Seeing Sound : Real-time Sound Visualisation in Visual Feedback Loops used for Training Musicians

Sam Ferguson*, Andrew Vande Moere°and Densil Cabrera*
*School of Architecture, Design Science and Planning, The University of Sydney
°Key Centre of Design Computing and Cognition, The University of Sydney
samferguson@ihug.com.au, andrew@arch.usyd.edu.au, densil@arch.usyd.edu.au

Abstract

Musicians in training need to understand the sound they are producing in order to improve its deficient aspects. Verbal feedback from musical masters is the usual method used for attaining this understanding. However, using real-time sound visualisation as a complementary form of feedback allows the large amounts of data typical of real-time acoustic analysis to be employed within training. This improves the efficiency of the feedback loop normally present within musical training and pedagogy. The implementation and effect of such a system is discussed.

Keywords—**Sound visualization, real-time visualization.**

1 Introduction

Musicians in training have a difficult problem. They must extend their skills as musicians, often without being able to rely on their personal aural appraisal of sound quality, tuning and loudness. This means it is difficult for them to distinguish between effective and ineffective training. Most musicians therefore seek verbal advice from a master of their particular instrument, who has more experience with appraising musical sound, locating improvement and specifying effective types of training for these students.

Verbal feedback is the primary tool used in instrumental and vocal pedagogy. It is the basis of many years of music tuition and is a very flexible pedagogic method in the hands of talented pedagogues. However, the number of data-points practically transferable using verbal feedback is limited compared to that of information visualisation. If we wish to communicate several specific pieces of information about the sound being produced by the musician, all of which change several times per second, then clearly verbal feedback is ill-suited to this information transfer task. By contrast, real-time acoustic analysis can achieve this rate of information supply, and an effective real-time visualisation can both communicate it and elucidate the patterns within it to the musician (Figure 1).

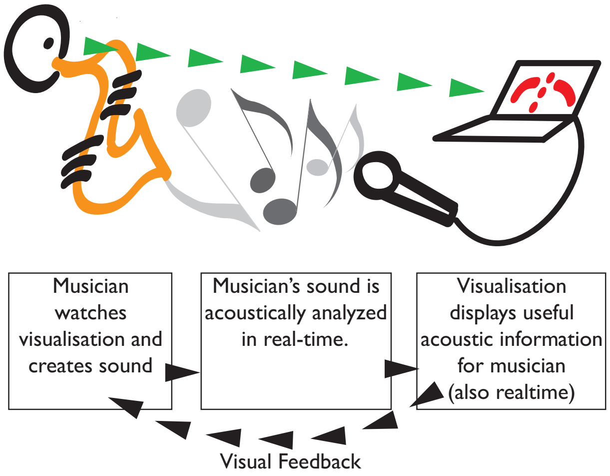


Figure 1: Visualisation helps musicians understand acoustic parameters of their sound. The display assists comprehension of multiple parameters, allowing instant changes to sound production methods to be made when deficiencies are interpreted.

Visualisation has been applied to sound previously for various purposes, but usually without utilising the principles of information visualisation. Visual integration of the various sound characteristics is essential for a useful display, so that the musician can build a complete conceptual understanding of their sound from the data. The authors are unaware of any other tool that attempts this aim.

This research documents the design and implementation of such a tool for instrumental musicians in training. A preliminary evaluation of its effectiveness is also provided.

2 Background

Information visualisation is concerned with navigating and finding patterns within large, often static, datasets. Dealing with real-time data visualisation is a different type of problem. If this visualisation's purpose were solely recognising patterns within a musician's sound, the dataset could be recorded and any one of many statistical visualisations could be applied in post-processing. However, for a musician this information would not be useful, as the various mouth, throat and other muscle positions affecting the sound would be forgotten by the time they were able to see the patterns.

A real-time visualisation's instantaneous feedback allows students to employ the information perceived about their sound in correcting or reinforcing their sound production methods. The visualisation must clearly show variables that change in time, and also allow understanding of the rate and trend of this change. It also needs to describe the *relationships* between each variable visually, allowing them to be understood in context.

