Analysis and assessment of musical sounds

Pollard, H. F. and Jansson, E. V.

journal: STL-QPSR
volume: 21
number: 4
year: 1980
pages: 062-098

http://www.speech.kth.se/qpsr
C. ANALYSIS AND ASSESSMENT OF MUSICAL SOUNDS
H.F. Pollard* and E.V. Jansson

Abstract

A complete description of musical sounds must include the changes in spectrum with time to describe the four major characteristics: duration, pitch, loudness and timbre as well as the microstructure within notes. To understand the perception of the musical sounds, the operating characteristics of the hearing process should be considered, such as the critical bands of hearing, masking effects, the time resolution needed, time constants for integration of signal parameters, and likely memory functions.

Long-time-average-spectra of organ ranks are shown to predict steady state parameters independent of starting transients, frequency range of a rank and its roughness. Sampled filter methods show that transients of different partials dominate during different portions of the total starting transient and that very high rate of changes may occur. By replacing the natural starting transients with exponential ones, the overall effect is not to make the sound completely different, it is a change in one quality factor of the tone, while lack of reverberation makes the tone sound unnatural.

The tri-stimulus method, employed in explaining the colour vision, is introduced as a method to analyse musical sounds. In this investigation the fundamental was used as the first coordinate, the following three partials resolved by the ear as the second, and the higher partials as the third. The tri-stimulus graphs display differences between ranks and between pipes of the same rank, which seems perceptually important.

In the following three appendices analysis methods are reviewed, results of previous analysis presented, and a synthesis prescription given to test the relevance of an analysis.

1. Introduction

Musical sounds form a special class that differ markedly in structure from either pure tones or noise, the two extremes that are usually considered by acousticians. Methods of analysis that have been devised for pure tones or noise are often unsuited to the analysis of a musical sound. In particular, changes that occur in the spectrum of the sound with time must be described in some way.

In the time domain, a musical sound consists essentially of three parts: (1) a starting transient in which complicated interference

* guest researcher at the Dept. of Speech Communication, Febr.-June, 1979. Address: School of Physics, University of New South Wales, Kensington, Australia
processes occur between the free vibrations of the modes comprising the instrument and the forced motion imposed by the source of energy, (2) a more or less steady region corresponding to the forced motion of the instrument. (This 'steady state' sound usually consists of a set of partial tones that may or may not be related harmonically, together with some background instrumental noise.) (3) The decay of free vibrations of the modes of the instrument when the forcing function has been removed. The decay pattern is usually complicated by interference between the instrument modes and room modes. Under normal music-making conditions, the room modes predominate during the decay. A major characteristic of musical sounds are the changes of state that occur with time. It is rare to encounter a sound in music whose descriptive parameters are not varying in some way with time.

A measurement system suited for analysis of musical tones must be capable then of providing not only a time description of the sound but also a detailed frequency-time analysis in terms of all the important spectral components. Furthermore, even the simplest piece of music consists of a sequence of notes. Therefore, the analysis procedure should be capable of dealing with at least a simple sequence, such as a diatonic scale.

The present investigation aims towards such a physical analysis procedure. But it is equally important that the analysis procedure is adequate for the perception by our hearing. Therefore properties of the hearing process are discussed in some detail. Musical sounds have certain signal properties and their perception determines which of these properties are important.

The report is the work by one of the authors (HFP, on sabbatical leave to the Department of Speech Communication, KTH) together with the second author. The report contains two main parts: one part reviewing previously used methods and investigations, and a second part with an analysis of organ tones done at the Department of Speech Communication. Furthermore, simple gating and filter experiments are described and a new method, "the Tri-stimulus Method", is introduced as a possible description for the timbre of musical instruments.
2. Characteristics of a musical sound

There are four major characteristics that are used to describe a musical sound: (1) duration, which involves the relative importance of the starting transient, the 'steady' sound and the decay; (2) pitch, which involves an assessment on a one-dimensional scale related to frequency; (3) loudness, which is a measure of the electric power transmitted by the auditory nerve (Howes, 1979); (4) timbre, which is a multi-dimensional entity.

In addition to the major characteristics of the sound, there are the dynamic fluctuations within each note (sometimes called the micro-structure), examples of which include: (a) variations in the level of the time envelope, (b) variations in both the level and frequency of spectral components, (c) the temporary appearance of inharmonic spectral components, and (d) the presence of noise components. All of these factors can be vital for a musical sound.

3. Perception of musical sounds

When it is desired to study loudness relationships and some aspects of timbre, the operating characteristics of the hearing process should be considered. While there is detailed physiological and psycho-physical information available concerning processes occurring in the region of the cochlea, there is much less known concerning the processing of neural impulses after they leave that region.

