the spectrum) represents the sound's spectral centre of gravity. The *spectral skewness* quantifies the asymmetry of the spectrum about its mean. Two further descriptors quantify tonal or noise components in the signal: *spectral flatness* compares the geometric and arithmetic means of the signal, using the insight that white noise produces a flat spectrum while a single sinusoidal component appears as a peak in the spectrum; *spectral crest* acts somewhat similarly, using a frame-by-frame comparison between the maximum value and the arithmetic mean of the spectrum. Finally, *spectral regularity* compares the degree of uniformity (or otherwise) in the amplitude of adjoining peaks in the spectrum.

Spectro-temporal analyses A function of both time and frequency, *spectral flux* computes the frame-by-frame difference observed in the spectrum as the sound evolves, highlighting temporal positions of important spectral changes. Secondly, *spectral roughness* estimates the sensory dissonance that a human listener would experience due to beating between pairs of peaks in the spectrum.

Harmonic analyses The final audio descriptor used in this approach computes a value for *key strength* that quantifies the probability that a given chord is associated with the major key of its corresponding root pitch (C or F♯, as appropriate). To achieve this, the spectrum was wrapped to redistribute the (re-normalised) energy into twelve equally-tempered intervals in a single octave range. The resulting chromagram was then cross-correlated with pre-defined profiles governing the expected pitches for each of the possible key candidates.