Proceedings of the Australian Acoustical Society Conference, Melbourne, 24-26 November 1999, pp 47-54.

‘PSYSOUND’: A COMPUTER PROGRAM FOR PSYCHOACOUSTICAL ANALYSIS

D. Cabrera

Department of Architectural and Design Science, University of Sydney

ABSTRACT

This paper outlines the capabilities of a computer program called ‘PsySound’ (written by the author), which implements a range of psychoacoustical models.

1. INTRODUCTION

The relationship between an acoustic stimulus and the corresponding sensation is no simple matter. While physical measurements are relatively easy to perform, psychoacoustical models are principally available in limited or expensive forms, and in that sense, are unavailable to all but the most specialised researchers. This is a matter of some frustration to researchers, students and educators with limited resources.

This paper describes a program called ‘PsySound’, in which several psychoacoustical models are implemented. PsySound is compiled for Macintosh PPC, and is freely available from this author’s web site (http://members.tripod.com/~densil/). This paper primarily refers to the current version of PsySound (version 2.x), which combined what were previously five separate programs into one.

PsySound reads 16-bit signed integer sound files, with a sample rate of 44100 Hz, and either 1 or 2 channels. Sound Designer 2, AIFF, and Microsoft Wave formats can satisfy these criteria. It analyses the files in a succession of overlapping 93 ms windows. Results are recorded in four tab-delimited text files: one for level, spectrum and cross-channel time series data; one for loudness and related time series data; one for pitch and related time series data; and one for summary data. The broad structure of the program is shown in Fig. 1.