2.1 Previous work

Interest in visualising sound is not new, and has existed at least since Lissajous and Helmholtz in the 19th century [1]. These pioneers developed systems to analyse and represent sound visually that were either optical or mechanical, proving quite difficult to devise. Modern computing power has increased the ease of both graphical representation and acoustic analysis. It has also allowed the two to be developed separately and connected arbitrarily, rather than requiring the engineering of an integrated machine.

Visualisations for singers are already available as commercial products, and have been researched for some time. One of the earliest attempts at visualisation for training singers was the 'SINGAD' system implemented by Welch [14]. This system was able to facilitate noticeable efficiency improvements in teaching children the ability to sing with control over their pitch [13].

Sing and See [4] is a more recent program that uses real-time spectral displays, metering and traditional notation to provide visual feedback for singing pedagogy. Enthusiastic responses to the software were received when a qualitative investigation was conducted in pedagogical situations [3].

Within speech pathology visualisation of acoustic measurements has been used widely. This is often useful for young patients, as treating speech deficiencies can be approached in the form of interactive computer animations using patient speech as a control mechanism (for instance *Speechviewer III* [9]).

Finally, sound visualisation has also been employed in providing aural awareness for hearing-impaired persons [6]. A visualisation was implemented that appeared in the corner of the computer display used in an office environment. After two weeks using this system a respondent could detect and distinguish speech, mobile phone calls, chair movement, typing, mouse movement, page turning, papers rustling, footsteps of people entering the office and even a truck that visited the workplace carpark once weekly.

2.2 Feedback loops and musicians

Music education is normally carried out using a master-student framework in single-student or small class lessons.

In these lessons an iterative feedback loop is used for the tuition in which; a) the student performs b) the master interprets the performance c) the master provides verbal feedback d) the student interprets the verbal feedback and gains Knowledge of Results (KR) e) this process is repeated (Figure 2) [11]. This loop is also present in the student's personal practice, except that the student mentally interprets their own performance rather than having the benefit of the master's appraisal.

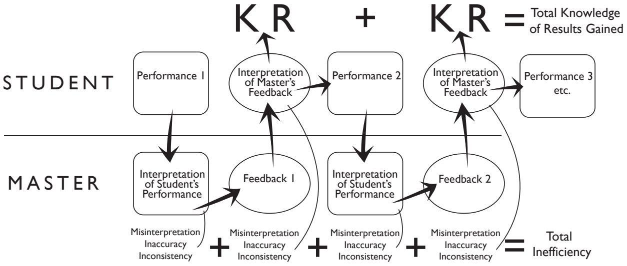


Figure 2: The process of private music lessons and personal musical practice can be described as a feedback loop.

Two processes of interpretation are relied on in this feedback loop, both of which can be inconsistent depending on the skills of master and student. Furthermore, feedback is usually provided after the student performs, meaning that it can be difficult for the student to relate muscle movements and sound production methods to their musical results, based only on their memory of the performance. Unfortunately these problems undermine the efficiency of pedagogy with verbal feedback.

Using real-time visualisation circumvents many of these problems. The repeatability of acoustic analysis strongly decreases the reliance on subjective interpretation skills for primary aspects of sound production (e.g. tuning, loudness, sound quality). The real-time nature of the visualisation also means that it is possible to receive instant feedback, allowing the student to relate production method to musical result with increased certainty. This instant feedback also allows the feedback loop to iterate at a much higher rate, greatly increasing the speed at which the student receives KR.

3 Implementation

Implementation of our visualisation involved i) acquiring and filtering the data to provide only the relevant data attributes, ii) considering the requirements and visual associations of training musicians, and iii) employing appropriate visual metaphors for the visualisation.

3.1 Data

This visualisation problem consists of integrating salient attributes of the data into a cohesive visual representation. Visualising only the data that benefits the musician when supplied in real-time or very recent time assists in simplifying the representation.

The primary data source consists of a digitised audio stream, which is then analysed using Fourier transform based methods to extract parameters that describe specific aspects of that stream. These parameters include :

Harmonic Content: The harmonic content of a note is what determines, to a great extent, its ‘sound character’. We used Puckette et al’s analysis system ‘fiddle∼’ [10] to obtain the magnitude and frequency of the first four harmonics within the sound.

Noisiness: The comparison between periodic and random elements within the data stream can be described as a ratio and also assists in describing the ‘sound character’. This algorithm was first described by Johnston [8], and this instance was implemented by Jehan [7].

