'PSYSOUND3': SOFTWARE FOR ACOUSTICAL AND PSYCHOACOUSTICAL ANALYSIS OF SOUND RECORDINGS

Densil Cabrera[1], Sam Ferguson[1,2] and Emery Schubert[3]

[1] Faculty of Architecture, Design and Planning, University of Sydney, NSW 2006, Australia
[2] MARCS Auditory Laboratories, University of Western Sydney, Locked Bag 1797 South Penrith, NSW 1797, Australia
[3] Empirical Musicology Group, University of New South Wales, NSW 2052, Australia
densil@usyd.edu.au

ABSTRACT

This paper describes a software project called PsySound3. This software provides an accessible platform for the analysis of sound recordings using procedures applied in acoustics and psychoacoustics. Acoustical analysis methods include a sound level meter module, as well as processes such as Fourier transform, cepstrum, Hilbert transform and auto-correlation. Psychoacoustical models include dynamic loudness, sharpness, roughness, loudness fluctuation, pitch height and pitch strength. Results are presented as numbers, auditory graphs and visual graphs. The software is modular, allowing additional analysis methods to be contributed. Several additional analysis modules are planned. The software is distributed freely via www.psysound.org. This paper illustrates some of the analysis possibilities by using auditory alarms as examples.

[Keywords: Sound analysis, Psychacoustics, Software]

1. INTRODUCTION

The analysis of sound recordings is an important component of research in virtually every discipline of acoustics and audio that involves measurement. Similar tools could be used in fields as diverse as musicology, mechanical services noise, bioacoustics, and auditory display. Often a researcher, teacher, student or enthusiast will find their analysis options limited by either the cost of high quality analysis devices and software, or by their limited time or ability to produce their own analysis software. Fortunately, there is a large community of researchers who have produced software for specialized sound analysis. However, this software tends not to be prepared in an easily accessible or standardized way. This paper describes a project that aims to improve this situation, by providing a freely available and easy-to-use software platform for sophisticated analysis of sound recordings. The software platform, initially released just prior to this conference, is called PsySound3, and it has been developed by the authors, with substantial contributions from several other researchers.

1.1. History of PsySound

The concept of the PsySound project began several years ago with the need of Cabrera to analyze sound recordings using psychoacoustical models for his PhD. At the time there was little software available for this other than very expensive 'sound quality' software. Another problem was that much software that was accessible provided graphical output, with little by way of numeric output. While free software was available from some researchers implementing psychoacoustical models, usually it was not able to read sound files. Written by Cabrera, PsySound1 emerged in 1998 as a suite of computer programs implementing various analysis procedures for direct analysis of sound files: spectrographic analysis (purely physical analysis), loudness and related measures, pitch and related measures, acoustic dissonance, binaural analysis, and post-processing routines. While these programs were useful and flexible, they were not easy to use.

PsySound2 integrated most of the functions of PsySound1 into a single program, providing a much simpler interface [1]. While output data were stored to text files for graphing in other programs, a crude ASCII-based graphical representation of various parameters was also given on screen. PsySound2 has been used for teaching psychoacoustics and sound quality (at least at the University of Sydney and Delft Technical University), for research into acoustical predictors of emotion in music [2], for industrial noise assessment [3], automated sound categorization [4], and modeling of timbral brightness [5]. The final update of PsySound2 was in July 2000 and its operating system (Macintosh 'Classic') has now been superseded.

In the present day, the availability of inexpensive and sophisticated software for acoustical, auditory and audio research has greatly expanded. Current implementations often are capable of sound file input. Prominent instances of such software include: Praat [6], which is oriented to phonetic analysis; STRAIGHT [7, 8] analysis/resynthesis software primarily for voice analysis; Marsyas [9], which extracts various dimensions of music sound files without specifically intending to extract features to which musicians may be accustomed (useful for automatic genre classification); and Aurora [10], which is oriented to impulse response measurement, analysis and inverse filtering.

1.2. Aim of PsySound3

PsySound3 aims to meet a need for a platform implementing psychoacoustical models at an easily accessible level. By 'accessible', we mean that the program should be usable by someone who does not have computer programming experience (for example, someone who has never used Matlab) and also should be inexpensive or free. The benefit of this should be to provide a software tool implementing otherwise inaccessible

models for researchers, educators, students and enthusiasts from a wide range of disciplines, having a wide level of engagement with this type of analysis (from beginning experimenter to in-depth research).