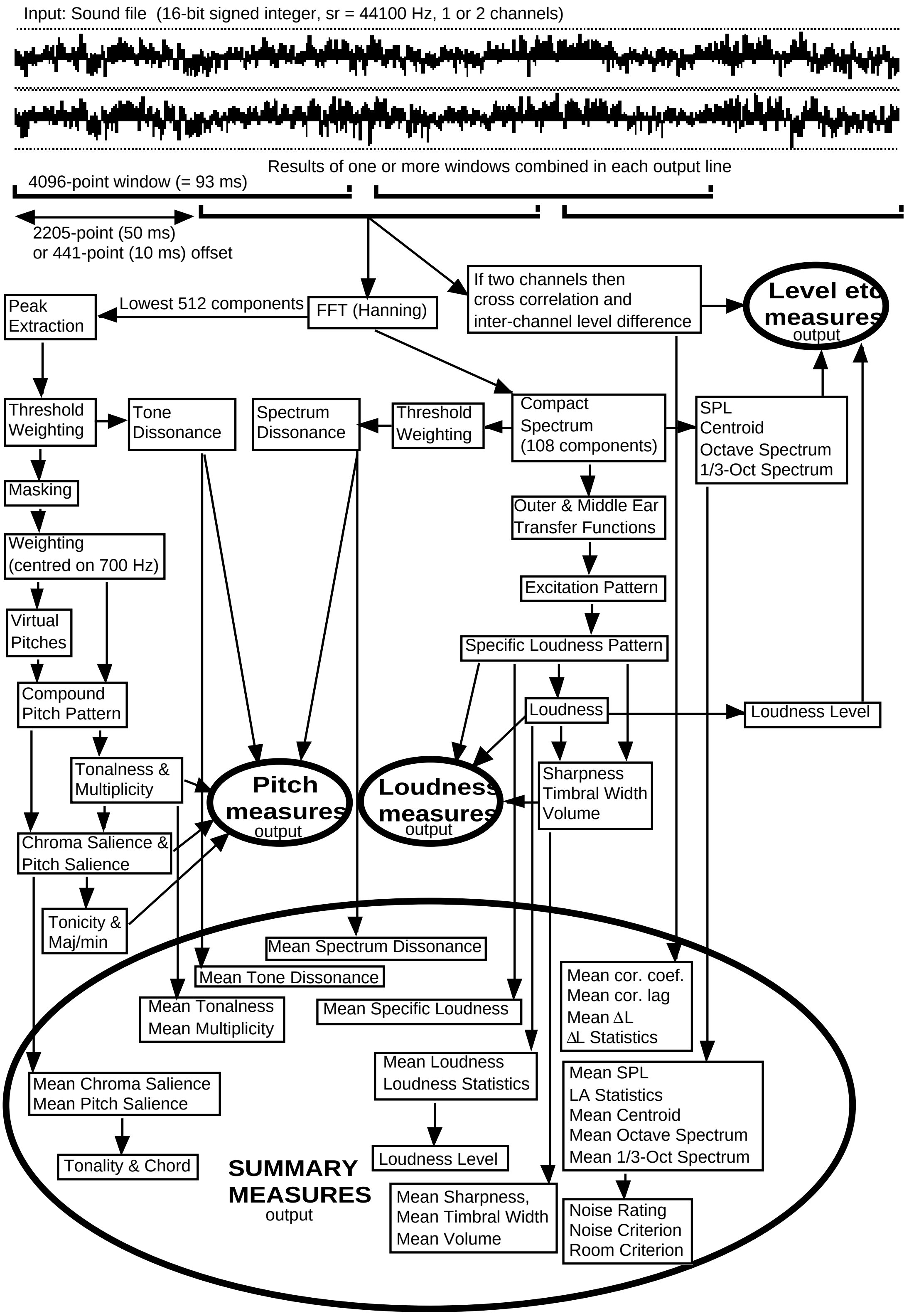

Fig. 1: Program flow diagram.

2. MEASURES

Level, spectrum and cross-channel measures

Frequency analysis is achieved initially by a 4096-point Fourier transform. In order to speed up processing, the frequency spectrum is reduced to a compact form: linear frequency distribution is retained at low frequencies, while twelfth-octave distribution applies to higher frequencies (reducing the spectrum from 2048 to 108 components). From this, octave (from 31.5 Hz) and third-octave (from 125 Hz) spectra are derived, as well as L_A, L_B, L_C, L_{lin}, and spectral centroid. Summary measurements also include statistical A-weighted levels and background noise measures (NC, NR, and RC).

All of the abovementioned measures are easily obtainable by other means. Nevertheless, their presence in PsySound output is important for program calibration. It can also be useful to compare physical and psychoacoustical measures in an effort to understand how various dependencies interact. For example, a comparison between NR and loudness level measurements for a range of sounds having the same SPL but varying bandwidth might show that NR decreases with a smoother spectrum, while loudness level increases.

The two audio channels of a sound file are compared by calculating the unweighted inter-channel level difference and their cross correlation function (for lags between ±1 ms). Only the maximum absolute value of the function is recorded, together with the corresponding lag time. A slight central tendency is applied in an effort to produce meaningful results for periodic cross correlation functions. In some circumstances, the lag time and level difference could be interpreted in terms of lateralisation, especially when their signs match. The cross correlation coefficient might be interpreted in terms of the spread of sound in the horizontal plane (in a manner similar to auditory spaciousness measures in auditoria). Statistical level difference measures are calculated for the summary. In some situations, ΔL_{10}-ΔL_{90} or ΔL_{20}-ΔL_{80} could be used to quantify the lateralisation range in the sound.

Measures related to loudness

The Australian Standard [1] procedures for calculating loudness have a number of deficiencies which have been accounted for by the ‘Cambridge School’ of psychoacoustics. Stevens’ method (ISO 532A) is only suitable for sounds without tonal components. Zwicker’s method (ISO 532B) has broader application, but is based on an auditory filter model that is challenged by more recent research. In addition to using the more sophisticated auditory filter model of Glasberg and Moore [2], the loudness model of Moore et al [3] takes a more detailed approach to ear transfer functions and the loudness function. This model is implemented in PsySound. The free field, diffuse field or field at the eardrum can be assumed.

Specific loudness is the loudness attributable to an auditory filter. The specific loudness function extends from the low frequency filters (with centre frequencies at around 50 Hz) to the high frequency filters (with centre frequencies around 15 kHz). A psychoacoustical ‘frequency’ scale accounts for the distribution of sound in the cochlea. Following Moore et al [3], the Erb unit is used for this ‘Erb-rate’ scale, with values ranging between 2 and 39 Erbs. PsySound calculates specific loudness at 0.25

Erb intervals.

