Psychoacoustically informed spectrography and timbre*

DAVID M. HOWARD and ANDY M. TYRRELL

Department of Electronics, University of York, Heslington, York YO1 5DD, UK
E-mail: dmh/amt@ohm.york.ac.uk

Pitch and loudness are subjective aspects of sound which can be described in terms of the observed abilities of subjects to rate them on a scale from 'low' to 'high'. Timbre is a subjective aspect of sound for which there is no such scale and neither qualitative nor quantitative descriptions are generally found that are widely accepted. The purpose of this paper is to shed light on some frequency domain aspects of the nature of timbre by making use of the results obtained from an analysis system which is designed to take advantage of contemporary psychoacoustical knowledge relating to human peripheral hearing. Results are presented which illustrate the relationship between contemporary psychoacoustic ideas relating to timbre and ideas first discussed by Helmholtz and later taken up by other researchers. Analyses by the system of a selection of sounds from acoustic musical instruments with clear timbral differences are also presented in order to place these discussions in a musical context.

1. INTRODUCTION

Musicians and engineers working in the areas of music technology are generally comfortable with the subjective concepts of *loudness* and *pitch*, but less so with that of *timbre*. Terms such as *piano*, *mezzo piano*, *mezzo forte* and *forte* are used to describe the loudness of individual notes or parts by musicians, and a direct relationship between an objective measure of acoustic intensity and the subjective ordering of loudness can be demonstrated. The notes of a musical scale can be ordered subjectively by their perceived pitch, and a direct relationship between an objective measure of fundamental frequency f_0 and the subjective rating of pitch can be demonstrated. For completeness, it should be noted that small variations in pitch may also be perceived when the intensity, duration or spectral content of a sound is varied whilst the fundamental frequency is kept constant.

Timbre might have descriptions such as *mellow*, *rich*, *bright*, *strident*, *harsh*, *sombre* and *lacklustre* associated with it, but there is no direct relationship between any objective measure and the subjective rating of timbre. Indeed, formal definitions of timbre reflect this: '*Timbre is that attribute of auditory sensation in terms of which a listener can judge two sounds similarly presented and having the same loudness and pitch as being dissimilar*' (ANSI 1960). In other words, two sounds which have the same perceived loudness as well as pitch (and, to be complete, duration), but are nevertheless perceived as being different, are said to differ by virtue of their timbre.

Timbre then, relates to those aspects of the note which can be varied without affecting the perceived pitch, duration or loudness of a note, such as the frequencies and amplitudes of individual components and how they vary during a sound. Of particular importance in relation to the perceived timbre is the nature of these changes during the onset and offset stages of a sound, or how the spectral components develop out of silence at the start and return to silence at the end, and a sound is often considered in terms of its *onset*, *steady state* and *offset*. In order to describe timbre differences in psychoacoustic terms, this paper will introduce and discuss results from an analysis system that incorporates knowledge of the human peripheral hearing system based on results from contemporary psychoacoustic research. Key elements of the human peripheral hearing system are first introduced to explain the basis of operation of the analysis system, as well as the ideas that underpin current understanding of human perception of timbre, and also pitch and loudness. This is followed by a description of the implementation of the system and the nature of the spectrographic output it provides. Results are presented which illustrate both the conclusions reached by Helmholtz in regard to timbre in terms of contemporary psychoacoustic theories, as well as the nature of the frequency domain representation of a selection of notes from a number of orchestral musical instruments which differ markedly in their timbre.

2. THE HUMAN PERIPHERAL HEARING SYSTEM

For the purposes of this work, the human peripheral hearing system is assumed to comprise the outer,

* The authors would like to thank Paul Murrin for help with the preparation of the figures. The development of the real-time transputer system was supported by UK-EPSRC research grant No. GR/J42267.

https://doi.org/10.1017/S1355771897009011 Published online by Cambridge University Press