The basilar membrane contains about 30,000 receptor hair cells and behaves as a set of 24 band pass filters (critical bands) each containing approximately the same number of hair cells (about 1300 per band). The bandwidths of the critical bands are approximately 100 Hz for frequencies below about 450 Hz and approximately one-third octave for higher frequencies (Zwicker & Feldtkeller, 1967). The bandwidths measured as pulsation threshold have, however, approximately half the above bandwidths (Plomp, 1967, Ch. 1). The maximum firing rate of neurons energised by the sound wave is 1000 per sec; the rate being dependent on the intensity of the sound. Individual neurons respond to an intensity range of 40 dB. The recovery time after each firing
is greater than 1 ms so that the minimum time required to detect a change in intensity is probably in the range 2-10 ms. There is some agreement that a time constant of 100 ms is involved in the aural assessment of the loudness of a steady sound (Zwicker & Feldtkeller, 1967). Details concerning the subsequent auditory coding of neural impulses arising from the ear are still somewhat speculative (Roederer, 1975).

Thus, for loudness determination, the analyser should preferably be the equivalent of a set of filters with bandwidths approximately the same as the critical bands of the ear, and with the added capability of providing an estimate of the loudness in each band after low pass filtering with a time constant of 100 ms. The firing rate implies that the filters should be sampled at least every 10 ms.

One characteristic of a steady sound is roughness (Terhardt, 1974). A musical sound with two or more partials within the same critical band will give rise to a sensation of roughness. The partials within the same band are not resolved but produce a beating effect which can be unpleasant. With only one partial in each occupied band, the musical effect is one of smoothness. The perceived roughness corresponds to a low-pass RC-filtering with a 13 ms time constant.

A strong sound component may make another component impossible to perceive. This phenomenon is called masking. In the frequency domain a low tone may mask a high tone, while the opposite effect is small. Therefore, simultaneous masking can be modelled as a masking threshold falling from the level of each band at a rate of approximately 10 dB/Bark towards higher pitches (c.f. Zwicker & Feldtkeller, 1967, Fig. 76.2). Masking phenomena are also present in the time domain. A suitable labelling pre-masking and post-masking were introduced in Zwicker (1977). The pre-masking for loudness is short relative to that of post-masking.

The relative importance of the starting transient and the steady state parts of a musical note have been studied by a number of workers (summary in Grey, 1975). The starting transient plays an important
role in the recognition of a musical sound, since it represents an abrupt change of state from previous sounds, even when these come from the same instrument. There is some argument as to whether the starting transient is perceived as a whole or whether there is time for the ear and brain to conduct a detailed analysis. It seems likely that the brain attempts to find overall feature changes during the transient rather than indulge in a detailed spectrum analysis as a function of time. Musicians are well aware of the importance of starting transients and frequently refer to the character of the 'attack' sounds. These are especially prominent in brass, the woodwinds, and in organ pipes. In the two first cases the starting transients are major clues to the identification of these instruments (Wedin, 1970). In the latter case, organ builders have evolved specific methods for either increasing or decreasing the presence and character of the starting transient in a pipe. The durations of the starting transients for most musical sounds lie in the range 5 to 360 ms (Jansson, 1977). For transient analysis the system must be capable of analysing segments of the waveform having durations of the order of 10 ms or less.

Recently some significant advances have been made in timbre research which has involved the use of multi-dimensional analysis. Both Plomp and Steeneken (1973) and Grey (1977) report that three factors seem to be sufficient to describe the timbre of steady sounds. Grey tentatively identifies these factors as: (a) the location of the centroid of the loudness distribution (this factor is also sometimes called sharpness), (b) synchronism, a term used to describe the behaviour of the partials in which the upper partials rise and fall in amplitude together at one end of the note (beginning or conclusion) and are independent at the other end, (c) the presence of higher order partials in the initial stage of the attack.

The pitch sensation of musical sounds, i.e. complex tones, are related to the fundamental frequency. However, even with missing lower harmonics, including the fundamental, a clear pitch sensation is achieved. One explanation of this fact is that the ear makes use of all partials that are resolved by the critical bands of hearing
(Plomp, 1976). Recent findings, that singular hair cells in the cochlea have a sharp frequency selectivity (Russele & Sellick, 1978) provide explanations for the highly selective pitch sensation.

Although most discussions of the problems involved in analysing musical sounds refer to single notes, it is well to bear in mind that even the simplest piece of music consists of a sequence of notes. To be realistic, the analysis procedure should be capable of dealing with at least a simple sequence, such as a diatonic scale.

Furthermore, music consists of several parts played by different instruments. Experiments have proved that spectral cues are relevant when two tones of the same temporal envelope are presented in perfect synchronisation. With a slight asynchronisation, the higher note starting 10 to 30 ms earlier than the low note, the separation threshold is lowered from -20 dB to -60 dB for the higher note, now largely independent on non-temporal cues (Rash, 1978). This asynchronisation corresponds to that generally found in playing of polyphonic music (Rash, 1979).

Thus, a fairly complex analysis procedure is needed to incorporate all mentioned phenomena. In the frequency domain we need to model a parallel filter bank with critical bands with the high frequency masking flanks. For the pitch perception a second step with high frequency resolution is needed. Regarding the loudness, the different loudnesses of the different filters should be added in an appropriate manner and integrated in the time domain. For the roughness the envelope output in each filter band should be low-pass filtered equivalent to an RC-network of a 13 ms time constant. These criteria are set by present knowledge of properties of hearing. There are, however, still much that must be incorporated in a final model.