When read directly, the specific loudness function shows the parts of the frequency spectrum that make the strongest contribution to loudness. The specific loudness function also contributes to several higher level measures: PsySound calculates loudness, sharpness, timbral width and volume from the specific loudness function.

Loudness is the integral of specific loudness, or the area under the specific loudness curve. Figure 2 shows the specific loudness functions produced by a narrow band and broad band sound, both presented at 60 dB(SPL). While the A-weighted level is a little lower for the pink noise, its loudness is more than four times that of the 1 kHz tone. These results were calculated by PsySound, assuming a free field.

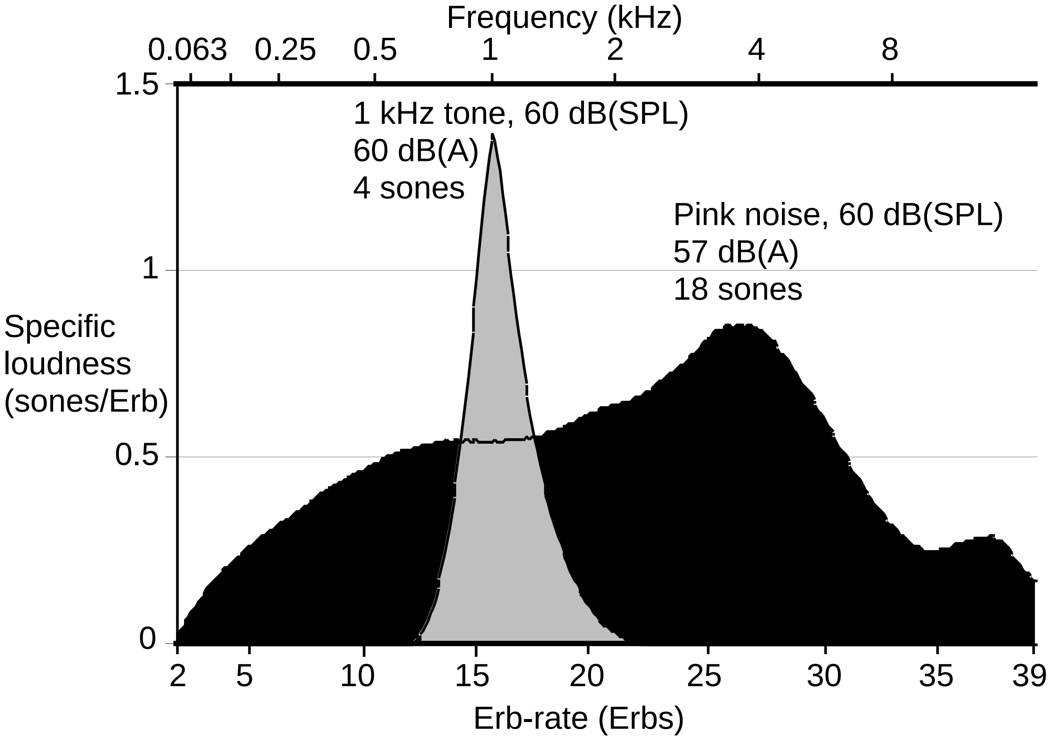

Fig. 2: Specific loudness functions of a 1 kHz tone and pink noise, both at 60dB(SPL).

Statistical loudness measures are calculated for the summary results. These might be manipulated in a manner akin to the more familiar statistical SPL measures, except that ratios are used instead of differences. Zwicker and Fastl [4] find a high loudness percentile, such as N_5 or N_{10}, usually gives the best representation of overall loudness, the louder moments of a sound being more salient than its quieter moments.

Sharpness is a subjective measure of sound on a scale extending from dull to sharp - sometimes it is thought of as a pitch-like (low-high) aspect of timbre. ‘Brightness’ and ‘density’ are two other terms that have been used to denote equivalent or closely related attributes by, for example, Boring and Stevens [5] and Lichte [6]. PsySound implements the sharpness models of Zwicker and Fastl [4] and Aures [7].