Many other concepts underlie the software platform, and are based on experience in applying psychoacoustical analysis to research problems. The ability to do calibrated analysis easily (for example, with a 94 dB SPL calibration tone on a measurement microphone) is necessary to allow the system to be used for analysing samples recorded with measurement equipment. The validation of implemented analysis modules (through test stimuli that match published output) is essential to be able to rely on the results and understand how they are calculated. Extensibility of the software by users (using the Matlab programming environment to add new analysis modules) means that both the application of the models and their improvement, alteration and comparison with other models is possible within a single environment. This also facilitates a large number of analysis options, including the possibility of multiple models for a single psychoacoustical parameter. The provision of detailed numeric output data for use in other software means that users are able to take the results and use them for whatever purpose they wish (significantly for statistical analysis). A logically sequenced user interface, provides the support for inexperienced users or students to use the program and experiment with psychoacoustical analysis methods without needing programming experience. Similarly, it helps make novice users aware of the basic concepts needed to utilize the sophisticated analysis algorithms successfully (e.g., the necessity for approximate calibration). The sonification and visualization of data (including animations) allows for better understanding of the large output of analysis results. Batch processing based around the logistics of working with audio research datasets assists users working with large numbers of files. Verifiable analysis processes allow the user to both repeat the analysis years after originally performing it, and to be able to step through the analysis process looking for problems.

2. STRUCTURE

The structure of the program is presented in Figure 1. The analysis process begins with the selection of audio files for analysis. These can be in many common formats, as they are converted to a wave file (.wav) prior to being analyzed. Whilst the file format is controlled, neither bit depth nor sampling rate are altered except where analysis algorithms have particular requirements.

The calibration of the signals is the next part of this structure, and there are a number of methods for achieving control over the level of signals to be analyzed. Calibration allows the program to measure sound pressure level (rather than an un-referenced relative level), and is particularly important for some psychoacoustical models because the relationship between sound pressure level and psychoacoustical parameters is not simple. One method for controlling level is to record a microphone calibration signal before recording the signals of interest. As long as the calibration signal can be reliably reproduced before every recording session this method provides precise control over the level of the signals. In acoustical and psychoacoustical contexts this approach is common (but requires special hardware). One alternative, for those who do not necessarily use calibrated recordings, is to standardize the signals being presented in some way. For example, if the user is analyzing telephone ring tones, they may wish to select a mean sound pressure level that is typical for such sounds in everyday environments, and then scale the gain all the recordings to be at this selected mean level. Alternatively, a constant gain adjustment can be chosen for every recording analyzed.

Following the input and calibration stages, the next phase is the digital analysis. For the purpose of comparisons and flexibility we wish digital analysis algorithms to be easily added to the program without the necessity for completely rewriting the program. To achieve this we have set up an architecture into which we can drop in new digital analysis algorithms that have a couple of 'hooks' defined that allow them to be used by the program. We have called these wrapped pieces of digital analysis code 'Analyzers' for the purposes of this program. To integrate into PsySound3, these 'Analyzers' need to achieve five tasks. They need to provide a user interface that gives the user enough descriptive information about the algorithm to be able to employ it successfully. The user interface also needs to provide the means for the user to make specific decisions about whatever analysis parameters can be configured for the algorithm. Once the user has provided the necessary information, the Analyzer needs to read the data from the audio file on the disc appropriately for the analysis it is about to perform. When the data are appropriately configured, the Analyzer sends the data to the appropriate analysis software. Finally, the output of this software is formatted into one or more of the standard data objects. These data objects are explained below in some detail, but it is sufficient to say that the data formats seek to save all relevant information about the analysis process and output in a single object. These objects can then be sent to graphing, export, conversion or other analysis procedures, so that their output can be used in many common ways.

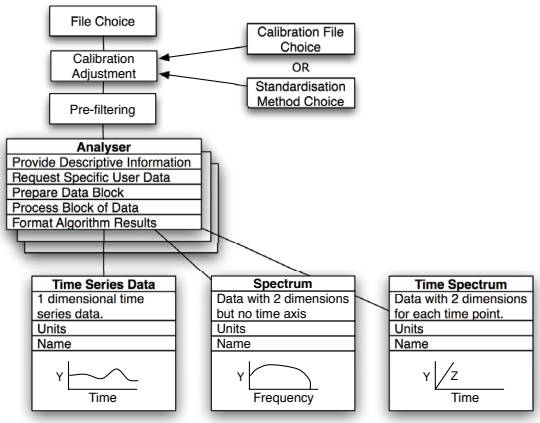

Figure 1. *The program takes files that are calibrated in some way, and processes them with a set of 'Analyzers' yielding acoustical or psychoacoustical data formatted in one or more of the three major data formats. These can then be graphed, exported or used in various other ways.*

3. ANALYZER OVERVIEW

This section outlines many of the currently implemented analysis functions, as well as functions that are planned for implementation in the near future. Where examples are given, the sounds are from a set of auditory alarms for air traffic control consoles [11]. The sound recordings used are for a simulation of the auditory alarm at the operator's head position.

3.1. Physical Signal Analysis

In a sense, every measure produced by PsySound3 is a physical measure, as it is derived algorithmically from the physical waveform, and has no subjective component. However the term 'physical measure' is used more narrowly here to refer to those measures that make little or no attempt to emulate subjective response. Such measures are generally simply defined, and their derivation is easy to comprehend. Physical signal analysis is an important part of psychoacoustical analysis because: (i) it allows signals to be calibrated; (ii) it allows signals to be described in well-understood standard ways (so that the signals can be reconstructed by others); (iii) processing the physical signal can be more efficient than psychoacoustic models (which may require significant additional computation, sometimes for marginally improved output); and (iv) physical signal measurements can be compared with the output of psychoacoustical models (for example, comparing loudness to A-weighed sound pressure level, sharpness to spectral centroid, or pitch to frequency).

