ACOUSTICAL FEATURES OF MUSICAL SOUNDS

Howard F Pollard
6 Wren Place, Cronulla 2230.

ABSTRACT: In searching for significant features of musical sounds, it is necessary to convert basic physical spectral data into psychophysical measures. Preliminary analysis and organisation of the data takes place in the cochlea where incoming acoustic waves are filtered and converted into digital nerve impulses. These impulses are passed on to the brain where they are processed in a number of stages of increasing sophistication. Time and frequency analysis occurs simultaneously allowing continuous assessment of both the characteristics of the starting transient and the eventual 'steady' sound. The tristimulus method of analysis emulates this process by measuring the changes that occur with respect to both time and frequency.

1. BASIC PHYSICAL ANALYSIS

In order to study the acoustical output of a musical instrument, a set of representative sounds may be recorded and displayed as a time display, a frequency display or a combination of both using a sonograph [1]. Such records are purely physical in that they are independent of the properties of the ear and brain. For studies of instrument mechanics or the effects produced by rooms, this type of record is adequate. Quantities that may be measured include the mean sound level of the complete sound, the durations of the starting transient, steady state and decay together with spectrum levels at specific times. To have relevance to musical problems it is necessary to conduct further processing, in particular, to convert the measurements into psychophysical units and subject them to data reduction.

2. PSYCHOPHYSICAL MEASURES

When the processing functions of the ear and the brain are taken into account, features of the sound may be extracted that can be described as being psychophysical, implying that they are a combination of physical measurements and data originating from experiments involving human judgments. Preliminary processing takes place in the cochlea where the frequency sensitive hair cells are organised into a set of band-pass filters, called critical bands. When excited, the hair cells generate digital nerve impulses that are transmitted via the acoustic nerve to a series of auditory centres in the brain. The critical bands play a primary role in determining loudness, pitch and timbre. They determine spectral weighting, masking and some timing functions leading to measures including the mean loudness of the complete sound, band loudness spectra at specific times and band loudness derivatives as a function of time (these provide estimates of start-times and rise-rates for the partial tones present in the sound). The critical bands have time constants of approximately 10 ms at low frequencies falling to less than 1 ms at high frequencies. Their bandwidths are such that, for most musical sounds, the first five or six partial tones fall within separate bands – an important occurrence for the assessment of both pitch and timbre. For the higher partial tones, two or more fall within a particular band requiring their intensities to be combined. *[In practice, it is often convenient to use a set of one-third octave band filters. Response times are similar to those for critical bands except for the range below 400 Hz where one-third octave filter response times are longer than those for corresponding critical bands].*

Processing by the brain is an involved and only partially understood domain. It is generally agreed that the organisation of neurons in the cochlea by their frequency sensitivity is maintained through all the auditory stages up to the auditory cortex. The brain includes stages involving computation, comparison, correlation and integration coupled to a very sophisticated memory system. The result is the ability to perform rapid evaluation of a range of features of a sound. Pitch, loudness and timbre are assessed as well as more subtle aspects such as sharpness [2], roughness [3] and features of the important starting transients such as early noise, inharmonic components and dominant tones. The overall effect of all this activity is to produce a 'sound image' or 'acoustic template' that is stored for future reference and identification of sounds.

Data reduction plays an important role in reducing each image to its essential properties. For instance, in assessing the pitch of a complex sound, the first few partial tones that fall in separate critical bands play a dominant role in allowing the brain to form a harmonic template of the sound and assign a pitch [4]. This has the effect of reducing a large amount of spectral data to a single entity. Similarly, in the case of timbre, it is important that the first 5 or 6 partial tones fall in separate critical bands before the signals are passed on to the brain for assessment. In the tristimulus method described below, the large amount of spectral data associated with each note at any given time is reduced to 3 timbre coordinates. A further important property of the ear-brain system that increases overall processing economy is adaptation wherein certain cells respond mainly to changing signals, reducing their activity when no new information is presented.