Zwicker and Fastl’s model is simply a weighted centroid of specific loudness, while Aures’ model is more sensitive to the positive influence of loudness on sharpness. Both models use the Bark (rather than Erb) scale of critical-band rate, so PsySound simply transforms Erbs to Barks, using the specific loudness function already generated, producing minor deviations from the original models.

Timbral width is a simple measure proposed by Malloch [8], inspired by Pollard and Jansson's [9] tristimulus method of timbre analysis. It measures the flatness of the specific loudness function. Its implementation in PsySound had to be changed from that of Malloch because Malloch used Stevens' [10] loudness calculation method, which has broader frequency components than Glasberg and Moore's auditory filters.

Volume is a subjective measure of sound on a scale extending from small to large. It is a rather old-fashioned measure, and was the subject of Stevens' PhD thesis [11]. A model of the volume of pure tones was developed by Terrace and Stevens [12]. Large volume is associated with low frequency, high intensity, and broad bandwidth. Volume and spaciousness have very similar subjective definitions (the apparent size of the sound), suggesting a binaural component to volume. Recently this author [13] has derived a preliminary model of auditory volume using the specific loudness function, which is implemented in PsySound.

Dissonance

Musical dissonance is determined by a combination of acoustic and contextual factors. The contextual factors relate to what might be called 'musical language'. To measure them would be a complex task, possibly suited to neural network modelling. However the measurement of the acoustical component of dissonance is relatively simple, and models have been proposed by Kameoka and Kuriyagawa [14], Hutchinson and Knopoff [15], Greenwood [16] and Sethares [17]. These models assume that acoustical dissonance is caused by interference (roughness) between frequency components within the auditory filters.

Dissonance is a type of roughness, and fairly sophisticated roughness models have been developed by Aures [18] and Daniel and Weber [19]. Unlike 'dissonance' models, these are sensitive to amplitude and frequency modulation effects, and model the effect of loudness on roughness. However the computational complexity of such models precluded them from the present version of PsySound.

PsySound calculates dissonance in four ways. It calculates dissonance using all the components of the 'compact spectrum', and also using just the tonal components extracted in the early stages of Terhardt et al's [20] pitch model. In both of these cases, the algorithms of Hutchinson and Knopoff [15] and Sethares [16] are applied. The former algorithm normalises the results, and uses linear intensity. The latter algorithm does not normalise the results, and uses scaled decibels (following personal communication with Sethares). When applied to the compact spectrum, these algorithms measure the noisiness of the sound; when applied to the tonal components, they come closer to measuring musical dissonance. Sethares [21] and Malloch [8] give detailed examples of how dissonance models can be used in music analysis.

Pitch Measures

Pitch, here, is used in the psychoacoustical sense: it refers to the *perceived* pitch(es) of sound. This type of analysis is to be distinguished from 'pitch tracking' (musical score or MIDI sequence extraction), and from frequency analysis.

Pitch is multidimensional, at least involving the components of pitch height and pitch strength (or salience). Shepard [22] has developed much more sophisticated models of the structure of 'pitch space', accounting for pitch height, octave similarity and the cycle of fifths. Including pitch strength and the temporal dimension renders futile the reduction of such structures to a flat piece of paper.

The pitch model of Terhardt et al [20] was chosen for PsySound, primarily because of its track record in music analysis, especially in the work of Parncutt [23]. This relatively simple model (based on frequency spectrum analysis rather than autocorrelation), predicts pitch height and strength, virtual pitches and pitch shifts. Additional measures proposed by Parncutt allow the estimation of two types of tonalness (how tone-like the sound is) and multiplicity (the number of pitches heard).