Loudness: The strength of the instrument’s sound pressure level as perceived by the listener is an essential parameter for providing the musician with information regarding their production of contrasts between loud and soft. By using psycho-acoustical *loudness* ratings, a better approximation of the *perceptual effect* is made. The algorithm used for this is described by Jehan [7].

Fine Pitch: The pitch of the note can be extracted relatively precisely. By comparing this pitch to the western musical scale we can measure any discrepancy. This data is again based on the pitch-picking abilities of ‘fiddle∼’.

These four data streams can often be related to each other as a by-product of the acoustic properties of the instrument or production mechanism. For example, an increase in loudness can also be accompanied by a change the harmonic content of the sound. They can also sometimes be totally separate. Therefore, simultaneous perception of all parameters is necessary.

3.2 Design rationale

Designing a sound visualisation for facilitating musical training has included considering the integration of prioritised information, using time-variance wisely and understanding the way in which a musician would mentally visualise sound.

In a real-time display of information that is constantly changing, it is impossible to visually scan the smaller aspects of the visualisation to reveal hidden detail. If the visualisation features many objects that need to be perceived separately, then only one of these objects can be understood at once, and other objects must be ignored. Therefore it is preferable that a visualisation feature a single object that has several perceivable attributes, rather than several different objects.

The real-time nature also affects the use of time as an axis. It is possible to provide a scrolling display of data over a recent time-span. This has the advantage of allowing comparison over time without relying on the viewer’s memory. However, this element also has the capacity to distract the attention from the instantaneous information. By placing the display of this information on the periphery, attention remains on the instantaneous information.

Relating the information display to the way that musicians are likely to mentally visualise sound is advantageous. This establishes a connection with the musician on a more associative conceptual level, and decreases the necessity to consciously interpret the display. Musical sound visualisation of a mental nature is heavily subjective, and relies implicitly upon the specific style of music concerned. However, certain associations can be relied on for some generality across across cultural groupings and musical style boundaries. Walker [12] has shown that when presented with simple tone groups, and given visual analogies on cards, the choices made followed previously understood patterns. Size was associated with amplitude, frequency with vertical position, duration with horizontal length, and waveform with pattern. It is not immediately apparent why these analogies may be anything but associative, but for some attributes of sound visualisation physical perceptual elements may have a role in sustaining these associations (see for instance [2]).

3.3 Technical implementation

Figure 3 shows the stages by which the program processes the primary data source, and the data mapping rules.

Implementation was completed in the Max/MSP environment using Jitter to program the graphical output [5]. Max/MSP and Jitter are oriented towards processes rather than data, and thus deal with real-time visualisation well.

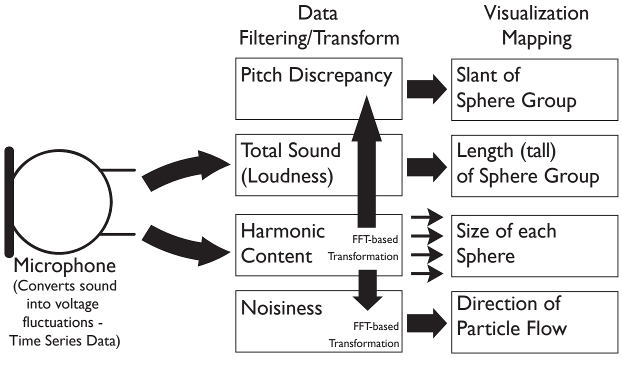


Figure 3: A schematic diagram of primary data source filtering and data mapping rules.