A basic task for the investigator is to identify the essential features in each type of sound. For musical sounds, the relative roles of the 'steady sound' and the transients must be assessed. *[The term 'steady sound' is not particularly accurate since there are significant variations of loudness, pitch and timbre forming components of vibrato to which the brain is sensitive].* In the case of timbre, some investigators (including Helmholtz [5]), have maintained that the brain uses different procedures for assessing steady-state and transient

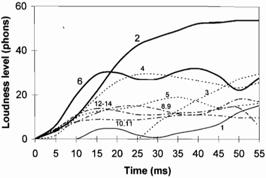

Fig 1 Starting transient growth curves for an open organ pipe (Principal 8', pitch G4, Marcussen organ, Spånga church, Stockholm). The time is measured from the first onset of sound from the pipe. For this pipe, there is some initial noise in the bands containing partial tones 8&9, 10&11, and 12-14. Initially, partial tones 6 and 2 (shown by heavy lines) are dominant. After a slow start, the fundamental increases in strength up to 200 ms. The second harmonic remains the strongest component.

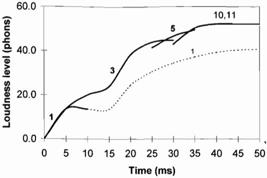

Fig 2 Dominant partials for the starting transient of an oboe note E4 flat. The order of dominance is the fundamental (0-5 ms), partial 3 (5-28 ms), partial 5 (28-34 ms), partials 10,11 (44 ms onwards). The later level of the fundamental is shown by the dotted curve, reaching a level of 40 phons in the steady state.

sounds. Other workers consider that these two aspects should be considered together, timbre being a composite property of the whole sound. It is not clear how the brain processes and coordinates information relating to the starting transient and the steady state as different parts of the brain take part. It will be shown below that both aspects of timbre can be combined in the same experimental procedure.

3. ANALYSIS PROCEDURE FOR LOUDNESS AND TIMBRE

For the analysis of a complete musical note, the sound is digitised, filtered into critical or one-third octave bands and stored in a computer for processing [6]. As the sound builds up, measurements are made at 5 ms intervals using a sliding Hanning window of 10 ms equivalent bandwidth [6]. Starting transients typically occupy between 30 and 80 ms with some string sounds extending to 300 ms. The output of each filter band is converted into linear loudness units (sones) and then into logarithmic loudness level units (phons). *[The logarithmic phon scale is more useful for displaying low-level values than the linear sone scale.]* A productive way to present the basic data is in the form of spectrum growth curves for the partial tones present.

In evaluating the critical band loudness response it is immaterial where a particular tone falls within a band or whether the tone is harmonic or not. If two or more tones fall within the same band they are not heard as separate tones; their intensities are summed to find the equivalent loudness within that band. For many musical instruments the partial tones are harmonic although they often deviate from this strict condition in the first few milliseconds as the standing wave system is being established. There are consequently initial non-harmonic tones present in some instruments, such as the 'mouth tones' heard in the flute and flue organ pipe [1]. Percussion instruments generally have non-harmonic partial tones. Under normal listening conditions, low-level initial tones are difficult to hear as they are easily masked by background noise, especially by wind noise in the case of organ pipes.

4. INITIAL PROCESSING OF LOUDNESS AND TIMBRE IN THE INNER EAR

Growth curves and dominant partials

An organ pipe has been chosen as an example of a musical sound having moderate harmonic development and which is not dependent on the manner of playing by a performer. A comparison will be made with a note played on a representative orchestral instrument, an oboe. Figure 1 shows a full set of growth curves from the onset of sound up to 55 ms for a Principal pipe, pitch G4, from the Marcussen organ in Spånga church, Stockholm. The G4 pipe was selected from the 8 ft Principal rank of pipes. The first 5 partials fall in separate critical bands, partials 6 and 7 in the next band, then 8 and 9 together, 10 and 11 , 12 to 14, 15 to 18, etc. For present purposes, there is too much information in Figure 1 from which essential features need to be extracted.

The dominant partial for a given time value, an important quantity in the assessment of the starting transient, is revealed by the envelope of the growth curves (shown as heavy curves). The dominant partials in the starting transient for the Principal G4 pipe are the 6th (from onset to 18 ms) , and the 2nd (from 18 ms onwards). The fundamental is slow to develop taking at least 200 ms to reach a level comparable with that of the 2nd harmonic.

There is no regular pattern in the order of appearance of the partial tones from pipe to pipe even within the same rank of pipes. For instance, for the C4 pipe from the same 8 ft

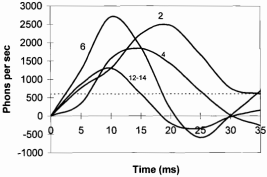

Fig 3 Derivative curves for the four fastest rising partials of the Principal G4 pipe. The dotted horizontal line at 600 phons/s is a threshold level for the measurement of start-times and times to steady state. The position of the peak of each curve gives the maximum rise-rate and the corresponding time.