3.1.1. Sound Level Meter and 1/3-Octave Band Analysis

These two analyzers provide a software implementation of the functions normally built into a precision sound level meter. They implement A, B, and C-weighting filters, and 1/3-octave band filters in the time domain, and perform 'fast' (125 ms) or 'slow' (1000 ms) temporal integration of the filtered and rectified signals, using leaky integrators. Octave band output is taken from the sum of the three constituent 1/3-octave bands. While this approach is a little more involved than fast Fourier transform based analysis, the analyzers' frequency and time response should match those of a standard sound level meter.

Values that can be obtained using this analyzer include the filtered signals (prior to rectification and integration), time series sound pressure levels (Fig 2), equivalent sound pressure level, and percentile distributions. This analyzer is also used internally by the program to calculate the level of the calibration files that are used to calibrate the files to be analyzed.

3.1.2. Other Physical Signal Analyses

PsySound3 implements analyses that are often associated with the field of digital signal processing, namely Fourier transform, cepstrum, Hilbert transform and auto-correlation.

Values that can be output from the Fourier transform include the time varying spectrum, long-term average power spectrum, and spectral moments (both time-varying and overall). Similar output is available from cepstral analysis, which can be useful in identifying fundamental frequencies of harmonic series present in the input signal. The Hilbert transform may be used to output the Hilbert amplitude envelope (Fig 2) and instantaneous frequency, both as time series. The auto-correlation function provides a time domain approach to periodicity analysis.

3.2. Loudness and Related Analysis

3.2.1. Loudness Models

There are many models for predicting the subjective sensation of loudness. In recent decades loudness models used in psychoacoustics can be divided between ones using Bark auditory filters (or critical bands) and Erb auditory filters, and also between steady state and dynamic models. The Bark and Erb scales are similar in concept, except that Erb auditory filters have narrower bandwidths than Bark filters, and the filter distribution differs particularly in the frequency range below 500 Hz. Steady state models account for spectral effects on loudness (for example, that greater bandwidth can yield substantially greater loudness for signals of identical sound pressure level). However, dynamic models also account for the effect of auditory temporal integration on loudness, and so are better suited for time-varying signals, especially when fine temporal detail is of interest. Dynamic models are more complex and computationally intensive.

PsySound2 used a steady state loudness model [12] to analyze successive 93 ms Hanning-weighted blocks. PsySound3 is much more flexible by providing both dynamic and steady state models using either Erb or Bark auditory filters. It implements the dynamic loudness models of Glasberg and Moore [13] (using Erbs) and Chalupper and Fastl [14] (using Barks). In the latter case, code was provided by Chalupper, and modified slightly to conform to the requirements of PsySound3. Steady state models are included in PsySound3 for the purpose of comparison. Zwicker's model (which is standardized in ISO532B) is included, along with Moore, Glasberg and Baer's model [12] (which was used in PsySound2, and on which ANSI S3.4-2005 is based [15]). Examples of loudness measurements are given in Figures 2, 3 and 5 and Table 1.

3.2.2. The Specific Loudness Pattern

Specific loudness is the loudness attributable to an auditory filter. The specific loudness pattern might be likened to a magnitude spectrum based on loudness. A psychoacoustical 'frequency' scale (either in Erbs or Barks) accounts for the distribution of sound in the cochlea (based on the characteristic frequencies of auditory filters), and the unit of specific loudness is sones per Erb or sones per Bark (Figs 3 & 5). When read directly, the specific loudness pattern can indicate the parts of the frequency spectrum that make the strongest contribution to loudness, and can also indicate the extent of masking. The specific loudness pattern also contributes to several higher level measures – most notably, loudness is its integral and sharpness is based on its centroid.

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 [16] and Lichte [17]. PsySound implements the sharpness models of both Aures [18] and Zwicker and Fastl [19] (Fig 4, Table 1). 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.