Thus, there is also the need to introduce some form of memory function, possibly at two levels: (1) a short-term memory process that operates within a given note and which would enable comparison to be made between the starting transient and steady-state parts of
the note together with an assessment routine for dynamic changes,
(2) a long-term memory which would enable the identification of the
instrument being played or the relationship of one note with another
in a sequence.

A further refinement that may become possible at a later date
is to use a system of analysis that follows more closely the audi-
tory coding processes used by the brain. These processes have yet
to be delineated with respect to hearing but advantage may possibly
be taken of the progress that has been made in relation to optical
coding.

Some initial studies for this report are summarized in Appen-
dices. Thus the Appendix A presents methods previously used for
analysis of musical sounds, Appendix B some results of previous anal-
yses, and Appendix C methods of synthesis for finding out the im-
portant sound parameters for the perception.

4. LTAS-measurements and filtering experiments

Recordings were made of a diatonic scale (C₄ to C₅) played on
three different organ pipe ranks at Spånga Church, Stockholm, using
four different note durations (very short, 0.5 s, 2 s, continuous
diatonic scale). For each of the ranks of pipes recorded and each
note set, an LTAS was computed (see Appendix A). Fig. IV-C-1a shows
the curves derived for a Principal 8' rank (short notes and contin-
uous scale), while similar curves are shown in Figs. IV-C-1b and 1c
for Gedackt 8' and Vox Humana 8' ranks, respectively. There is little
difference in the shapes of the curves obtained for each rank, ex-
cept for the shortest notes played on the Principal 8'. In the latter
case, the curve for the shortest time shows a reduced level at
low frequencies likely because the fundamental is slow to develop.
It is concluded that where the LTAS is not very sensitive to the char-
acter of the starting transients of the notes, the LTAS is a repre-
sentative average for the quasi-steady state portion of the tones —
i.e., the physical parameter defining steady state timbre.
Fig. IV-C-1. LTAS-curves for a diatonic scale (C₄ to C₅) on
(a) Principal 8' rank of organ pipes,
(b) a Gedackt 8' rank,
(c) a Vox Humana 8' rank.
The organ was a Marcusson in Spånga church, Stockholm.
The curves shown are for a set of very short duration notes (---) and for a continuous diatonic scale (-!-).
A pre-emphasis curve was used in the analysis.
From the LTAS:es in Fig. IV-C-1 we can predict possible roughness sensations. The LTAS:es describe approximately the spectrum envelope for different complex tones and thus we can estimate the upper frequency limits for the different musical tones.

For the three organ pipes having pitch C₄ (262 Hz), the high frequency limits are approximately: Principal C₄, 4 kHz; Gedackt C₄, 3 kHz; Vox Humana C₄, 8 kHz.

Principal and reed pipes as Vox Humana produce a full set of harmonic partial tones and for a fundamental frequency of 262 Hz, which is also the frequency difference between partials, two or more partials will fall within the same critical band above 1.5 kHz. This will give rise to possible roughness, which is especially so for the Vox Humana, with strong partials above 1.5 kHz.

Closed pipes produce only odd-numbered partials. For instance, for a Gedackt C₄ pipe, only partials above 3 kHz can contribute to roughness. As noted above, there is little radiated sound above this frequency. Listening to recorded tones reveals that the Gedackt C₄ sounds smooth, the Principal C₄ sounds slightly rough, and the Vox Humana C₄ very rough. Low-pass filtering at 1.5 kHz removes the roughness sensation, which confirms the previous conclusions regarding roughness.

5. Growth curve measurements and gating experiment

Steady signals are frequently analysed with the aid of a set of analog filters (either 1/3 octave bandwidth or narrow band) whose outputs are scanned in series or parallel. For the analysis of transient signals, as in speech and music, a parallel bank of filters is required whose outputs can be sampled at short time intervals, at least every 10 ms. The filter bandwidths may be variable or set to 1/3 octave values. Digital filters would be an alternative but to date they do not appear to have been used in the analysis of musical sounds. A pre-emphasis network may be desirable if the magnitudes of important high frequency components are low. The weighting network can also be designed to be an approximation to loudness weighting functions (Stevens, 1972; Zwicker & Feldtkeller, 1967). The sampled filter outputs are stored in a computer for subsequent analysis and display.
5.1 Sampled Filter Method

A set of 51 adjustable filters was available for this analysis. The filters were set to have a centre frequency every 100 Hz starting from 100 Hz, the bandwidth was set to 125 Hz, smoothing filter time constant 13 ms, sampling rate 100 Hz. In addition, a pre-emphasis network was used (Fig. IV-C-2) that permitted the recording of greater amplitudes at higher frequencies and also gave a weighting that was an approximation to loudness weighting functions (Stevens, 1972; Zwicker & Feldtkeller, 1967). If required, unweighted filter responses could easily be derived from which loudness values in sones could be computed.

Figs. IV-C-3, IV-C-4, IV-C-5 show results obtained for a Gedackt C₄, Principal C₄, and Vox Humana C₄ pipe, respectively (Spånga church organ). In each case the weighted total level is given together with curves for some of the partial tones (only the first six are shown for the Principal and Vox Humana). For loudness computations the first six partials are treated separately, after which the partials are grouped since more than one of the upper partials will fall into each critical band.

