PsySound3: a program for the analysis of sound recordings

Densil Cabrera[a], Sam Ferguson[a], Farhan Rizwi[b] and Emery Schubert[b]

[a]University of Sydney, Faculty of Architecture, Design and Planning, NSW 2006 Sydney, Australia
[b]Empirical Musicology Group, University of New South Wales, NSW 2052 Sydney, Australia
densil@usyd.edu.au

This paper describes the sound analysis software PsySound3, which was written by the authors. The software currently includes a range of general sound analysis techniques (e.g., spectrum, cepstrum, autocorrelation, Hilbert transform, sound level meter emulator), as well as implementations of psychoacoustical algorithms often associated with sound quality (e.g., loudness, sharpness, loudness fluctuation, roughness, pitch, binaural attributes). In some cases, PsySound3 makes available multiple models of the one auditory attribute – for example it implements dynamic and static loudness models using Erb- and Bark-based auditory filters. The program is extensible, and so has the potential to allow researchers to share their analysis models using a common interface. PsySound3 is written in Matlab, and will also be available as a stand-alone program. The software is freely available via www.psysound.org.

1 Introduction

The analysis of sound recordings is an important process in many sub-disciplines of acoustics, including sound quality, architectural acoustics and musical acoustics. While many commercial solutions exist, for many researchers and students these can be out of reach, or else not flexible enough for the particular problem at hand. Nevertheless, for the most part, the analysis method is a tool to facilitate answering a question, and creating that tool may distract a researcher from their main focus. This paper briefly describes a freely available open source software environment for sound analysis which in some contexts may facilitate research and teaching in fields such as psychoacoustics, sound quality, musical acoustics and architectural acoustics. This software was written in the context of university environments, for which needs and priorities may differ considerably from those of commercial contexts.

PsySound3 was written by the authors with many contributions, some substantial, from several others. It uses the Matlab environment but has a graphical user interface that requires no special knowledge of Matlab to use. At the time of writing it is not available as a stand-alone program, but this is intended for the future.

This paper describes the major features of the program in the sequence that a user of the program would encounter them, with an emphasis on features of interest. A more detailed and illustrated description of the program is given by Cabrera et al. [1].

2 Files and calibration

PsySound3 has soundfiles as its input, and does not perform real-time audio analysis. It accepts a large variety of file formats. It can be set up to analyse individual files or collections of files (batch processing), and each file is analysed in its entirety.

Gain sensitivity is an important aspect of using psychoacoustical models, because in most cases there is a complex relationship between input level and the output parameter values. There are several calibration options for setting gain sensitivity in the analysis including calibrating using calibration tones recorded in particular files. Another method allows for the A, B, C or unweighted equivalent sound pressure level of each of the files being analysed to be set to a particular value. A third method allows the maximum, median or minimum equivalent sound pressure level for a set of files to be specified, and all files in that set are adjusted by the same gain offset to meet that specification. Finally, a fixed gain offset with respect to full scale can be specified. While the first method of calibration would be most common in acoustic measurements (where microphone calibrators are used), our experience is that researchers in disciplines such as music psychology may prefer to use one of the other calibration methods because microphone calibrators may not be used, and the listening level does not need to be set accurately for the analysis goals to be met. These alternative methods are also useful in the systematic exploration of the effect of signal gain on psychoacoustical parameters.

3 Audio Analysers

Audio analysers are largely self-contained modules within the PsySound3 program, thereby allowing new analysers to be added by anyone who has the time, inclination and moderate programming ability required. In some cases existing analysers are based around pre-existing code, which has been 'wrapped' with some additional features and inserted into the program directory structure. Having many analysis methods driven by the one program has the advantage of being able to coordinate and, to an extent, automate analyses in which the outputs of multiple analysis methods are compared. The analysers mentioned in this section are examples of what can be done, but there is the potential for many other analysers to be added in the future.

