Comparing human and machine listening for consonance/dissonance rating of isolated chords

Yuko Arthurs* and Amy V. Beeston†

* Department of Music, University of Sheffield, Jessop Building, 34 Leavygreave Road, Sheffield S3 7RD
† Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello, Sheffield S1 4DP

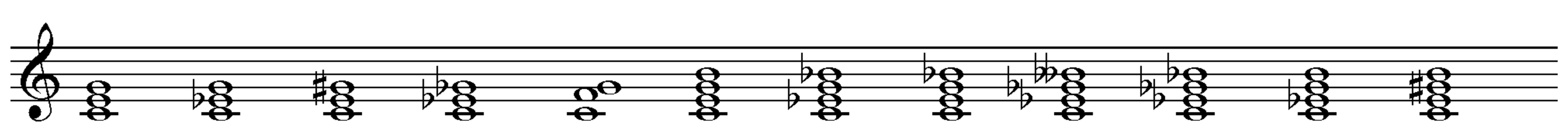
Motivation

- Machine listening is 'unreliable' and can cause frustration in live performance.
- Machine listening is little developed compared to human listening.
- Human auditory system compensates for the listening environment to achieve
  – remarkable sensitivity to some indicative aspects of sound,
  – insensitivity to other trivial aspects of variation (Watkins, 2005).
- Machine listeners have little context-sensitivity (Beeston & Summers, 2013)
  – adversely affected by change in performative or acoustic conditions,
  – cannot distinguish indicative and trivial types of variation in signal.
- With signal analysis techniques and computational modelling of sound perception, we query whether it is possible to reconcile human and computer 'experiences' of sound.
- We believe such work may eventually benefit performers and composers alike.

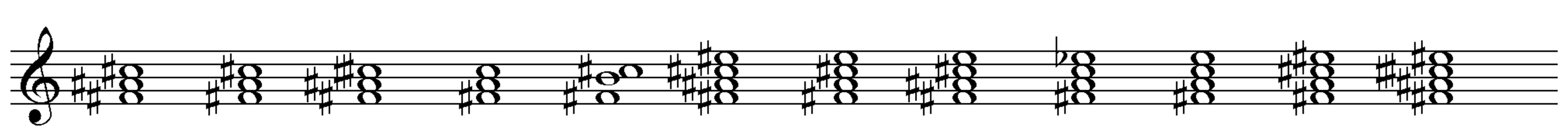
Consonance and dissonance

- The perception of consonance and dissonance (C/D) of chords has been attributed to sensory features, for example
  – roughness (Vassilakis, 2001),
  – harmonic spectrum of chords (McDermott et al, 2010).
- A listener's familiarity with types and timbres of chords might also be an influence on the perception of C/D.
- We explore the way in which acoustic features and familiarity influence the C/D perception of chords heard in isolation.
- We ask how well calculated acoustic values can predict human perception.

Datasets



Root C, triads tetrads



Root F#, triads tetrads

Chord stimuli

- 48 different chords selected
  – 12 chord types
  – 2 key-roots (C, F#)
  – 2 instrumental timbres (piano, organ)
- Chords synthesised in Cubase, stereo files, constant MIDI velocity values.

Chord familiarity

- Participants rated familiarity with each instrumental timbre.
- Frequency of occurrence of chord type was counted in 30 songs of the Beatles.