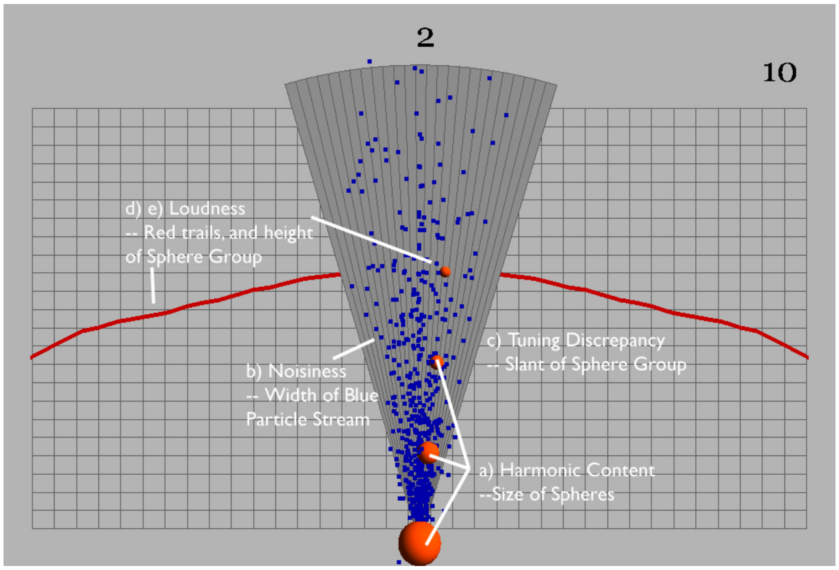


Figure 4: The visualisation display prototype.

Sophisticated signal processing extensions to the Max/MSP environment implemented by Puckette et al [10] and Jehan [7] are the basis of the audio analysis stage.

The power of modern computer systems allows this program to be run from a single laptop computer. This allows a realistic use of this system within teaching or practice situations without modification or specialised hardware.

4 Results

The prototype visualisation is shown in Figure 4.

Harmonic Content is mapped to each sphere's radius, with the fundamental frequency being the lowest sphere, and the harmonics displayed equally spaced above. Sounds with harmonic content containing a strong fundamental frequency will display as a large sphere with smaller spheres above it (Figure 4(a)).

Noisiness is represented by a 'fountain' of particles, which are placed directly behind the spheres. These particles originate from the same point as the sphere that represents the fundamental frequency. The average direction of the particles is always directly vertical, however there is also a random horizontal force that determines the width of the particle stream. This width is mapped to the noisiness values, and thus the distribution of the particles in the horizontal direction is very narrow when the noisiness value is low, and very wide when it is high (Figure 4(b)).

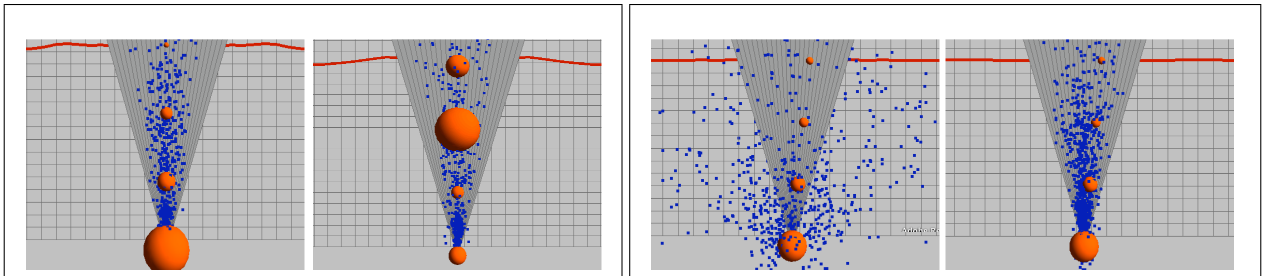
Fine Pitch is represented using a balance metaphor; the spheres are balanced vertically if the pitch of the sound agrees with the model pitch. This forms a line of spheres that moves in a manner reminiscent of the needle display in common electronic chromatic tuners (Figure 4(c)).

Loudness is represented by the overall height to which the group stretches (its size). This follows the size analogy validated by Walker (Figure 4(d)). It is also shown over the recent timespan by the red trails that are located either side of the central group of spheres (Figure 4(e)). Smooth or sudden changes in dynamics are common and planned in musical performance, and the visualisation helps describe the rate of change and its linearity. Display over the recent timespan for the other parameters mentioned above is not usually as important, as their planned change over time is much less common.

4.1 Information revealed

When two tones are played consecutively with markedly different harmonic content the sphere's sizes change accordingly (Figures 5(a) and 5(b)). This can seem quite dramatic, with spheres suddenly disappearing or appearing when certain notes are played. Generally speaking, large changes in sound quality between notes near to each other are not desired by musicians, and this visualisation helps locate these problem notes. Notes or groups of notes that would be referred to by musicians as having a 'muffled', 'honky' or 'tinny' sound character are clearly distinguished by this visualisation behaviour.