Many analysers work by dividing the soundfile into successive (or overlapping) frames, from which spectral patterns or single-number parameters are derived. Having stepped through the entire sound file, the main analyser output formats are time-series objects (showing how a single parameter varies over time), spectrum objects (showing how a parameter is distributed across a non-time dimension such as frequency) and time-spectrum objects (such as a spectrogram). Some analysers can yield an output rate equal to the audio sampling rate of the file being analysed, potentially leading to a substantial increase in the size of the analysis compared to the file size of the input wave. However, the program allows the output of multiple analysers to be 'synchronised', meaning that high output rate data are downsampled, low output rate data are upsampled (although in practice that is rare), and the step size of discrete analysis frames is set to a given value. Synchronised output data allow for easy comparison between time series and may avoid excessive data density for a given application.

3.1 Sound level, spectrum and related analyses

PsySound3 includes analysers that are not especially concerned with psychoacoustics, but could be thought of as general purpose audio analysers. The collection includes:

- A sound level meter emulator (with A, B, and C weighting, and 'fast' and 'slow' integration times using exponential temporal integration).
- An implementation of the revised low-frequency B weighting analysis method proposed by Soulodre [2] and Soulodre and Lavoie [3] for loudness monitoring of broadcast audio. Unlike the sound level meter, this uses a rectangular temporal window for rms signal determination.
- A Hilbert transformer, yielding the 'instantaneous' sound pressure level and frequency of the input waveform.
- A fast Fourier transform (FFT) analyser, yielding the time-varying spectrum, as well as spectral metrics as time series objects, and the time-averaged power spectrum. This analyser also allows a 'zoom' spectrum to be calculated based on the chirp-z-transform.
- Real and complex cepstrum analysers. In the latter case there is an option to lifter the cepstrum, so as to return spectrum values. This is typically used in separate analyses of fine and coarse spectral features (for example, harmonic series versus formants in analyses of voice recordings).
- Auto-correlation analysis, which includes a crude calculation of pitch and pitch strength as a function of time. A second auto- and cross-correlation analyser is also included, which is mentioned in the section on binaural analysis.
- Constant percentage bandwidth analysis (octave, 1/3-octave, 1/12-octave).

3.2 Loudness and related analyses

Loudness modelling is one of the most important aspects of sound quality software, but there are many ways in which loudness can be modelled. PsySound3 currently implements two main loudness models – that of Chalupper and Fastl [4] and of Moore, Glasberg and Baer [5]. PsySound3 also implements the ISO532B steady state loudness model using 1/3-octave band long term average sound pressure levels as input.

The model of Chalupper and Fastl uses an adaptation of code provided by Josef Chalupper. Calculations use the Bark scale of critical bands, and auditory temporal integration is included in the model. Outputs include the time-varying specific loudness pattern, time-averaged specific loudness pattern, loudness, sharpness, and loudness fluctuation. Loudness fluctuation [6] is related to fluctuation strength, and the model was specifically designed and evaluated for fluctuating broadband and real life sounds.

The model of Moore et al. uses the narrower Erb scale of auditory filters, and the temporal integration procedures given by Glasberg and Moore [7] remain to be implemented. In fact the code used is based on the PsySound2 code (dating from 1998-2002), with some minor improvements. Outputs are similar to those of the Chalupper and Fastl loudness analyser, except that loudness fluctuation is not given. Additional parameters include an auditory 'volume' model of Cabrera [8], and acoustic dissonance (also known as 'roughness') models of Hutchinson and Knopoff [9] and Sethares [10]. A separate audio analyser provides an implementation of Daniel and Weber's psychoacoustical roughness model [11].

Loudness models such as these are particularly useful in analysing sound with widely varying bandwidth. For a given sound pressure level, greater bandwidth is often associated with substantially greater loudness. Sound pressure level models do not capture this effect.

3.3 Pitch analysis

PsySound3 has two types of pitch analyser: (i) implementations of 'pitch estimation' algorithms such as SWIPE [12], SWIPE' [12], Praat [13, 14] and YIN [15]; and (ii) a model capable of characterising multiple simultaneous pitch percepts based on algorithms of Terhardt et al. [16] and Parncutt [17]. The pitch estimation algorithms provide time series data for pitch-height and pitch-strength (of a single pitch percept at any given time). The pitch algorithm of Terhardt et al. [16] is based on template matching of harmonic series to find the heights and strengths of multiple virtual pitches (at the respective fundamentals of the viable harmonic series), as well as 'spectral' pitches (such as the pitch of a pure tone). These are interpreted using processes proposed by Parncutt [17] to yield the multiplicity, complex tonalness and pure tonalness of the sound as a function of time. The implementation in PsySound3 also identifies chords based on chroma salience patterns derived from octave-spaced tones, as suggested by Parncutt.