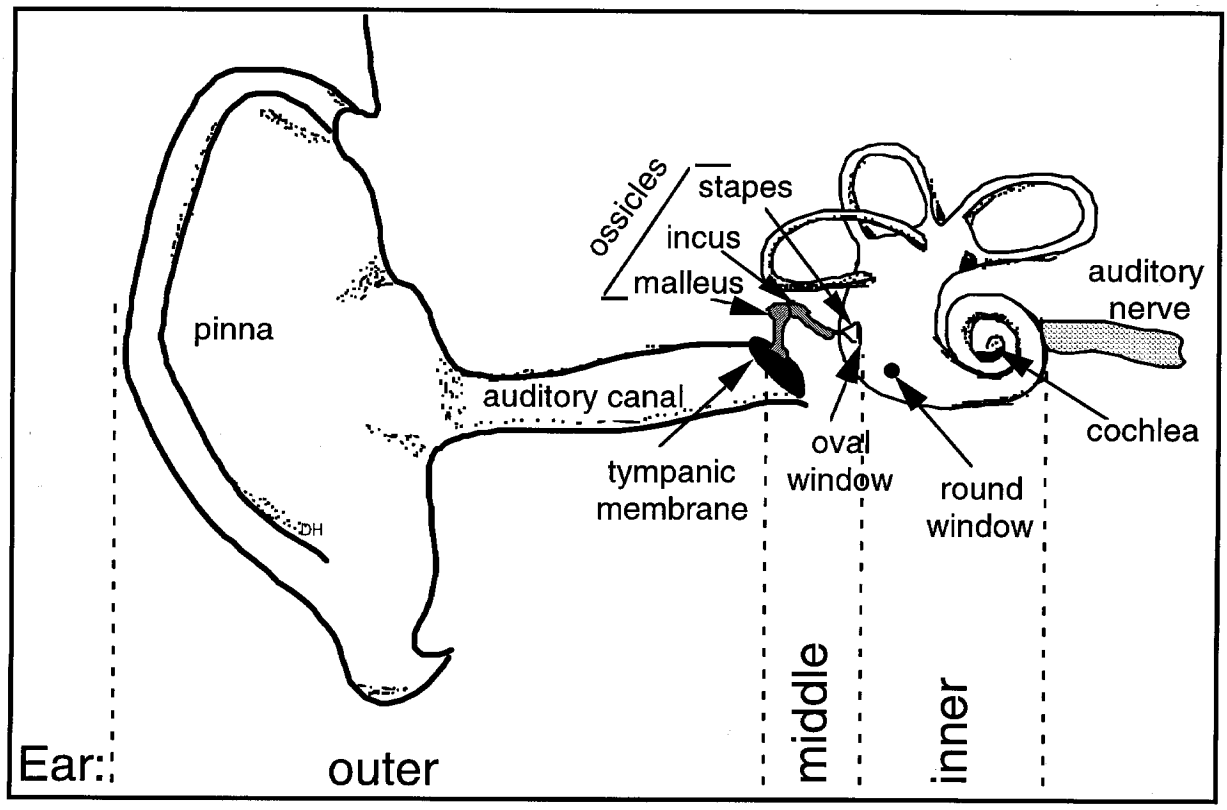


Figure 1. The human peripheral hearing system.

middle and inner ear, to the point of neural transduction in the organ of corti of the cochlea prior to transmission via neural links to higher centres of the brain (see Pickles 1982). Figure 1 illustrates the main elements of the peripheral hearing system. The outer ear consists of the pinna and auditory canal, and the boundary between it and the middle ear is formed by the tympanic membrane, or eardrum. Our pinnae help us to locate sound sources, since high-frequency sounds, above approximately 4.5 kHz, are collected more effectively from in front of the head than from behind it, due to the size and geometry of the pinnae relative to the wavelength at these frequencies. The auditory canal acts as an acoustic resonator with a dominant resonant frequency at approximately 3.5 kHz. The tympanic membrane converts acoustic pressure variations to mechanical vibrations.