Principal rank, the dominant partial order is 6th, 2nd , 3rd and fundamental while for the E4 pipe the order is 5th, 2nd , and fundamental. The voicing in this organ is in accord with baroque organ styling. A different dominant partial order would be expected from pipes voiced in a 'romantic' manner [1].

Figure 2 shows the starting transient dominant partials for a note (E4 flat) played on an oboe (data from Moorer and Grey [7]). The order of appearance of dominant partials is the fundamental (0-5 ms), 3rd (5-28 ms), 5th (28-34 ms) and partials 10 and 11 (44 ms onwards). As with the organ pipe, this is a record of a single sound and is not necessarily characteristic of all oboe notes. On wind and string instruments, there are considerable tonal differences between different notes played on the same instrument or between the same notes played by different players.

Properties of the starting transient

For a given sound, the rate of rise of loudness for each partial tone is different. There are two related parameters that are important in the assessment of the starting transient [5]: the maximum rise-rate and the time to reach a steady sound for each partial. Both quantities may be measured from derivative curves, based on the measured loudness differences at 5 ms intervals. Figure 3 shows smoothed derivative curves for the four fastest rising partials of the Principal G4 pipe. From these curves the following values are found:

	start-time (ms)	max rise-rate (phons/s)	time to max rise-rate (ms)	time to steady sound (ms)
Partial #6	2.5	2750	10.5	18.5
Partials #12-14	3.0	1300	9.5	15
Partial #2	4.0	2500	19	33
Partial #4	6.5	1800	14	25.5

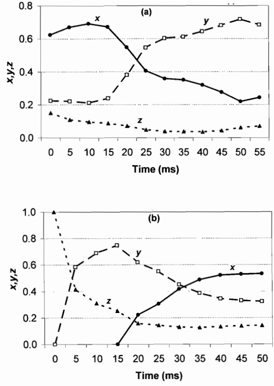

Fig 4 Variation of tristimulus coordinates *x, y, z* with time for (a) Principal G4 open organ pipe, (b) oboe note E4 flat. For (a), *x* starts high then decreases, y starts low then increases, *z* remains relatively steady. For (b), *z* starts high then falls rapidly, *y* increases rapidly at first while *x* starts late but then dominates.

The starting time for a given partial is taken as the time when the loudness level reaches 3 phons above background (a difference of 3 phons in a time interval of 5 ms is equivalent to 600 phons per second). Maximum rise-rate corresponds with the first peak of the curve (zero slope) while the duration of the starting transient corresponds with the return of the curve to the threshold value of 600 phons per second.

For the Principal G4 pipe, the starting transient duration is less than 50 ms for most partial tones but is approximately 120 ms for partial #2 and in excess of 200 ms for the fundamental. Further examples of starting transient durations include a stopped organ pipe (Gedackt G4) 40 ms, the previously quoted oboe note 35 ms, a reed organ pipe (Vox Humana G4) 30 ms, a clarinet note 45 ms and a viola note 65 ms.

Transition from starting transient to steady state

With musical sounds, time and frequency aspects need to be considered together. One of the remarkable properties of the ear and brain is their ability to process time and frequency information simultaneously. As Gabor [8,9] pointed out, there

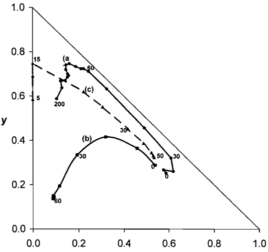

Fig 5 Tristimulus diagram with y plotted against x for (a) the Principal G4 pipe, (b) a Gedackt G4 pipe, and (c) the oboe note E4 flat. The Principal pipe starts with a bright sound (x=0.59, y=0.25, z=0.16) and finishes with stronger y and increased fundamental (x=0.10, y=0.60, z=0.30). The Gedackt pipe starts brightly (x=0.54, y=0.29, z=0.17) and finishes with a strong fundamental in the steady state after only 60 ms (x=0.09, y=0.14, z=0.77). The oboe note starts with fundamental only and finishes brightly (x=0.51, y=0.33, z=0.16).

is no limit to the accuracy of a time or frequency measurement carried out in isolation (1 variable) but when carried out simultaneously (2 variables) the results are limited by the uncertainty principle. Gabor introduced the concept of an 'elementary signal' or 'logon' of area $\Delta t.\Delta f$, where Δt is the effective duration and Δf is the effective frequency resolution. A logon is the smallest allowable quantum of information governed by the uncertainty principle $\Delta t.\Delta f \geq 1$. A profitable application of logons is in the analysis of musical starting transients [10].