3.4 Binaural analysis

Currently binaural analysis is limited to one audio analyzer, which implements auto- and cross-correlation analysis to yield parameters described by Ando [18]. The cross-correlation parameters can be used for investigating spatial hearing (e.g., auditory lateralization and image width). The core code for this analyzer was provided by Shin-ichi Sato. We envisage more binaural and multichannel analysers in the future.

4 Post processing

Once audio has been analysed, the results of the analysis can be used in several ways within the post processing section of PsySound3. This includes representing or exporting data and further analysis.

4.1 Output data

Like most sound analysis programs, PsySound3 produces output data formatted as single numbers, arrays of numbers (e.g., time series, time-spectra or simple spectra) and visual charts. Numeric data can be exported for further analysis or charting in other programs, and charts generated by PsySound3 can be exported for further editing or use elsewhere. PsySound3 also produces auditory displays of data, which are briefly described in the next subsection.

4.2 Exploratory sound analysis

One of the distinguishing features of PsySound3 is that it provides tools for sonifying the results of analysis. The purpose of this is to allow people to hear the phenomena that are the basis of the analysis – which can be useful in developing an understanding of how output data relate to the original recording. Exploratory sound analysis (ESA) is an approach to the sonification of sound recordings which is achieved primarily by fragmenting and rearranging the original recording [19], rather than using conventional auditory graphing techniques. This idea draws on the concept of 'exploratory data analysis' developed by Tukey [20], and is implemented using concatenative synthesis techniques similar to Schwarz's implementation [21].

One example of this is the cumulative distribution function. If, for example, a sound recording is analysed for its time-varying loudness, the loudness time series can be re-ordered from the quietest to loudest value. This is sonified by playing the respective audio frames (fragments of the original audio file) from the quietest to the loudest. The same process can be done for pitch, sharpness, or indeed any of the time series outputs of PsySound3.

One purpose of sonifying results in this way is that, unlike visualization, it inherently describes (or explains) the parameter being represented. For example, if sharpness is the parameter under investigation and the user has little experience with the concept of sharpness, the auditory display from least to greatest sharpness demonstrates the concept of sharpness through an experience of the changing sound quality (but only if there is substantial variation in that parameter in the recording being analysed). In addition to explaining the abstract concept of the sonified parameter, this technique also shows much more powerfully than a visual representation how the parameter is associated with particular parts of a sound recording. For example, in a recording of speech, the auditory cumulative distribution of sharpness will allow the user to hear that certain phonemes are associated with greater or lesser sharpness, and indeed how the sharpness of background noise between speech fragments compares to that of the spoken sound. Another advantage of this is that correlates of the parameter (either particular to the recording, or applicable to general situations) can be heard. For example, if there is a correlation between loudness and pitch in a musical recording, an auditory cumulative distribution of one of these parameters will reveal that the other parameter follows it. Alternatively lack of correlation, or inverse correlations, can also be heard using this technique. This technique is particularly suited to psychoacoustical parameters, because, by definition, it should be possible to hear how such parameters vary.

Other exploratory sound analysis techniques are also implemented for univariate, bivariate, time series and spectral data. More details about the concept and implementation of ESA are given by Ferguson and Cabrera [19].

4.3 Further analysis

PsySound3 has the beginnings of a collection of data analysers (as opposed to audio analysers), which perform transformations of data produced by the audio analysers (or indeed by data anlysers). As an example, for rhythm analysis, a Fourier transform or auto-correlation of time series data can be performed. Other processes such as differencing can also be easily performed.