Although the difference between vibrato (cyclical frequency modulation) and tremolo (cyclical amplitude modulation) is not always obvious to the ear, the visualisation manages to distinguish these. It sways from left to right if vibrato is used, or stretches and shrinks upwards and downwards in a spring-like manner if tremolo is used . The trails of the visualisation also display the smoothness and speed of the tremelo (Figure 5(g)). Tremelo or vibrato quality can

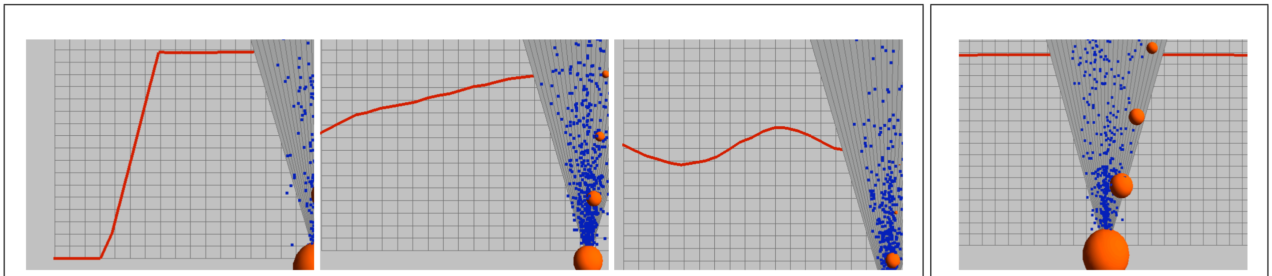


(a) An ‘oo’ sound has harmonic content in the fundamental and 1st harmonic, hence lower spheres are largest.

(b) An ‘or’ sound has harmonic content in the 2nd and 3rd harmonic, hence upper spheres are largest.

(c) Noisy sound causes the particle ‘fountain’ to be indiscriminate about which direction particles are fired towards.

(d) In contrast to Figure 5(c), if sound contains little noise particles fire straight upwards.



(e) The red trails on either side of the sphere group show a large loudness contrast occurring.

(f) They can also show a slow increase in loudness occurring.

(g) Tremelo (cyclical amplitude modulation) speed and smoothness can be visualised.

(h) The rightwards slant of the sphere group indicates fine pitch is slightly sharp. Vertical alignment represents correct tuning.

Figure 5: Revealed Information.

be very important aspects of the musician’s sound quality.

The display of overall loudness can give the performer an indication of their dynamic range. This can be very important for ensuring adequate ability for dynamic contrast is achieved. The time varying ‘tail’ of the visualisation allows the user to see these dynamic contrasts over the recent time-span (Figure 5(e) and 5(f)), and understanding the smoothness and consistency of their musical use of loudness is important for musicians.

This display of fine pitch (Figure 5(h)) functions in a similar way to an everyday electronic chromatic tuner, but it is essential to understand tuning within the context of the manipulation of these other parameters. The interconnected display helps bring out interconnected patterns of performance. Musicians often develop habits of sacrificing one musical attribute (often tuning) to achieve another, not necessarily realising they’ve adopted this behaviour. By using a display that shows tuning as well as other parameters behaviour of this nature may be more easily recognised.

5 Discussion

It can be seen from the results that this visualisation allows musicians to understand the outputs of many types of acoustic analysis and immediately relate this information to their musical sound production. This should increase the efficiency of musical training and prove a complementary tool for musical pedagogy.

5.1 Possible effect of interpretation

This system is interactive in a different way to most information visualisation. Users interact with the visualisation by creating new data in a different way, rather than by adjusting their view of a static set of data. The visualisation may play roles apart from the pure communication of information, and this raises the question of interpretation. Specifically, information may be communicated with a visual context that is not impartial, and rather suggests its modification. This is trivial in static visualisations, as this possibility usually does not exist. However, considering that the process of musical training is undertaken over hundreds of hours, the musician may spend a great deal of time staring at this visualisation. Therefore, it may be possible to utilise the impulse to 'tidy' a visual scene for the purpose of recommending the shift of parameters towards optimums, or conversely a badly designed visualisation may accidentally recommend arbitrary changes.