The middle ear consists of the ossicles, three small bones comprising the malleus, incus and stapes, and is bounded at the oval window where the stapes makes contact with the cochlea. The ossicles function as an impedance transformer to provide sufficient force at the oval window to move the fluid which fills the cochlea in response to external acoustic vibrations. The inner ear consists of the cochlea in which the mechanical vibrations passing through the

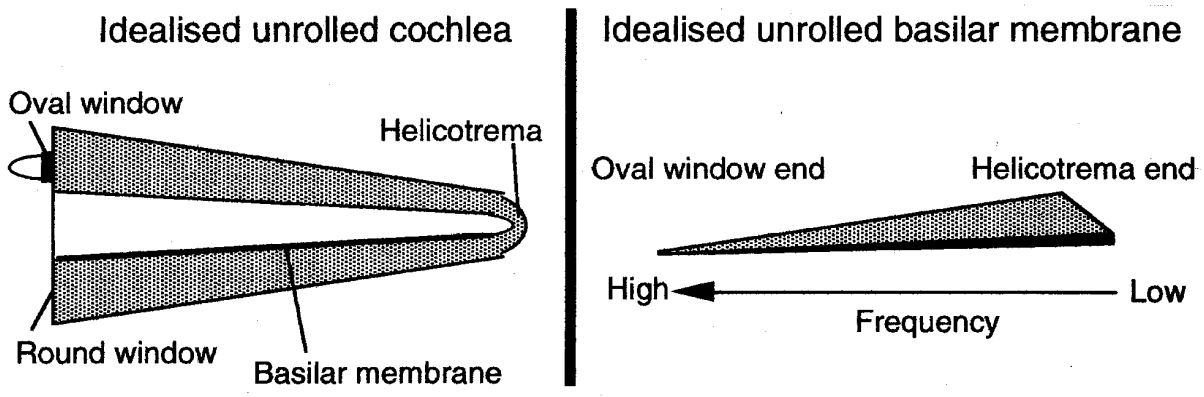


Figure 2. Idealised unrolled cochlea (left) and idealised unrolled basilar membrane (right).

https://doi.org/10.1017/S1355771897009011 Published online by Cambridge University Press

oval window are transduced into neural firings. Figure 2 shows an idealised unrolled cochlea. The cochlea is filled with an incompressible fluid, shown shaded in the figure, and inward movements of the oval window cause movement of this fluid along the length of the cochlea, via a small hole at the far end known as the *helicotrema*, and back to the round window which moves outward to compensate. Similarly, outward movements of the oval window are compensated for by inward movements of the round window.

Running around the cochlear spiral is the basilar membrane at the position illustrated on the left hand side of figure 2. This is the cochlear structure responsible for carrying out a frequency analysis of incoming sounds, based on its mechanical properties. The basilar membrane is narrow and thin at the oval window end, becoming wider and thicker towards the helicotrema, as illustrated on the right hand side of figure 2. Its mechanical properties are such that it responds well to high frequencies towards the narrow and thin end, and to low frequencies towards the helicotrema end. Maximum amplitude of movement of the membrane occurs at different positions, or *places*, along the length of the membrane depending on the component frequencies of the input sound.

The frequency analysis capability of the basilar membrane can be modelled as a bank of a large number of band-pass filters operating in parallel (Moore 1982), which in themselves do not exhibit symmetrical response curves in the frequency domain. One model often employed for modelling the shape of auditory filters is the GammaTone filter, which was introduced to describe the shape of the impulse response function of the auditory system as estimated by the reverse correlation function of neural firing times (Patterson 1976). In the time domain, the GammaTone filter is defined as

$$(1)\quad \mathrm{gt}(t) \propto t^{n-1} \exp(-2\pi bt) \cos(2\pi f_c t + \phi), \qquad (t \geqslant 0),$$

where n is the filter order, b is the impulse response duration, f_c is the carrier frequency, and ϕ is the carrier phase. The form of the function is that of an amplitude-modulated carrier tone of frequency f_c (Hz), with an envelope proportional to the Gamma distribution $[t^{n-1} \exp(-2^1 bt)]$.

The frequency selectivity of this filter bank can be described by means of a perceptual measurement of its *equivalent rectangular bandwidth* (ERB), which is that of an idealised filter with a rectangular bandwidth that would have the same output as an auditory filter if they were both stimulated with a flat spectrum input. This bandwidth is otherwise referred to as the *critical bandwidth* (Scharf 1970), due to the methods employed to measure it. The frequency selectivity of the filters is summarised by the following equation (Moore and Glasberg 1983), which relates the ERB for any particular filter to its centre frequency f_c:

$$(2)\quad \mathrm{ERB}(f_c) = (6.23 \times 10^{-6}) f_c^2 + (93.39 \times 10^{-3}) f_c + 28.52,$$

where $100\ \mathrm{Hz} < f_c < 10{,}000\ \mathrm{Hz}$.