5 Conclusion

Although it is at a relatively early stage of development, PsySound3 provides new opportunities for research and education requiring a 'one-stop shop' for many typical and specialist sound-recording analysis needs. Its flexible, extensible interface will allow it to be further developed and applied by the research community. The availability of multiple analysis models, together with the visualization and sonification options, make PsySound3 an excellent educational tool for students of sound design, music acoustics, psychoacoustics and related disciplines.

PsySound3 is intended as a public domain tool, and so its continued development will depend in part on the interest it generates among the teaching and research community. In addition to providing code for analysers and post-processors, the community may also interact with PsySound through a Wiki site accessed via www.psysound.org.

Acknowledgments

Many people have contributed to PsySound3, including Jens Brosbol, Arturo Camacho, Josef Chalupper, Matt Flax, Dik Hermes, Shin-Ichi Sato, Alex Tarnopolsky and Ernst Terhardt.

References

[1] D. Cabrera, S. Ferguson, E. Schubert, "PsySound3: Software for acoustical and psychoacoustical analysis of sound recordings", *13th International Conference on Auditory Display*, Montreal, Canada (2007) (via www.icad.org)

[2] G.A. Soulodre, "Evaluation of objective loudness meters", *116th Audio Engineering Society Convention*, Berlin (2004)

[3] G.A. Soulodre, M.C. Lavoie, "Development and evaluation of short-term loudness meters", *121st Audio Engineering Society Convention*, San Francisco (2006)

[4] J. Chalupper, H. Fastl, "Dynamic Loudness Model (DLM) for Normal and Hearing-Impaired Listeners", *Acta Acustica united with Acustica* 88, 378-386 (2002)

[5] B.C.J. Moore, B.R. Glasberg, T. Baer, "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness", *J. Audio Eng. Soc.* 45, 224-240 (1997)

[6] J. Chalupper, "Modellierung der Lautstärkeschwankung für Normal- und Schwerhörige", *DAGA 2000*, 254-255 (2000)

[7] B.R. Glasberg, B.C.J. Moore, "A model of loudness applicable to time-varying sounds," *J. Audio Eng. Soc.* 50, 331-342 (2002)

[8] D. Cabrera, "The size of sound: auditory volume reassessed," MikroPolyphonie 5 (1999)

[9] W. Hutchinson, L. Knopoff, "The Acoustical Component of Western Consonance", *Interface* 7, 1-29 (1978)

[10] W. Sethares, *Tuning, Timbre, Spectrum, Scale*, Springer (1998)

[11] P. Daniel, R. Weber, "Psychoacoustical roughness: implementation of an optimized model," *Acta Acustica united with Acustica* 83, 113-123 (1997)

[12] A. Camacho, *SWIPE: A Sawtooth Waveform Inspired Pitch Estimator for Speech and Music*, PhD thesis, University of Florida (2007)

[13] P. Boersma, D. Weenink, "Praat: Doing phonetics by computer (Version 4.4.04) [Computer Program]," Version 4.4.04 ed. vol. (2006) (via www.praat.org)

[14] P. Boersma, "Accurate short-term analysis of the fundamental frequency and harmonics-to-noise ratio of a sampled sound", *Proceedings of the Institute of Phonetic Sciences* 17, 97-110 (1993)

[15] A.D. Cheveigne, H. Kawahara. "YIN: A fundamental frequency estimator for speech and music", J. Acoust. Soc. Am. 111, (2002)

[16] E. Terhardt, G. Stoll, M. Seewann, "Algorithm for extraction of pitch and pitch salience from complex tonal signals", *J. Acoust. Soc. Am.*, 71, 679-688 (1982)

[17] R. Parncutt, *Harmony: A Psychoacoustical Approach*, Springer (1989).

[18] Y. Ando, *Architectural Acoustics*, Springer (1998)

[19] S. Ferguson, D. Cabrera, "Exploratory sound analysis: sonifying data about sound", *14th Int. Conf. Auditory Display*, Paris (2008) (via www.icad.org)

[20] J.W. Tukey, *Exploratory Data Analysis*, Addison-Wesley (1977)

[21] D. Schwarz, *Data-driven Concatenative Sound Synthesis*, PhD thesis, University of Paris 6 (2004)