One of the major effects produced by the starting transients of sounds is to draw the attention of the listener to the sound. It is of interest therefore to plot the derivative of the sound level versus time as in Fig. IV-C-6, for the Gedackt C₄. From inspection of such graphs it is possible to ascertain which partial tones are likely to be dominant in gaining the attention of the ear. Such overall features may be shown more clearly by plotting only the dominant slopes in each case, as shown in Fig. IV-C-7. For instance, in the case of the Gedackt C₄ pipe, the fundamental tone has its fastest rate of change between 0 and 11 ms, the fifth partial between 11 and 18 ms, followed by the fundamental again from 18 to 42 ms.

For the Principal C₄ pipe, the second partial is dominant between 0 and 23 ms, the 11th and 12th partials between 23 and 31 ms, the third partial between 31 and 48 ms, followed by the fundamental.
Fig. IV-C-2. (a) Pre-emphasis filter characteristic for sampled filter set.
(b) Loudness weighting function (Stevens, 1972),
(c) Loudness weighting function (Zwicker, 1967).

Fig. IV-C-3. Analysis of starting transient for a Gedackt C_4 organ pipe. Partials 1-9.
Fig. IV-C-4. Analysis of starting transient for a principal C₄ organ pipe. Partial 1-6.

Fig. IV-C-5. Analysis of starting transient for a Vox Humana C₄ organ pipe. Partial 1-6.
Fig. IV-C-6. Derivative of loudness level versus centre time of 10 ms band for a Gedackt C₄ organ pipe. The numbers indicate partial tones.

Fig. IV-C-7. Derivative of loudness level versus centre time of 10 ms band for (a) Gedackt C₄, (b) Principal C₄, (c) Vox Humana C₄ organ pipes. The partial number is marked against each graph segment.
In the case of the Vox Humana, the fundamental is dominant from 0 to 17 ms, the third partial from 17 to 32 ms, the 11-13th partials from 32 to 43 ms, etc. While such graphs do not constitute the whole picture regarding the aural assessment of transients, they are at least useful in revealing the partials which have the greatest rate of change of loudness at each interval of time. It is interesting to note the high rates of change involved, in some cases over 2000 dB/s.

The above analysis implies that there is an analog of the effect which is applicable to rates of change of signal level. It would be of interest to conduct further psycho-acoustic tests relating to this supposition.

5.2 Gating Experiments

Effects of three kinds of gatings were tried. In the first one modifications of the starting transients were investigated. The transients were cut with an exponential fast gate to avoid click sounds. The initial portions of tones were cut off in steps from 50 to 300 ms. The effects of the cuttings varied from small to considerable (especially for the Vox Humana) but did in no case make the sound completely different. The overall effect is to change one quality factor of the tone.

In the second series the effect of cutting away the room reverberation was tried. With the previously used gating system, tones with and without reverberation were compared. Lack of reverberation makes the tone sound unnatural. In the third series two tones were added with the start of one delayed relatively to the second.* The delays tried were 0, ± 25, ± 50 ms. The effect of this delay was fairly small even with the 50 ms delays although the difference in starting transients were perceived in these cases. A 50 ms time delay corresponds to 16 m difference in distance to the pipes, i.e. a fairly large distance difference seldom encountered by a listener.

5.3 Fourier Transform Method

Similar information to that provided by the sampled filter method may be obtained by digitising and storing the signal and then using a

* The additions were implemented by Ake Olofsson on the computer of the Department of Technical Audiology, Karolinska Institutet.
fast Fourier transform to compute the spectrum of selected segments of the signal. If a sampling rate of 25 kHz is used, then each 10 ms slice of the signal will contain 256 points and a corresponding frequency spectrum is obtained with a frequency resolution of 100 Hz. Problems that arise with this method include the definition of the starting point of the analysis and the choice of a suitable window function. Some overlap of time slices is desirable - at least 50% is often used. This method needs further development before it can compete with the sampled filter method for convenience of operation.

6. Tri-stimulus Method

A basic problem in devising a satisfactory system of measurement for transient sounds is to relate the acoustic procedures to those used by the ear and brain. Unfortunately, there is little detailed information available concerning the processing of neural impulses after they leave the region of the cochlea. This situation is in contrast with optical research where there is considerable information available concerning the coding of optical signals. Despite the enormous number of receptors in the retina, colour vision depends on only three different types of cone receptors. Some of the possible coding operations that occur between the retina and the visual cortex have been discussed by Land (1977).

Some clues as to possible auditory coding procedures come from psychoacoustic experiments relating to the timbre of musical sounds. Grey (1977) describes experiments on a set of 16 music instrument tones in which perceptual similarities were determined for all pairs of stimuli. Application of multidimensional analysis showed that three dimensions were sufficient to give a satisfactory representation of the data (see also Section 3). Plomp and Steeneken (1973) describe an experiment in which the relation between timbre and sound spectrum for selected steady sounds was studied at various points in a diffuse sound field. A dissimilarity matrix was constructed which showed that the ten test stimuli could be represented satisfactorily by a three-dimensional space. Thus present tentative information, such as that provided by the above experiments, suggests
that the final assessment of auditory signals is in terms of a small number of derived parameters. Whether the number of such parameters can be as small as three will depend on the outcome of more extensive experimentation. On a related topic, Möller (1979) finds that six measurement parameters are needed to obtain reasonable correlation between objective and subjective descriptions of high fidelity audio systems.