3. PSYCHOACOUSTICALLY INFORMED ANALYSIS SYSTEM

Our model of the human peripheral hearing system currently concentrates on modelling the basilar membrane by implementing a bank of band-pass GammaTone filters. The output of this particular implementation is a spectrographic display, which has enabled its output to be directly compared with conventional spectrograms which are usually implemented by means of an FFT for speech (Howard, Hirson, Brookes and Tyrrell 1995) and musical sounds (Brookes, Tyrrell and Howard 1996). The model incorporates a band-pass filter to model the frequency response effects of the outer and middle ear.

The model is based on a bank of GammaTone filters whose bandwidths are determined on an ERB basis using equation (2). These filters operate in parallel, mapping closely the operation of the ear to our model. Such a model cannot be implemented particularly efficiently by means of conventional sequential processing techniques due to limitations on processing speeds and the need to carry out the calculations for each additional channel serially. Tyrrell, Howard and Beasley (1992) suggest that at least 24 filter channels are required to provide a useful basis for the observation of perceptually relevant display of acoustic cues, and that a real-time version of such a system would only therefore be possible using parallel processing techniques.

The filters themselves are implemented by means of transputers, which are processors designed to operate most efficiently in parallel. There are two variants of transputers used to implement the processing in this work: T800s and T9000s. The T800 transputer integrates a central processor, a floating point unit, 4 KB of static random access memory plus an interface for external memory, and a communications system onto a chip about 1 cm^2 in area. A transputer can be used in a single processor system or in networks to build high-performance concurrent systems. A network of transputers and peripheral controllers, as required in this work, is easily constructed using the point-to-point communication mechanisms uniquely provided by the transputer. The T800 is a 32-bit CMOS microcomputer with a 64-bit floating point unit and graphics support. The processing speed of a device can achieve instruction throughputs of 15 MIPS (millions of instructions per second). The 64-bit floating point unit provides single and double

https://doi.org/10.1017/S1355771897009011 Published online by Cambridge University Press