Human listening

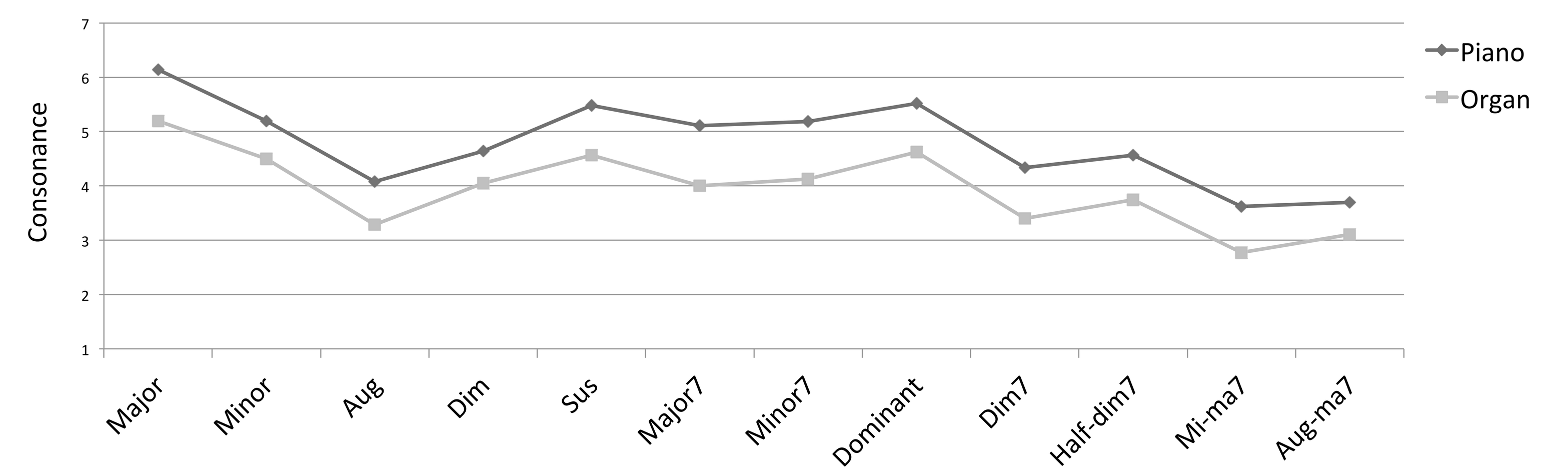
Methods

- 33 participants, listening through a pair of speakers
- C/D of chords and familiarity of timbre each rated on a 7-point scale

Dissonant Unfamiliar						Consonant Familiar
Very	Moderately	Slightly	Neither	Slightly	Moderately	Very
1	2	3	4	5	6	7

Results

- Piano was rated significantly more familiar than organ timbre.
- Major triads were most consonant; comprised 60% of the Beatles' dataset.
- Minor-major seventh chords were most dissonant; 0.17% of Beatles' dataset.



Machine listening

Data preparation

- Two Matlab-based toolboxes used for audio analysis.
- Audio analysis used 2-channel equal-RMS versions of stimuli.
- Differs from human presentation (no loudspeaker or room).

MIR Toolbox (Lartillot & Toiviainen, 2007)

- 5 analyses selected from Eerola et al, 2012: envelope centroid; ratio of high/low frequency energy; spectral skewness; spectral regularity; spectral flux.
- Additional measure: roughness.

Timbre Toolbox (Peeters et al, 2011)

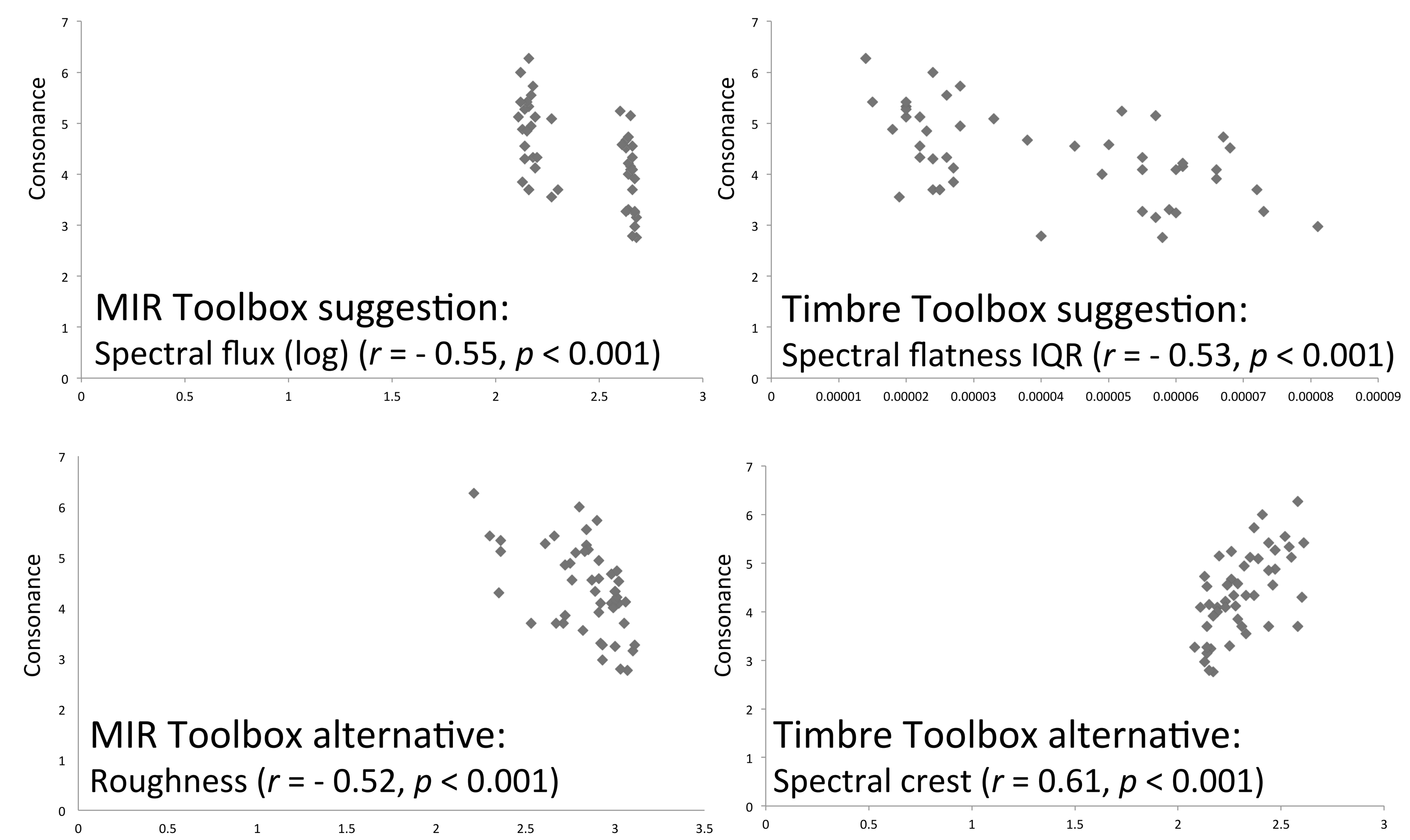
- 4 features selected from Peeters et al (2011): envelope variability; ratio of high/low frequency energy; variability in high/low frequency energy ratio; spectral flatness inter-quartile range (IQR) .
- Additional measure: spectral crest.