PsySound quantises the results to fit the 12-tone equal temperament scale (saliences of out-of-tune pitches are shared between the adjacent pitch categories). By default it does not implement Terhardt's pitch shifts, as these degrade the results for such a coarse quantisation. Pitch salience patterns are expressed linearly over the pitch height range, and circularly over the chroma range. Fig. 3 shows how these aspects of pitch might be combined. It shows the mean pitch salience of the entire final movement of Mahler's Ninth Symphony, as analysed by PsySound. Even in such a chromatic movement, its tonic of D flat is clearly discernible, although whether the key is major or minor is unclear.

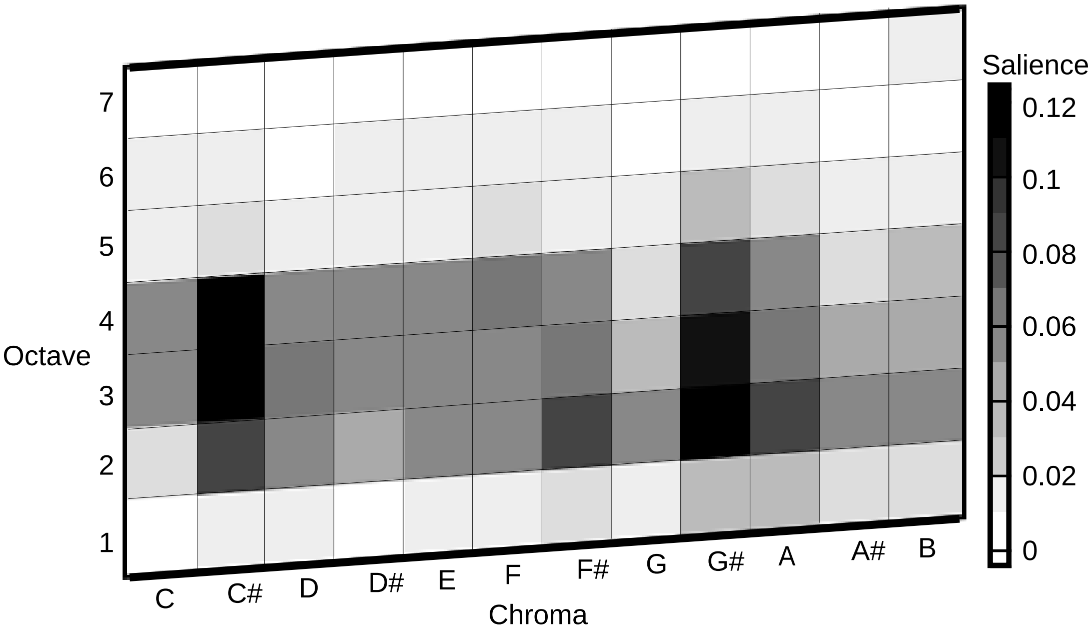

Fig. 3: The mean pitch salience of the fourth movement of Mahler's Ninth Symphony.

PsySound correlates measured chroma saliences with those of 24 keys and 27x12 octave-spaced chords, thereby conducting a crude automatic harmony analysis.

3. APPLICATIONS

PsySound was designed for use in research and education, particularly in areas related to music. Earlier versions of PsySound have been used as part of a course ‘Musical Applications of Psychoacoustics’, taught as an elective at the University of Sydney’s Conservatorium of Music in 1988.

In research it has been applied by Jeong [24] in the analysis of stimuli used in just-noticeable difference experiments in room acoustics. A very early version was used by Schubert [25] in a time series analysis of musical emotion. It is being used by Neil MacLachlan (RMIT University) for the perceptual modelling of bells with harmonic series. This author is using PsySound to analyse music recordings, as well as for the analysis of listening test experiment stimuli (using a dummy head microphone).

4. CONCLUSION

PsySound aims to provide an accessible platform for psychoacoustical analysis, oriented towards musical measures. It combines several models in a single program. It provides interfaces so that the models can be applied directly to sound files, and produces relatively detailed results for spreadsheet, statistical and graphing programs.

The large number of measures implemented in PsySound could mislead a reckless correlation hunter. Nevertheless they allow several measures of related phenomena to be compared, and in this capacity PsySound is an efficient exploratory and educational tool. While PsySound almost presents itself as a ubiquitous analysis program, it would benefit from the inclusion of more sophisticated spatial measures, as well as roughness, fluctuation strength and rhythm models. Proper temporal integration of loudness, and pitch streaming would greatly enhance the present measures.