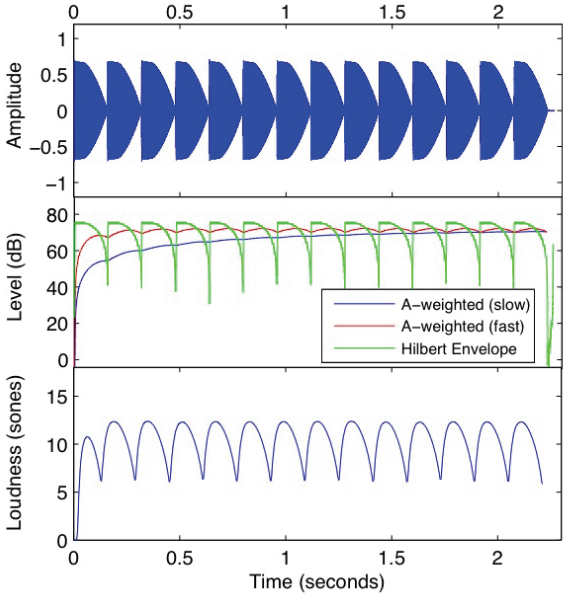

Figure 2. *Time series representation of a high priority auditory alarm used in air traffic control consoles. The top chart shows the waveform (or at least its apparent envelope, because details of the waveform are not visible on this scale). The middle chart shows the A-weighted sound pressure level using slow (1 s) and fast (125 ms) integration times, as well as the un-weighted Hilbert envelope (which responds almost instantly to the input waveform). The bottom chart shows dynamic loudness (using the model of Chalupper and Fastl), and the first pulse is modeled as slightly quieter than the subsequent ones because it is preceded by silence.*

3.2.3. Loudness Fluctuation

'Fluctuation strength' refers to the subjective sensation of the strength of fluctuations in sound. Limited models of fluctuation strength are presented by Zwicker and Fastl [19]. Modulation frequencies around 4 Hz make the strongest contribution to fluctuation strength, with very little contribution for frequencies less than 0.5 Hz. Fluctuation strength increases with modulation depth, reaching saturation at 30 dB. The code for loudness fluctuation (a general model of fluctuation strength accounting for amplitude modulation, but not frequency modulation) was supplied by Chalupper [20]. This model was specifically designed and evaluated for fluctuating broadband and real life sounds. Example results are in Table 1.

Overall values for loudness and sharpness are better derived from percentile analysis than by averaging over time. The short term loudness value exceeded 5% of the time (i.e., the 95th percentile) is often used for overall loudness [14, 19]. PsySound3 allows percentiles to be calculated for individual time series (such as loudness) and also for each component of the specific loudness pattern.

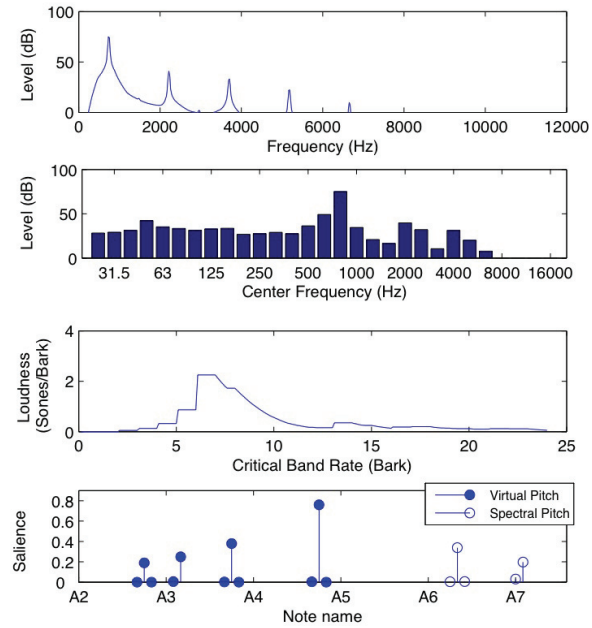

Figure 3. *Examples of the spectral data type derived from the high priority air traffic control console auditory alarm. The top chart is a fast Fourier transform magnitude spectrum, in which a harmonic series is clearly visible. The second chart is the 1/3-octave band spectrum, showing substantial low frequency noise, which the short time window of the FFT was not sensitive to. The third chart is the specific loudness pattern, showing the loudness attributable to auditory filters from low to high Bark values. The bottom chart is the chromatic pitch pattern quantized from Terhardt's pitch model, showing both spectral and virtual pitches.*

3.3. Roughness and Dissonance

Roughness is a sensation associated with rapid fluctuations in the sound received by auditory filters. As such, it can be caused by beats between tone components, amplitude modulation, and frequency modulation, with peak roughness sensitivity being for modulation around 70 Hz in the mid and high auditory filter frequency range. Models of roughness for simple stimuli are given by Zwicker and Fastl [19]. A model of roughness applicable to arbitrary stimuli was developed by Aures [21] and optimized by Daniel & Weber [22]. This model is implemented in PsySound3, using code provided by Dik Hermes.

The concept of 'acoustic dissonance' has been used in musicology and musical acoustics to explain part of the concept of musical dissonance [23-25]. Acoustic dissonance models predict the roughness due to tone-pair interactions in the frequency domain – hence acoustic dissonance can be considered to be a subset of roughness. PsySound2 implemented models of acoustic dissonance: 'spectral dissonance' used all Fourier components, while 'tonal dissonance' used a peak

extraction algorithm prior to dissonance calculation. PsySound3 offers users the opportunity to compare acoustic dissonance and roughness model results. Hutchinson and Knopoff's model of acoustic dissonance [23] is implemented following peak extraction (using the same peak extraction algorithm as for Terhardt's pitch model [24]).