During the starting transient, the signals change rapidly with time, hence maximum time resolution is required. In this phase, a typical logon for sampled filter measurements would have Δt = 10 ms and Δf = 100 Hz. Once the steady state is reached, changes with respect to time slow down. Sensitivity to pitch changes then becomes more important requiring maximum frequency resolution. In the steady state, accurate pitch recognition would require $\Delta f \leq 10$ ms with a consequent expansion of the time resolution to $\Delta t \geq 100$ ms.

It can be concluded that the cochlea filters play a fundamental role in analysing both the starting transient and steady state parts of a musical sound. The brain then interprets this basic data in terms of a number of more sophisticated concepts.

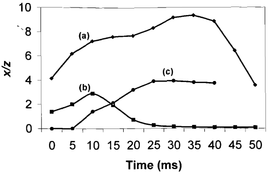

Fig 6 Graph of brightness (x/z) as a function of time for (a) the Principal G4 pipe, (b) the Gedackt G4 pipe, and (c) the oboe note E4 flat. For both organ pipes the brightness diminishes as the fundamental tone grows whereas the oboe note behaves in the opposite sense.

5. TIMBRE PROCESSING IN THE BRAIN

Tristimulus method

The human brain is particularly adept in reducing the large amount of spectral information involved in listening to music. In the musical assessment of timbre, even though at least 10 separate factors may be involved [10], three parameters have been found to lead to an adequate practical description [10, 11]. In the tristimulus method, band spectrum measurements are reduced to three normalised coordinates derived from the following three groups [12]:

Group 1: the loudness of the fundamental tone in sones,
Group 2: the loudness of partial tones 2, 3 and 4,
Group 3: the loudness of partial tones 5 and greater.

Because of its importance as a reference point both for pitch and timbre, the fundamental is the only tone included in group 1. For groups 2 and 3, the loudness of each group is computed using Stevens Mark 7 procedure [13].

Three tristimulus coordinates, x, y and z, are then defined as:

$$x = N(5,n)/N, \quad y = N(2,4)/N, \quad z = N(1)/N \qquad (1)$$

where $N(1)$ is the loudness of the fundamental, $N(2,4)$ is the loudness of partials 2 to 4, $N(5,n)$ is the loudness of partials 5 to n, and $N = N(1) + N(2,4) + N(5,n)$. The coordinates x, y and z describe timbre only; they are normalised for loudness (since $x+y+z=1$) and for pitch (since the fundamental tone is used as a reference pitch). Two-dimensional graphs may be drawn for any pair of coordinates such as x versus y or x versus z. The method may be applied to steady sounds or to changes occurring during starting transients.

Figure 4 shows the variation of x, y and z with time for (a) the Principal G4 open organ pipe and (b) the oboe note E4 flat. The tonal behaviour during the starting transient is quite different for these two sounds. For the Principal pipe, the proportion of high-frequency partials (x) falls rapidly up to 100 ms, the mid-frequencies (y) grow rapidly during the first 60 ms and then remain at a high level while z remains relatively low, reaching its maximum value at 200 ms. For the oboe note, z starts high, since the fundamental is the first partial to start, the mid-frequencies (y) rise rapidly then diminish while the high-frequency components (x) start slowly but dominate after 30 ms. The tone is bright in the steady state where the coordinate order is x, y and z.

Figure 5 shows a tristimulus diagram in which y is plotted against x for (a) the Principal G4 open pipe, (b) a Gedackt G4 stopped pipe by way of comparison, and (c) the oboe note. The Principal pipe starts with a bright sound ($x = 0.59$, $y = 0.25$, $z = 0.18$) and finishes in the steady state (200 ms or later) with much reduced x, a stronger y and an increase in z ($x = 0.10$, $y = 0.60$, $z = 0.30$). The Gedackt pipe starts brightly ($x = 0.54$, $y = 0.29$, $z = 0.17$) and finishes with a strong z in the steady state after only 60 ms ($x = 0.09$, $y = 0.14$, $z = 0.77$). The oboe note starts with fundamental only and finishes brightly at approximately 40 ms ($x = 0.51$, $y = 0.33$, $z = 0.16$).