Time variations of amplitude and frequency are important in the perception of sounds. Associated with these effects are memory processes, both short-term and long-term. At present there is no satisfactory method for incorporating such processes into acoustical measurement procedures.

As a starting point for the development of a simplified method for representing musical sounds, it was decided to use a 3-coordinate method based on the spectral information derived from the sampled filter set. This method produces the level of each partial as a function of time. The partials are grouped if more than one falls within the same critical band. A choice must now be made of the three spectral regions that are to represent the three coordinates. It was decided to use the fundamental as one coordinate.

A choice must also be made of the unit in which to express the filter outputs. Since a number of arithmetic operations are to be made, it was decided to convert the filter outputs into loudness units in sones. If the unweighted output levels are used, it is convenient to compute loudness values using Stevens (1972) Mark VII method. When weighted values were used, it was assumed that these values were the equivalent of PLdB values from which sones were easily obtained using the table given in Stevens (1972).

As a first approximation, the output of each band may be considered to be independent, so that the total loudness of the sound may be expressed as

\[
N_t = \sum_{i=1}^{n} N_i + \sum_{i=1}^{n} N_i + N_f \quad (1)
\]
If an analogy is drawn with the optical tri-stimulus method, each term on the right-hand side of Eq. (1) may be regarded as a tri-stimulus value from which a set of normalised coordinates may be found:

\[
x = \sum_{i=1}^{n} \frac{N_i}{N_c} \quad \text{and} \quad y = \sum_{i=1}^{4} \frac{N_i}{N_c} \quad \text{and} \quad z = \frac{N_5}{N_c}
\]

From these coordinates, it is now only necessary to select two in order to draw a graph since \(x + y + z = 1\). Fig. IV-C-8 shows a graph of \(x\) versus \(y\) with the significance of the main areas indicated.

Fig. IV-C-9 shows a tri-stimulus graph for the behaviour of the starting transients of the following organ pipe sounds:

(a) Gedackt 8' C₄, (b) Principal 8' C₄, and (c) Vox Humana 8' C₄. Times measured from the onset of the sound are shown beside each curve. The final point in each case corresponds with the steady sound. A number of interesting features may be observed. It is immediately apparent how the spectral character of the sound changes between onset and steady state. The Gedackt C₄ has strong high partials initially and then progresses to a more fundamental type of tone. The Principal C₄ starts with strong high partials and then progresses to an evenly balanced sound. The Vox Humana C₄ starts with a predominance of lower partials and then moves towards a predominance of higher partials.

This method may be used to investigate the transient behaviour of different pipes within the same rank of pipes. For instance, Fig. IV-C-10 shows normalised tri-stimulus plots for three notes of the Gedackt 8' rank, while Fig. IV-C-11 and 12 show plots for the same three notes of the Principal 8' rank and Vox Humana 8' rank, respectively. While the Gedackt notes show a similarity of behaviour, the Principal and Vox Humana notes show marked differences in transient behaviour, although the steady state values are
Fig. IV-C-8. Acoustic tri-stimulus diagram.

Fig. IV-C-9. Acoustic tri-stimulus diagram showing transient behaviour of a Gedackt C₄, Principal C₄ and Vox Humana C₄ organ pipe. The numbers indicate time in milliseconds after onset of sound. The larger circles correspond to the steady state.
**Fig. IV-C-10.** Tri-stimulus diagram showing the transient behaviour of three Gedackt 8' organ pipes.

**Fig. IV-C-11.** Tri-stimulus diagram showing the transient behaviour of three Principal 8' organ pipes.

**Fig. IV-C-12.** Tri-stimulus diagram showing the transient behaviour of three Vox Humana 8' organ pipes.
in approximately the same area of the diagram. In listening to the recorded sounds, the Principal \( E_4 \) pipe had a slowly developing fundamental while the Principal \( G_4 \) pipe had a prominent high-frequency 'chiff' in the starting transient. The Vox Humana \( E_4 \) and \( G_4 \) notes have a similar starting sound whereas the \( C_4 \) note is markedly different.

7. Discussion

Methods are now being evolved for the analysis of complex sounds, such as the starting transients of musical sounds, that take account of the known characteristics of the hearing process. Factors that are relevant to such a system of analysis include (a) an initial frequency analysis related to the critical bandwidths of the ear, (b) a knowledge of the response times of the hearing mechanism, (c) masking effects, (d) the possible operation of the precedence effect, and (e) memory processes.

In our experiments we have shown that long-time-average-spectra give a representation of the "steady-state" properties of tones from an instrument. We have shown that the properties of starting transients are suitable to analysis in sampled parallel filter bands and a new method "the tri-stimulus method" has been introduced as a simple way to summarise these properties.

Not all of the factors (a) to (e) above are sufficiently well defined to be incorporated in a system of measurement (see Appendices A and B). For instance, there is a vagueness in the literature concerning the minimum time required to register a change in loudness or of timbre. A knowledge of such times would be important in defining the frequency of sampling and averaging of time-dependent transients. There have been few attempts to incorporate memory functions into analysis procedures despite their importance in subjective evaluation of transient sounds.