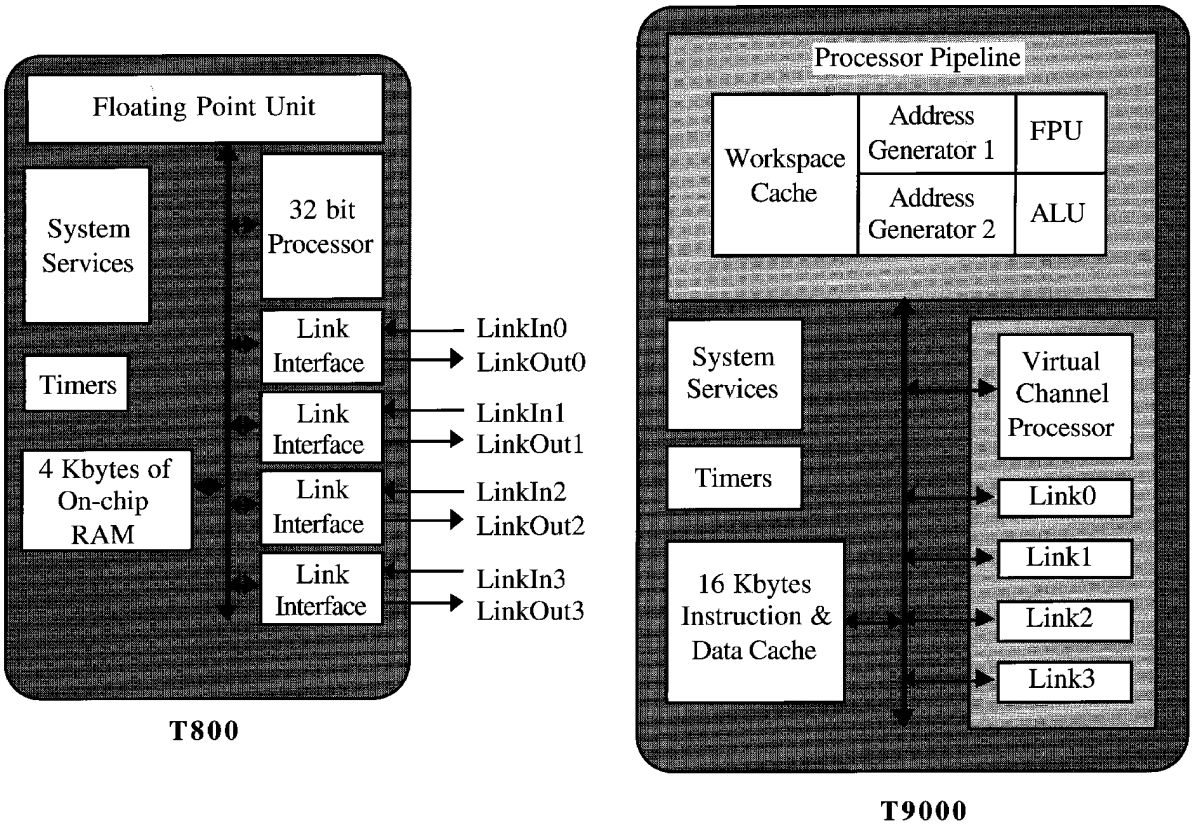


Figure 3. Transputer functional blocks.

length operations with a sustained rate of up to 2.25 Mflops (millions of floating point operations per second). The standard communication links allow networks of transputers to be connected and support bi-directional data rates of up to 2.35 MB/s. The basic functional blocks provided by a T800 transputer are shown in figure 3 (left).

The T9000 is a second-generation transputer; it has a superscalar processor, a hardware scheduler, 16 KB of on-chip cache memory, and an autonomous communications processor. The T9000's scheduler allows the creation and execution of any number of concurrent processes. The processes communicate by passing messages over point-to-point channels. Channels are unidirectional, and message passing is synchronised and unbuffered; the sending process must wait until the receiving process is ready, and the receiving process must wait until the sending process is ready. The use of this type of message passing removes the need for message queues and message buffers. Messages are passed over these links by the autonomous communications processor, the virtual channel processor. The T9000 gives a sustained integer performance of 60 MIPS and a floating point sustained performance of 10 Mflops. The basic blocks involved in a T9000 are shown in figure 3 (right).

The unique architecture of the transputers allows considerable parallelism to be achieved on a single processor. Each transputer, T800 and T9000, has four autonomous communication channels with their own processor. The integer and floating point processing units are also autonomous; thus a single transputer could be performing integer arithmetic, floating point arithmetic and communicating on all four links concurrently. If we add to this the possibility of building networks of these processors, the processing capabilities are immense.

Our first system (Swan, Tyrrell and Howard 1994) ran in real time to give an overall analysis bandwidth of 2,250 Hz with a 32-channel filter bank. The speech pressure waveform was input via an analogue-to-digital (AD) converter and processing was concurrently carried out by means of a bank of forty-two T800 transputers. The output was plotted as a conventional grey-scale spectrogram on the screen of the host PC compatible computer.

There were, however, a number of limitations and disadvantages with this system for many applications. These included: limited frequency bandwidth of the model, limited number of filters to maintain resolution up to 8 kHz, limited temporal resolution of display, and limited accuracy due to small sine and cosine lookup tables. A second generation system was designed and constructed using a mixture of T800 and T9000 transputers in an attempt to alleviate the problems of the first system.

https://doi.org/10.1017/S1355771897009011 Published online by Cambridge University Press