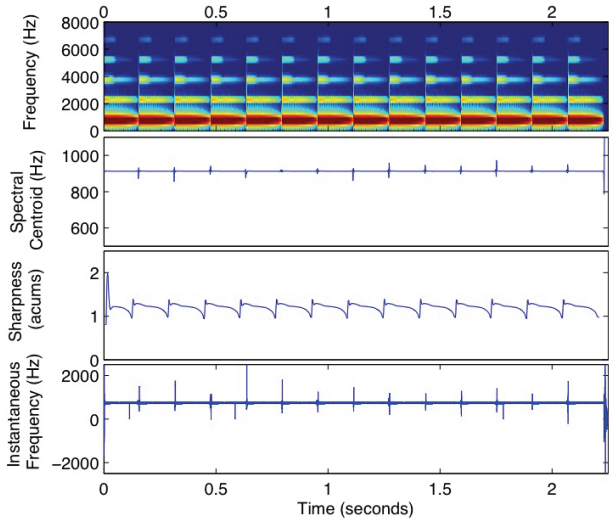

Figure 4. *Spectrographic and time series representation of a high priority auditory alarm used in air traffic control consoles. The top chart shows the FFT spectrogram in which the harmonic content of the pulsed tone can be seen. The second chart shows the spectral centroid (derived from the spectrographic data), which scarcely varies during the duration of each pulse due to the dominance of the fundamental frequency component (a magnitude centroid would vary more –a power centroid is shown). The third chart shows time-varying sharpness, which gives a much clearer representation of how timbre varies with time. The bottom chart shows instantaneous frequency, derived from the Hilbert transform.*

3.4. Pitch Analysis

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 [26] has developed much more sophisticated models of the structure of 'pitch space', accounting for pitch height, octave similarity and the cycle of fifths.

The pitch model of Terhardt *et al.* [27] has been used in all versions of PsySound, primarily because of its track record in music analysis, especially in the work of Parncutt [28]. This relatively simple model (based on frequency domain template matching of harmonic series rather than auto-correlation), 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). The implementation of these measures in PsySound extends the application of Parncutt's model to the analysis of sound (whereas Parncutt restricted his attention to the analysis of 12-tone equal temperament pitch categories). PsySound3 quantizes the results by default 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 quantization. Pitch salience patterns are expressed linearly over the pitch height range, and circularly over the chroma range. Figs 3 and 5 give examples of pitch results.

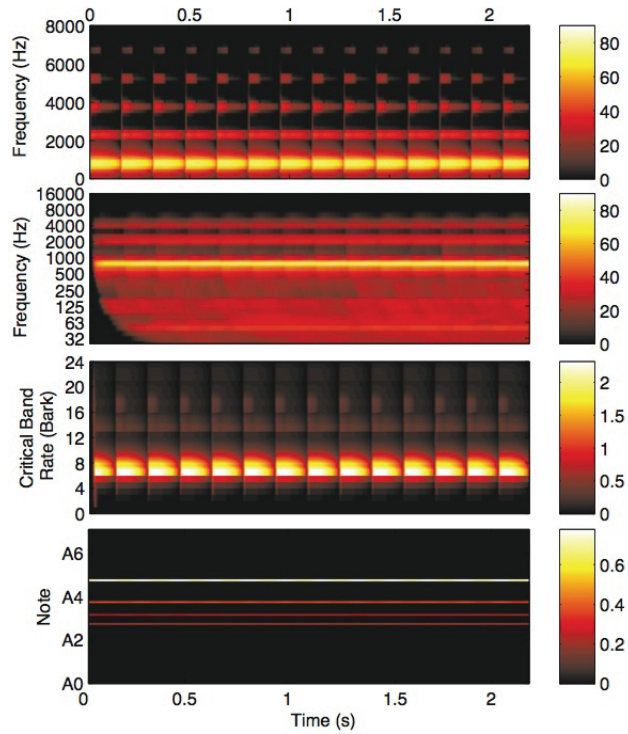

Figure 5. *Spectrographic and time series representation of a high priority auditory alarm used in air traffic control consoles. The top chart shows the FFT spectrogram in which the harmonic content of the pulsed tone can be seen. The second chart shows the 1/3-octave band spectrogram (at the time of writing, low frequency filters have greater latency than high frequency filters, but we plan to change this to constant latency). The third chart is the time-varying specific loudness pattern (units are sones per Bark). The final chart is the time-varying pitch analysis (although in this case the pitch pattern is constant despite the substantial changes in sound pressure level and loudness).*

3.4.1. Harmonic Content

For 12-tone equal temperament musical sound, PsySound2 calculated the most likely tonic of the key (using long term average pitch weights) and chord (using short term pitch weights), and these functions will soon be implemented in PsySound3. Parncutt [28] presents chroma profiles for the 24 major and minor keys based on octave spaced tones (Shepard tones). PsySound calculates the Pearson correlation coefficient (r) of the chroma salience pattern to each of these key profiles, and the one with the highest coefficient is output. The r^2 value is also output - a value close to unity indicates a close correlation, while a low value suggests that the key of best fit fits poorly.