Figure 6 shows the value of x/z as a function of time for the sounds shown in Fig 6. The ratio x/z measures the contribution of the upper partials compared with the fundamental and is a useful quantity describing the tonal balance or 'brightness' of the sound. *[While some writers dislike the term 'brightness' when applied to a sound, it is a term often used by musicians in contrast to a 'dull' sound. Bright sounds have high x/z values, dull sounds have low value.]* Both the Principal G4 and Gedackt G4 pipes show early high values that gradually decrease as the fundamental tone becomes more dominant. The behaviour is the reverse for the oboe note.

Sensitivity of the tristimulus method

The tristimulus coordinates are very sensitive to small changes in the spectrum of the sound. The method is particularly useful for studying small tonal changes due to differences in playing techniques or in instrumental conditions. It is possible to measure changes as small as the just noticeable differences (JND) in timbre for instrumental tones. According to Coltman [14], for skilled musicians, the JND in timbre for flute sounds corresponds to a change of only 1 dB in the level of a given harmonic. For less skilled listeners, the JND corresponds with a change of about 3 dB. Coltman found that the smallest JND values were observed with strong harmonics such as the second.

Consider the following spectrum (typical of a flute note played forte):

Harmonic:	1	2	3	4	5	6	7
Level (dB):	40	30	25	23	10	10	5

For this sound, $x = 0.115$, $y = 0.442$, $z = 0.443$.

If the level of the second harmonic is progressively reduced by 1, 2 and 3 dB, the y coordinate reduces by 0.015, 0.029 and 0.044 respectively with corresponding percentage changes of 3.3%, 6.6% and 9.9%. With stable sounds, the tristimulus method can record changes of the order of 2-3%.

Changes in the level of the second harmonic directly affect only data in the second group and hence the y coordinate. However, changes also occur in x and z as a consequence of the normalising condition $x + y + z = 1$.

6. CONCLUSIONS

Tonal properties of both the starting transient and the steady state can be measured using a single method in which sampled spectrum data are recorded at 5 ms intervals. Conversion into psychophysical measures facilitates interpretation of the roles played by the cochlea and the brain. Corresponding to frequency analysis performed by the cochlea, the growth of the partial tones with time, the presence of dominant partial tones and time measures associated with the starting transient may be extracted. Further analysis in the brain produces estimates of loudness and pitch (both one-dimensional) and of timbre (multi-dimensional). Application of a tristimulus method reduces timbre to three dimensions allowing for useful graphical presentation.

RFERENCES

[1] Castellengo M, "Acoustical analysis of initial transients in flute-like instruments", *Acustica* **85,** 387-400 (1999).

[2] Bismarck G von, "Sharpness as an attribute of the timbre of steady sounds", *Acustica* **30,** 159-172 (1974).

[3] Plomp R, "Aspects of tone sensation", Academic Press, London, 1976.

[4] Hartmann WM, "Pitch, periodicity and auditory organisation", *J. Acoust. Soc. Am.* **100,** 3491-3502 (1996).

[5] Helmholtz HLF von, *On the sensations of tone,* 4th German edition 1885, Dover reprint, NY, 1954.

[6] Pollard HF & Jansson EV, "Analysis and assessment of musical starting transients", *Acustica* **51,** 249-262 (1982).

[7] Moorer JA & Grey J, "Lexicon of analysed tones", *Computer Music* J. **2,** 23-31 (1978).

[8] Gabor D, "Theory of communication", *J. Inst. Elec. Eng.* **93** (Part III), 429-457 (1946).

[9] Gabor D, "Acoustical quanta and the theory of hearing", *Nature* **159,** 591-594 (1947).

[10] Pollard HF, "Feature analysis of musical sounds", *Acustica* **65,** 232-244 (1988).

[11] Plomp R & Steeneken HJM, "Place dependence of timbre in reverberant sound fields", *Acustica* **28,** 50 (1973).

[12] Pollard HF & Jansson EV, "A tristimulus method for the specification of musical timbre", *Acustica* **51,** 162-171 (1982).

[13] Stevens SS, "Perceived level of noise by Mark VII and Decibels (E)", *J. Acoust. Soc. Am.* **51,** 575-601 (1972).

[14] Coltman J, "Just noticeable differences in timbre of the flute", *J. Catgut Acoust. Soc.* **3,** 26-33 (1996).