5.2 Other Applications

This visualisation allows the study of sound quality from a musician's point of view. In acoustics research subjective sound quality is often investigated by asking human subjects to rate a set of sounds, and then attempting to correlate these sounds against acoustic parameters extracted from the sound. However, by providing an easily interpretable display it is possible for the musician to tell the acoustician what they personally believe to be good sound by defining it as an object shape.

Whilst this visualisation benefits musicians, the underlying principles are applicable to other areas directly. These may include any situation where a sound source is better monitored visually rather than aurally, or when multiple sound sources need to be monitored concurrently. As an example, the sound produced by each of an aeroplane's jet engines would be too distracting to monitor aurally within a cockpit, and with multiple engines in some aircraft the distinction between these would also be difficult to make with concurrent sound signals. However these sounds may contain sound 'signatures' of faults in or inconsistencies between the functioning of these engines. With visual display these sounds may be described in a non-distracting but still useful manner.

6 Conclusion and Further Work

In conclusion, this paper has presented a real-time visualisation system that communicates data extracted from real-time acoustic analysis to musicians. This system improves the efficiency of musical training by clarifying aspects of musical sound perception.

Much further work is likely for this system. Practical usage of the system will be evaluated quantitatively and qualitatively with musicians in training. The system will also be applied in the inverse, by asking musicians to discuss what constitutes good sound based on experimenting with the system, and correlating these subjective values against those accepted by the literature. This may also be a useful approach in the discussion of concert hall acoustics, and specifically stage acoustics. Further opportunities for research in this area include investigating a similar system designed for multiple musicians, extending and validating the sound quality representation with more complex data extraction and visualisation algorithms, and designing interactive creative works based around similar ideas.

7 Acknowledgements

We gratefully acknowledge the assistance provided by an Australian Postgraduate Award and a Department of Architectural and Design Science Supplementary Scholarship.

References

[1] Frederick J Ampel. The history of audio and sound measurement. In *Proceedings of 94th Convention of the Audio Engineering Society*, Berlin, Germany, 1993.

[2] Densil Cabrera, Sam Ferguson, Steven Tilley, and Masayuki Morimoto. Recent studies on the effect of signal frequency on auditory vertical localization. In *Proceedings of International Conference on Auditory Display*, Limerick, Ireland, submitted 2005.

[3] Jean Callaghan, William Thorpe, and Jan van Doorn. The science of singing and seeing. In R Parncutt, A Kessler, and F Zimmer, editors, *Proceedings of Conference of Interdisciplinary Musicology*, Graz, Austria, 2004.

[4] Cantare Systems. *Sing and See*. www.singandsee.com, 2005.

[5] Cycling74 Inc. *Max/MSP/Jitter*. www.cycling74.com, 2005.

[6] F Wai-ling Ho-Ching, Jennifer Mankoff, and James A Landay. Can you see what I hear? The design and evaluation of a peripheral sound display for the deaf. *CHI*, 5(1):161–168, 2003.

[7] Tristan Jehan. *Perceptual Synthesis Engine : An Audio-Driven Timbre Generator*. Masters thesis, Massachusetts Institute Of Technology, 2001.

[8] J D Johnston. Transform coding of audio signals using perceptual noise criteria. *IEEE on selected areas in communications*, 6:314–323, 1988.

[9] International Business Machines. *Speechviewer III*. International Business Machines, New York, USA, 2005.

[10] Miller Puckette, T Apel, and D Zicarelli. Real-time audio analysis tools for PD and MSP. In *Proceedings of International Computer Music Conference*, University of Michigan, Ann Arbor, 1998.

[11] C. William Thorpe, Jean Callaghan, and Jan van Doorn. Visual feedback of acoustic voice features: New tools for the teaching of singing. *Australian Voice*, 5:32–39, 1999.

[12] Robert Walker. The effects of culture, environment, age, and musical training on choices of visual metaphors for sound. *Perception and Psychophysics*, 42(5):491–502, 1987.

[13] Graham F Welch, D M Howard, and C Rush. Real-time visual feedback in the development of vocal pitch accuracy in singing. *Psychology of Music*, 17:146–157, 1989.

[14] Graham F Welch, C Rush, and D M Howard. The SINGAD (SINGing Assessment and Development) system: First applications in the classroom. *Proceedings of the Institute of Acoustics*, 10(2):179–185, 1988.