Detailed information concerning auditory coding of signals is not yet available. Eventually this type of information could lead to marked changes in analysis procedures. A satisfactory method of analysis should be applicable to sequences of musical notes since the
analysis of isolated notes is of little interest outside the laboratory. A critical test of any analysis procedure is to use the analysis data as the basis for comparison by synthesis (c.f. Appendix C).
APPENDIX A

METHODS OF ANALYSIS FOR TRANSIENT AND STEADY STATES

Available methods of analysis are nearly all forms of frequency analysis. Little progress has been made with analysis procedures in the time domain.

A.1. Modified Fourier Series Method

The Fourier series method assumes that the signal is periodic and infinite in extent. One period of the signal may then be digitised and a line spectrum computed consisting of harmonics of the assumed period. If the signal is mildly aperiodic or is varying slowly with time, as in a slowly growing or decaying transient, this method may be approximated by assuming that each period of the transient may be isolated and treated as if it were part of an infinite train of such periods.

Thus, it is assumed that the sound pressure as a function of time can be represented by

\[ p(t) = \sum_{m=1}^{M} C_m(t) \sin \left[ m\omega_1 t + \phi_m(t) \right] \tag{A1} \]

where \( C_m(t) \) is the amplitude of the \( m \)th partial at time \( t \),
\( \phi_m(t) \) is the phase angle of the \( m \)th partial at time \( t \).
\( \omega_1 \) is the fundamental frequency
\( M \) is the order of the highest partial computed

The values of amplitude and phase so computed are assumed to exist at the centre point of the chosen period. Sources of error include the definition of the period boundaries, the assumption that the time function is quasi-periodic and limitations of numerical integration. The method does not detect the presence of inharmonic tones or give a measure of background noise.

This method has been used by Keeler (1972a, b) for the determination of the growth in amplitude of the partial tones of organ pipes. As pointed out by Moorer (1975), when this method is used for extracting phase and frequency information, the values may vary...
APPENDIX A

discontinuously from one period to another, which gives rise to problems if the data is to be used later for synthesis. In addition, Keeler made use of Simpson's rule rather than use direct summation which could introduce aliasing problems for sounds having greater harmonic development than organ pipes.

A.2. Discrete Fourier Transform Method

Freedman (1967) used the following function to represent a musical note:

\[ p(t) = \sum_{m=1}^{M} h_m(t) \left[ \sin W_m(t) + \phi_n \right] \]  \hspace{1cm} (A2)

where \( h_m(t) = \sum_{m=1}^{N} A_{mn} u(t - \tau_{mn}) \left[ 1 - e^{-\alpha_{mn}(t-\tau_{mn})} \right] \) \hspace{1cm} (A3)

\[ W_m(t) = \omega_m(t-\tau_{1m}) + \sum_{l=1}^{M} u(t-\tau_{lm}) \omega_{lm}(t-\tau_{lm}) \]  \hspace{1cm} (A4)

The function \( h_m(t) \) is a simple cascading of exponential attack functions and \( W_m(t) \) represents a frequency which changes by a discrete step \( \omega_{lm} \) at time \( \tau_{lm} \). \( A_{mn} \) represents amplitude, \( \tau_{mn} \) onset time, \( \alpha_{mn} \) an attack time constant, \( \phi_n \) initial phase angle of the \( m \)th component, and \( u(t-\tau_{mn}) \) is a step function whose value changes from zero to one at time \( t = \tau_{mn} \).

In order to evaluate the parameters, two transforms are used. The

D transform, defined by

\[ D(t, \omega) = \frac{1}{t} \int_{0}^{t} f(t)e^{-j\omega \tau} \, dt \]  \hspace{1cm} (A5)
is used to search through the transient to determine signal frequencies, harmonic or inharmonic. In computation, the value of $t$ is gradually increased to sort out genuine spectral peaks from spurious ones. Partial tones actually present then persist throughout the evaluation.

A G-transform is defined by

$$G(t;\omega) = \int_{t-T}^{t+T} f(\tau) e^{-j\omega \tau} d\tau$$  \hspace{1cm} (A6)

For the steady state part of the signal this transform yields the coefficients of the Fourier series representation of the signal. It can also be used to extract the physical parameters of the signal during the transient part.

A.3 Fast Fourier Transform Method (Bariaux et al, 1975)

The time function is digitised (2048 points) and divided into slices of 8.5 ms duration (256 points per slice) using a sampling rate of 35 kHz. The discrete Fourier transform for each slice is computed using an FFT algorithm. The autocorrelation function is then formed from which the power spectrum is derived. The cross-correlation function is also computed between the spectral lines of successive time intervals. Leakage problems are minimised by an iteration method by means of which it is possible to evaluate the contribution of each signal to each spectrum line.

The method produces amplitude and frequency values for each spectral component as a function of time and does not require the duration of each slice to be an exact multiple of the fundamental period. It will therefore measure inharmonic as well as harmonic components and may be used with a variety of slice duration and sampling frequencies.