Chroma profiles of 27 common chords (the monad, and every dyad, triad and tetrachord) have been calculated for octave spaced tones. With 12 transpositions, this makes 324 chords (except that some are duplicated when transposed). The same correlation-based matching procedure is used to determine the most likely chord.

3.5. Articulation

Articulation is of interest to musicians and indicates the amount of time for which a musical note is sounded with respect to the note interonset interval. A note held up until the commencement of the next note is said to have legato articulation, whereas a note sounded for a relatively short time with respect to the interonset interval is usually said to be played with staccato articulation. Articulation calculations can be performed using PsySound3 based on the Average Silence Ratio [29]. However, subjective articulation is generally more complex due to perceptual non-linearities and effects of 'key overlap time' [30] and future versions will implement systems based on algorithms with a more statistical approach, such as that described by Brosbøl and Schubert [31].

3.6. Binaural Analysis

Spatial analysis of sound recordings is not currently implemented, but will be done assuming binaural recordings. A stereo-to-binaural converter may be used to convert recordings that would normally be listened to in 2-channel stereo. This is simply a 2x2-channel convolution (left and right loudspeakers to left and right ears) using dummy head head-related transfer functions for azimuths of ±30º. The binaural analysis will consist of cross correlation between the signals of the two ears for each auditory filter (with lags of ±1 ms). The peak height of this function may be used to indicate apparent source width, and the lag time of the function may be used to indicate lateralization. Interaural level differences will be calculated. PsySound2 implemented binaural analysis in this way, but without any spectral analysis.

4. RESULTS ANALYSIS

4.1. Data Format

The use of the data that are gained from each of the algorithms presents particular challenges. To avoid the laborious process of reformatting, cleaning and re-labeling data before its use, we are following an object-oriented approach to data storage. Each analyzer's output is formatted in a data object before being saved. These objects allow all the relevant pieces of information about the data to be saved in a single location and in a standard hierarchical format. For example, such information may include the name of the algorithm or the units its output is measured in, choices the user has made about the algorithm's process, or even methods or functions for performing tasks like averaging. Also, non-analysis information, such as independent parameters that are being varied between samples, may be stored within the data object. By storing all this information in a single location we reduce the user's reliance on their records or memory to successfully repeat the data analysis process.

Data objects are defined relatively abstractly, to allow maximal reuse of functions that accept these objects as arguments. An advantage of this object-oriented approach is that that functions may be defined that respond in the most appropriate way for each data object type (also called 'overloading' a function). This makes it easy to define a new algorithm and its output and still have access to the full set of analysis, graphing and export functionality that may be applied to similar algorithms that are already included in the program. Similarly, new analysis, graphing or export functionality may be added to the program without needing to address each specific algorithm output individually.

There are currently 3 main data objects. The 'time series' object is the simplest object, and stores a single value that changes with time (for instance broadband A-weighted sound pressure level). A 'spectrum' object accepts data that is two dimensional in nature, but is not time varying (e.g. one-third octave band levels). A 'spectrum' object, may not necessarily be a spectrum in the defined sense, but is named this way in order to signify its most likely use. A 'time spectrum' object is the third major type of object, and defined for spectral data that changes over time (e.g. spectrogram data). With three main data objects we provide frameworks for storing the outputs of all the algorithms we intend to include at this stage. By simplifying and codifying the possibilities for data storage it is hoped that less errors will be caused during the results analysis process.

4.2. Data Compression/Thinning

When there is a large number of long files to be analyzed, and some analyses (e.g., spectral analysis) may yield dimensional enlarging of the data size, it is very important to build methods for thinning the data down to a practical size. Also, in many situations high-resolution data is often no more useful than low-resolution data is. This thinning is achieved before the data objects are saved by using a user specified down-sampling ratio (incorporating anti-aliasing filters), and a window and time step to step through the data and group (average) blocks of data together. Another method used to reduce the data quantity is to choose an upper frequency limit within the output format. For example, by omitting the upper two octaves of a full bandwidth linearly distributed spectral analysis, the data quantity is reduced to one quarter of its original size, whilst retaining most of the most useful data.

4.3. Data Object Processing

There are a number of procedures that can be undertaken on each of the data objects outlined above. They can be summarized by the following categories:

Export: Exporting data to a text file to be analyzed in another program.

Graphing: Visualizing the data in a graphical format.

Data Conversion: Changing the data from one data object to another, usually by reducing its dimensionality. This includes averaging results down to single number ratings of various aspects of a sample using various descriptive statistics methods.

Sonification: Whilst visual graphing dominates data analysis in acoustics, sonification is an excellent alternative, especially for this type of data.