Toolbox consistency: spectral centroid

- Said to capture 'brightness' (Grey & Gordon, 1978).
- Measures spectral centre of mass, or high/low frequency balance.
- Toolboxes strongly correlated (r = 0.98, p < 0.001) using
  – MIRToolbox/spectral_centroid_Mean
  – TimbreToolbox/STFTpow_SpecCent_median

Comparison human/machine

- Human C/D ratings and acoustic analyses were compared in SPSS.
- For suggested analyses in each toolbox,
  – envelope features weakly correlated (even when equalised across all stimuli).
  – best remaining: MIR Toolbox spectral flux; Timbre Toolbox spectral flatness.
- Other correlating parameters exist, for example
  – MIR Toolbox roughness; Timbre Toolbox spectral crest.



Discussion and conclusion

- Several machine listening methods correlated with mean human C/D ratings.
- MIR Toolbox revealed that less consonant (more dissonant) sounds contained higher degrees of difference between successive spectral frames (spectral flux), or had increased beating between close frequency peaks (roughness).
- Timbre Toolbox analyses were consistent with these results. Stimuli with less peaky spectra (decreased spectral crest) or with more inherent variability (larger spectral flatness IQR) were rated as less consonant (more dissonant).
- Such features are likely to be useful in human/computer music as they capture details of signal content and variability to which human listeners are sensitive.

References

Beeston, A.V. and Summers, M.A.C. (2013). Groundwork for a resource in computational hearing for extended string techniques. In: 10th International Symposium on Computer Music Multidisciplinary Research (CMMR), Marseille, France, 15-18 Oct., pp. 662-669.

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 Perception, 30 (1), pp. 49–70.

Grey, J. M., Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres, J. Acoust. Soc. Am. 63 (5), pp. 1493–1500.

Lartillot, O., and Toiviainen, P. (2007). A Matlab toolbox for musical feature extraction from audio. In: 10th International Conference on Digital Audio Effects, Bordeaux, France, 10-15 Sep., pp. 237–244.

McDermott, J.H. Lehr, A.J. and Oxenham, A.J. (2010). Individual Differences Reveal the Basis of Consonance, Curr Biol. 20 (11), pp. 1035–1041.

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

Vassilakis, P. N. (2001). Perceptual and Physical Properties of Amplitude Fluctuation and their Musical Significance. Doctoral Dissertation. Los Angeles: University of California, Los Angeles; Systematic Musicology.

Watkins, A. J. (2005). Perceptual compensation for effects of reverberation in speech identification, J. Acoust. Soc. Am. 118 (1), 249–262.