A.4 Heterodyne Filter Method

Moorer (1973, 1975) describes an adaptation of the discrete Fourier transform which is used as a filter to determine the amplitude and frequency of quasi-periodic waveforms. The waveform is represented by:
Appendix A

\[ X(n) = \sum_{m=1}^{M} A_m(n) \sin\{nT (\omega + 2\pi F_m(n))\} \]  

where \( X(n) \) is the signal at time \( nT \), \( n \) is a time index, \( T \) is the time between successive samples, \( \omega \) is the fundamental radian frequency of the note, \( m \) is the partial number, \( A_m(n) \) is the amplitude of partial \( m \) at time \( nT \), \( M \) is the number of partials, and \( F_m(n) \) is the frequency deviation of partial \( m \) at time \( nT \).

For the method to function properly, it is necessary to estimate the fundamental frequency of the waveform within a 2% deviation.

Each harmonic is extracted in turn and plots made of amplitude and frequency as functions of time. The method is thus equivalent to a bank of filters.

A.5 Sampled Analog Filter Method

Steady signals are frequently analysed with the aid of a set of analog filters (either 1/3 octave bandwidth or narrow band) whose outputs are scanned in series or in parallel.

For the analysis of transient signals, as in speech and music, a parallel bank of filters is required whose outputs can be sampled at short time intervals, at least every 10 ms. Preferably, the bandwidth of the filters should be variable. For certain analyses a pre-emphasis network may be required if the magnitudes of important high frequency components are low.

A modified form of this method is to record the output of a set of filters in turn on a transient waveform recorder. Although this procedure is slow, it is feasible to record at least 256 sample points for each 10 ms of the waveform. Subsequent signal processing is performed by an interfaced computer.

A.6 Long-Time-Average-Spectra (LTAS)

Jansson and Sundberg (1975) and Jansson (1976) describe a technique for deriving an average spectrum of sounds or sequences of sounds that have a total duration of approximately 20 s. A bank of 23 filters is used having bandwidths similar to those of the critical bands of the ear.
Appendix A

The output from each filter is rectified, smoothed with a time constant set to 13 ms, and converted to a logarithmic scale. The logarithmic output from each filter is sampled at a frequency of 160 Hz, which means that several sampling points are obtained even of tones as short as 50 ms duration. Thereafter, the sampled output measures are quantized in steps of 1 dB, transformed into power measures, and added in storage cells in a computer, one cell for each filter. Finally, the contents of every cell are normalized with respect to the analysing time, transformed into dB measures, and plotted as function of frequency giving an LTAS. Thus an LTAS provides a record of sound-power band-level as function of frequency. The procedure allows the results to be displayed with essentially no delay after the sampling is completed.

As a consequence of the procedure just described, the LTAS is dependent on the fundamental frequencies, the spectrum envelopes, and the intensity levels contained in the sound analyzed. It is noteworthy, that tones of high sound pressure level will contribute more than tones of low, as the LTAS averages the sound-pressure amplitude squared times the duration of each frequency component.

The computer program allows comparison to be made of any of the stored LTAS records. For instance, a dissimilarity measure may be computed between two given LTAS:es.
APPENDIX B

PREVIOUS ANALYSES OF TRANSIENTS

There are relatively few publications available with detailed analyses of musical notes. With the advent of digital methods of analysis an increasing number of papers are appearing. In the following notes, investigations have been grouped according to traditional classes of instrument.

B.1 Pipe Organ

One of the earliest investigations was that of Trendelenburg (1936) who measured the starting transients of a number of organ pipes by means of oscillograms of the build-up of sound in successive one-octave bands. Nolle and Boner (1941) also examined starting transients, without using filters, by photographing an oscilloscope display with a movie camera. Deductions were made from the visual appearance of the oscillograms. For instance, it was observed that, with open diapason pipes, the second partial was initially dominant during the starting transient. With bourdon pipes, the fifth partial dominated initially. Reed pipes start promptly with the fundamental prominent from the outset. They also found that the time required to reach steady speech required approximately the same number of cycles regardless of the pitch of the pipe. They confirmed Trendelenburg's observation that inharmonic partials may be present during the starting transient state.

Richardson (1954) recognised the importance of the starting transient with respect to instrument tone and conducted an investigation of the initial build-up of sound in organ pipes. The microphone output was displayed on an oscilloscope, together with a timing signal, which was photographed by a moving film camera. Also photographed was a record of pressure changes in the pipe foot. Portions of the film, at approximately 10 ms intervals, were enlarged and subjected to harmonic analysis (amplitude and relative phase).

Caddy and Pollard (1957) investigated the effect of the playing action on the starting transient of a principal pipe. The factors controlling the starting transient and steady state were delineated.
Appendix B

Sundberg (1966) made a thorough study of the physical behaviour of organ pipes including an examination of starting transients. Oscillograms are shown for principal and flute pipes in which the growth of the fundamental and of the upper partials have been separated by filtering.

Keeler (1979b) applied a modified Fourier series method to the computer analysis of growth curves for a number of flute, principal and reed pipes. With this method there is the need to identify each cycle in the sound emitted by the pipe, which is difficult to do in the initial stages of the transient. In his published curves Keeler shows plots of pressure amplitude versus time. When these are converted into loudness plots, the relative roles of some of the partials are altered. Fig. IV-C-B.1 shows loudness plots for the first few partials of a gedackt, principal and trumpet pipe.