Spectral and Time Series Analysis: Much of this data is well-suited to spectral and time-series analysis. Thus, while PsySound3 does not aim to be a fully featured statistics tool, it is worthwhile noting that many spectral options are available to process the time series data objects. These may include options such as percentile distributions, statistical moments, fast Fourier transform, cepstrum, auto-correlation, and cross-correlation functions.

5. APPLICATIONS IN AUDITORY DISPLAY

PsySound3 has many possible applications, some of which are in auditory display. One possibility is the analysis of existing auditory displays (such as auditory alerts) in terms of their physical and psychoacoustical characteristics. Such an analysis might be helpful in improving the consistency or contrast between sets of auditory display sounds (such as auditory alerts), or in examining masking and streaming effects for multiple sounds. Table 1 gives an example of psychoacoustical measurements of four auditory alarms used in air traffic control consoles. Although these alerts were not designed using psychoacoustical models, the measurements show a general tendency for the scale values to increase with urgency.

Table 1. *Selected psychoacoustical measurements (averaged over time) of four auditory alarms used in air traffic control consoles, representing four levels of urgency (1 is highest, 4 is lowest).*

Urgency	Loudness (sone)	Loudness Fluctuation (vacil)	Sharpness (acum)
1	7.45	1.77	1.16
2	5.42	1.50	0.92
3	3.13	1.37	0.69
4	3.39	0.90	0.59

Sensory pleasantness may be a desirable attribute of auditory displays. The field of 'sound quality' develops models of sensory unpleasantness (usually for appliances and machinery noise) based on psychoacoustical models, which typically combine loudness, sharpness, roughness and tonalness [32]. A similar approach might be taken to the refinement of auditory display sounds.

As proposed previously by the authors [33], auditory graphs can be developed based on psychoacoustical parameters, instead of simple signal parameters. This is not done by directly inverting psychoacoustical models (which, for arbitrary signals, is not possible), but by generating a large set of parametrically defined stimuli that are measured using psychoacoustical models, and then the information from that is used for an indirect model inversion. The requisite matrix of stimulus parameter values can be developed using PsySound3. A similar approach could be taken in alert design. Design principles for hierarchical alert schemes currently use simple signal properties such as pulsation rate, tone fundamental frequency, inharmonicity, and so on [34]. An alternative approach could be to use psychoacoustical signal parameters to control urgency, although research is required to determine the relationship between such parameters and perceived urgency.

It is desirable in some psychophysical experiments in auditory display to have stimuli equal in terms of psychoacoustical parameters (eg equal loudness, equal sharpness etc). PsySound3 can be used to do this through iterative measurement, analysis and adjustment of each stimulus. Finally, PsySound3's sonification of analysis data is itself an auditory display application. This idea of sonifying data pertaining to sound is discussed further by the authors elsewhere [35] (if accepted).

6. FUTURE DIRECTIONS

Future versions of PsySound will include modules for music analysis, such as vibrato and trills [36], beat mapping and tempo curves [37]. Furthermore, initial part extraction of voices [38] will allow separate analysis of each part from a polyphonic sound recording using some of the analysis modules. This will include stream segregation of the kind described by Temperley [39]. Drawing results of multiple analysis modules together could be a kind of computational auditory scene analysis, and this may be considered in the future. The open source architecture will facilitate the flexibility of the program, and therefore PsySound is likely to see many further augmentations.

7. ACKNOWLEDGMENTS

Many people have made significant contributions to this project, including Jens Brosbøl, Josef Chalupper, Peter Daniel, Dik Hermes, Matt Flax, Alex Tarnopolsky and Ernst Terhardt. There are also many others who have contributed useful feedback and suggestions. This project was funded by the Australian Government through the Australian Research Council LIEF scheme.

8. REFERENCES

[1] D. Cabrera, "`PsySound': A computer program for the psychoacoustical analysis of music," *MikroPolyphonie*, vol. 5, 1999.

[2] E. Schubert, "Modeling perceived emotion with continuous musical features," *Music Perception*, vol. 21, pp. 561-585, 2004.

[3] P. Susini, S. McAdams, S. Winsburg, P. Ivan, S. Viellard, and X. Rodet, *Characterizing the sound quality of air-conditioning noise*, report, 2007.

[4] J. Ricard, "Towards computational morphological description of sound," Barcelona: Universitat Pompeu Fabra 2004.

[5] E. Schubert and J. Wolfe, "Does timbral brightness scale with frequency and spectral centroid?," *Acta Acustica United with Acustica*, vol. 92, pp. 820-825, 2006.

[6] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer (Version 4.4.04) [Computer Program]," Version 4.4.04 ed. vol. 2006, 2006.

[7] H. Kawahara, "STRAIGHT, exploitation of the other aspect of VOCODER: Perceptually isomorphic decomposition of

speech sounds," *Acoustical Science and Technology*, vol. 27, pp. 349-353, 2006.

[8] H. Kawahara, I. Masuda-Katsuse, and A. d. Cheveigne, "Restructuring speech representations using a pitch-adaptive time- frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a reptitive structure in sounds," *Speech Communication*, vol. 27, pp. 187-207, 1999.