ACKNOWLEDGMENTS

The author is thankful to Stephen Malloch, Brian Moore, Richard Parncutt, William Sethares and Ernst Terhardt for their advice and permission to include their algorithms. This research was conducted under an Australian Postgraduate Award and a Department of Architectural and Design Science Supplementary Scholarship.

REFERENCES

[1] Australian Standard AS 3657-1996, Acoustics-Expression of the subjective magnitude of sound or noise, Part 2: Method for calculating loudness level, Standards Australia, 1996 (ISO 532:1975, Acoustics-Method for calculating loudness level)

[2] B Glasberg and B Moore, ‘Derivation of auditory filter shapes from notched noise data’, Hearing Research, 47, pp 103-137

[3] B Moore, B. Glasberg and T Baer, ‘A model for the prediction of thresholds, loudness, and partial loudness’, Journal of the Audio Engineering Society, 45/4,

1997, pp 224-240

[4] E Zwicker and H Fastl, *Psychoacoustics* (Springer, Berlin, 1999)

[5] E Boring and S Stevens, 'The nature of tonal brightness', Proceedings of the National Academy of Science, 22, 1936, pp 514-521

[6] W Lichte, 'Attributes of complex tones', Journal of Experimental Psychology, 28/6, 1941, pp 455-480

[7] W Aures, 'Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale', Acustica, 59, 1985, pp 130-141

[8] S Malloch, *Timbre and technology* (PhD thesis, University of Edinburgh, 1997)

[9] H Pollard and E Jansson, 'A tristimulus method for the specification of musical timbre', Acustica, 51, 1982, pp 162-171

[10] S Stevens, 'Perceived level of noise by mark VII and decibels (E)', Journal of the Acoustical Society of America, 51/2, 1972, pp 575-601

[11] S Stevens, *The volume and intensity of tones* (PhD thesis, Harvard University, 1933)

[12] H Terrace and S Stevens, 'The quantification of tonal volume', American Journal of Psychology, 75, 1962, pp 596-604

[13] D Cabrera, 'The size of sound: auditory volume reassessed', Proc 1999 Australasian Computer Music Association Conf, Wellington NZ, pp 26-31

[14] A Kameoka and M Kuriyagawa, 'Consonance theory, part II: Consonance of complex tones and its calculation method', Journal of the Acoustical Society of America, 45, 1969, pp 1451-1459

[15] W Hutchinson and L Knopoff, 'The acoustical component of western consonance', Interface, 7, 1978, pp 1-29

[16] D Greenwood, 'Critical bandwidth and consonance in relation to cochlear frequency-position coordinates', Hearing Research, 54/2, 1991, pp 164-208

[17] W Sethares, 'Local consonance and the relationship between timbre and scale', Journal of the Acoustical Society of America, 93/3, 1993, pp 1218-1228

[18] W Aures, 'Ein Berechnungsverfahren der Rauhigkeit', Acustica, 1985, 58, pp 268-281

[19] P Daniel and R Weber, 'Psychoacoustical roughness: Implementation of an optimized model', Acustica, 83, 1997, pp 113-123

[20] E Terhardt, G Stoll and M Seewann, 'Algorithm for extraction of pitch and pitch salience from complex tonal signals', Journal of the Acoustical Society of America, 71/3, 1982, pp 679-688

[21] W Sethares, *Tuning, timbre, spectrum, scale* (Springer, London, 1998)

[22] R Shepard, 'Geometrical approximations to the structure of musical pitch', Psychological Review, 89/4, 1982, pp 305-333

[23] R Parncutt, *Harmony: a psychoacoustical approach* (Springer, Berlin, 1989)

[24] D Jeong, *Just noticeable differences in tonal quality as a measure of room acoustics* (PhD thesis, University of Sydney, 1998)

[25] E Schubert, *Measurement and time series analysis of emotion in music* (PhD thesis, University of New South Wales, 1999)