Fletcher (1976) has developed a theory for the onset of sound in an organ pipe which involves non-linear coupling between the air jet system and the set of pipe normal modes. A sensitive factor in developing a solution is the rate of rise of pressure in the pipe foot. In many cases, the second partial of the pipe is found to develop more rapidly than the fundamental. The frequencies of the upper partials are not initially harmonic being partly determined by the resonances of the pipe alone. In the steady state all modes become locked into a harmonic relationship.

B.2 Piano and Harpsichord

Fletcher et al (1962) examined the decay of sound in a grand piano using a wave analyser to isolate the partial tones. The rapid attack precluded any detailed analysis of the starting transients.

Weyer (1976, 1976/77) made a thorough examination of both the time and frequency behaviour of the starting transients and decay of various noted from harpsichords, upright and grand pianos.

Alfredson and Steinke (1978) used the Fourier transform method to examine the decay of a whole piano note but did not extend the method to examine temporal dependence.
Appendix B

B.3 Strings

Beauchamp (1974) describes the analysis of violin tones that were recorded in an anechoic room and analysed by computer using a modified Fourier series technique. The analysis provides fundamental frequency, harmonic amplitudes and harmonic phases as a function of time throughout the duration of each sound.

Grey (1975) describes the application of the heterodyne filter method to analyse variations in amplitude and frequency of each partial tone of a musical sound as a function of time. Examples are given for a number of different sounds including that of a cello note. From the original analyses, straight line segments are derived to show in as simple a manner as possible the time history of each partial tone. It is found that each of the time functions can be satisfactorily represented by between 4 and 8 line segments, as shown in Fig. IV-C-B.2. This technique significantly reduces the amount of data storage necessary for later use in synthesis of sounds.

B.4 Woodwinds

Grey (1975) also gives some examples of woodwind tones reduced to line segments. Fig. IV-C-B.3 shows notes of the same pitch from a clarinet, flute and oboe.

Using the method described in section A.3 of this report, Bariaux et al (1975) give an application of their method to measurement of the time histories of the first and third partials of a clarinet tone.

B.5 Brass

Rissett and Mathews (1969) used a modified Fourier series method to analyse trumpet tones. Since they were interested in comparing the original with synthesised tones, line segment functions were derived for the time histories of each partial tone. Factors arising from the analysis that were found to be important were:

1. the spectrum envelope,
2. the attack transient which lasts for approximately 20 ms with faster build-up of lower order harmonics than higher-order ones.
3. a high-rate, quasi-random frequency fluctuation.

It was also found that low-order harmonics have a longer decay.
Fig. IV-C-B.2. Heterodyne filter analysis of a cello note (D₄, frequency 314 Hz). Each partial curve has been approximated by straight line segments (Grey, 1975).

Fig. IV-C-B.3. Heterodyne filter analysis of notes from (a) an alto, (b) a flute, (c) a clarinet. Each note has the same pitch, D₄ (Grey, 1975).
Appendix B

B.6 Percussion

There are few examples of the study of percussion instruments, probably because of the complexity of the spectra which are often more noise-like than tone-like. Moorer (1977) gives an example of the overall analysis of a bass drum note. Fletcher and Bassett (1978) describe the analysis of sounds from the bass drum using computer-modelled band-pass filters with 3 Hz bandwidths. The sounds were analysed in 1 Hz steps from 30-1000 Hz and frequency as a function of time, peak sound pressure level and decay rate determined for each major component.
APPENDIX C

Analysis-by-synthesis is a powerful method, which traditionally is used at the Department of Speech Communication. It was early used for speech sounds (Fant, 1959, 1960). It has been used for synthesis of violin tones (Jansson, 1966), wood-wind tones (Fransson, 1964), a generative grammar for music (Lindblom and Sundberg, 1970), and is the motivation for the construction of the singing machine MUSSE (Sundberg, 1978).

Bell et al (1951) describe one variant of the analysis-by-synthesis method which provides a severe test for any system of acoustical analysis. The method consists of six essential steps:

1. Recordings of notes from musical instruments are analysed by computer,
2. A physical description of each note is extracted as a set of physical parameters,
3. A synthesized note is constructed using these parameters,
4. Subjective comparison is made between the original and the synthesized notes,
5. If the synthesis is judged to be unsatisfactory, further or different analysis is required in order to provide a better set of parameters,
6. Systematic variations in the parameters are made to test the relevance of the synthesis prescription.

The method is thus an extension of computer modelling techniques in that the ear is used as the final arbiter in comparing the synthetic model with the original sound. Once a satisfactory model has been found, other notes may be generated by applying scaling rules. A considerable saving in memory space is made compared with the direct storage of all fully digitised notes of interest. Considerable use of this method has been made by Grey (1975) in producing satisfactory synthesis recipes for a variety of different instrumental sounds.

Fig. IV-C-c.1 gives an example of an amplitude prescription for the initial part of a Gedackt 8' C4 note, analysed by the sampled filter method.
Fig. IV-C-C.1. Line segment representation of starting transient of a Gedackt C4 organ pipe.
REFERENCES