[9] G. Tzanetakis and P. Cook, "Marsyas: A framework for audio analysis," *Organised Sound*, vol. 4, 2000.

[10] A. Farina, "Aurora [Computer Program]," 2006.

[11] D. Cabrera, S. Ferguson and G. Laing, "Considerations arising from the development of auditory alerts for air traffic control consiles," in *Proc. 12th Int. Conf. Auditory Display*, London, Great Britain, 2006.

[12] B. C. J. Moore, B. R. Glasberg, and T. Baer, "A model for the prediction of thresholds, loudness, and partial loudness," *Journal of the Audio Engineering Society*, vol. 45, pp. 224-240, 1997.

[13] B. R. Glasberg and B. C. J. Moore, "A model of loudness applicable to time-varying sounds," *Journal of the Audio Engineering Society*, vol. 50, pp. 331-342, 2002.

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

[15] ANSI S3.4-2005 *American National Standard – Procedure for the Computation of Loudness of Steady Sounds*. Melville NY: Acoustical Society of America, 2005.

[16] E. G. Boring and S. Stevens, "The nature of tonal brightness," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 22, pp. 514-521, 1936.

[17] W. Lichte, "Attributes of complex tones," *Journal of Experimental Psychology*, vol. 28, pp. 455-480, 1941.

[18] W. v. Aures, "Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale," *Acustica*, vol. 59, pp. 130-141, 1985.

[19] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*. Berlin; New York: Springer, 1999.

[20] J. Chalupper, "Modellierung der Lautstärkeschwankung für Normal- und Schwerhörige," in *DAGA 2000*, 2000, pp. 254-255.

[21] W. v. Aures, "Ein Berechnungsverfahren der Rauhigkeit," *Acustica*, vol. 58, pp. 268-281, 1985.

[22] P. Daniel and R. Weber, "Psychoacoustical roughness: implementation of an optimized model," *Acustica*, vol. 83, pp. 113-123, 1997.

[23] W. Hutchinson and L. Knopoff, "The Acoustical Component of Western Consonance," *Interface*, vol. 7, pp. 1-29, 1978.

[24] R. Plomp and W. J. M. Levelt, "Tonal consonance and critical bandwidth," *Journal of the Acoustical Society of America*, vol. 38, pp. 548-560, 1965.

[25] W. Sethares, *Tuning, Timbre, Spectrum, Scale*. London: Springer, 1998.

[26] R. Shepard, "Geometrical approximations to the structure of musical pitch," *Psychological Review*, vol. 89, pp. 305-33, 1982.

[27] E. Terhardt, G. Stoll, and M. Seewan, "Algorithm for extraction of pitch and pitch salience from complex tonal signals," *Journal of the Acoustical Society of America*, vol. 71, pp. 679-688, 1982.

[28] R. Parncutt, *Harmony: A Psychoacoustical Approach*. Berlin: Springer, 1989.

[29] Y. Feng, Z. Yueting, and P. Yunhe, "Music information retrieval by detecting mood via computational media aesthetics," in *Proceedings of the IEEE/WIC International Conference on Web Intelligence (WI'03)* 2003.

[30] B. H. Repp, "Acoustics, perception, and production of legato articulation on a digital piano," *Journal of Acoustical Society of America*, vol. 97, pp. 3862-3874, 1995.

[31] J. Brosbol and E. Schubert, "Calculating articulation in solo music performances," in *9th International Conference on Music Perception and Cognition*, Alma Mater Studiorum University of Bologna, 2006, pp. 1362-1366.

[32] H. Fastl, "The Psychoacoustics of Sound Quality Evaluation," *Acustica - Acta Acustica*, vol. 83, pp. 754-764, 1997.

[33] S. Ferguson, D. Cabrera, H.-J. Song, and K. Beilharz, "Using psychoacoustical models for information sonification " in *Proceedings of the 12th International Conference on Auditory Display*, London, UK, 2006.

[34] E. J. Hellier, J. Edworthy, and I. Dennis, "Improving auditory warning design: Quantifying and predicting the effects of diffferent warning parameters on perceived urgency," *Human Factors*, vol. 35, pp. 693-706, 1993.

[35] D. Cabrera and S. Ferguson, "Sonification of Sound: Tools for teaching acoustics and audio," in *Proceedings of the 13th International Conference on Auditory Display*, Montreal, Canada, 2007.

[36] J. C. Brown and P. Smaragdis, "Independent Component Analysis for Automatic Note Extraction from Musical Trills," *Journal of Acoustical Society of America*, vol. 115, pp. 2295-2306, 2004.

[37] S. Dixon, "Automatic extraction of tempo and beat from expressive performances," *Journal of New Music Research*, vol. 30, 2001.

[38] P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2003.

[39] D. Temperley, *The Cognition of Basic Musical Structures*. Cambridge, Mass: MIT Press, 2001.