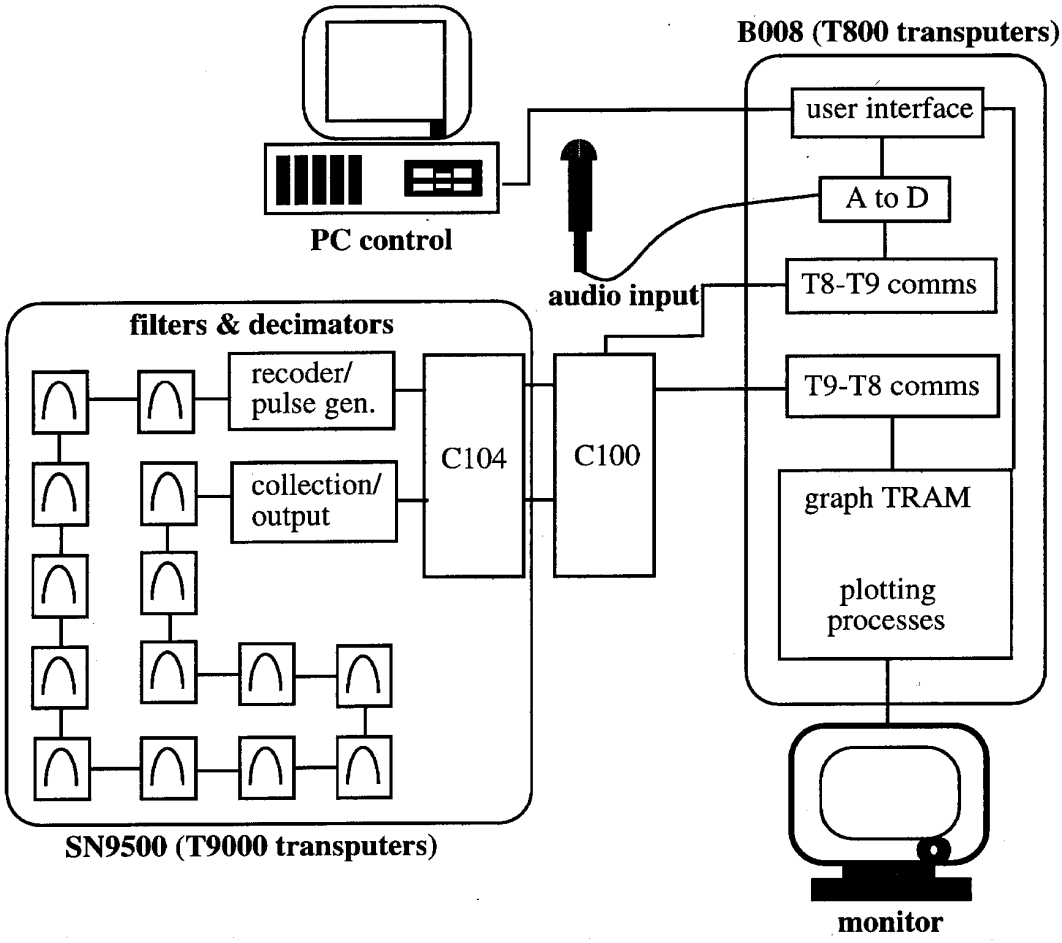


Figure 4. Block diagram of transputer processing system.

This system has significantly increased the processing power and the overall system bandwidth. It now consists of eight T800 transputers providing control, interfacing and graphics functions, together with sixteen T9000 transputers which provide the filter-bank processing. This complete system offers considerable processing power and communications bandwidth, the maximum processing achievable being approximately 1,080 MIPS and 178 Mflops, and the total system communications bandwidth being 240 MB/s. It should be born in mind that these could all be executing concurrently.

The input is again derived from an AD converter. The spectrographic output, however, is now provided by a dedicated G300 graphics chip which is controlled by one of the T800s; this complete structure is illustrated in figure 4. This hardware allows a bank of 64 filters to run in real time with an overall analysis bandwidth of 8 kHz. Additionally, the new system provides enhanced, mouse-driven, user control and a continuously scrolling output display. In order to implement 64 filters on 16 processors, optimisation of the available processing power was required. The GammaTone filter bank was designed to operate at multiple sample rates, the initial sampling frequency being 16.67 kHz. However, a number of the filters are able to operate at half that rate with no reduction in accuracy, and consequently they will have less data to process and require less processor time. The decimation process that halved the sampling rate was designed to remove from the audio signal all frequencies greater than the reduced Nyquist frequency. The half-rate data is produced by a decimation process which simply discards alternate samples of a low-pass filtered (to remove aliasing distortion) version of the original signal. Examples of the system output, which is produced in real time, are shown later in this paper.

4. PSYCHOACOUSTICS OF TIMBRE

Psychoacoustic descriptions of human perception of different aspects of sounds, such as pitch, loudness or timbre, are all based on the frequency analysis carried out by the basilar membrane, often referred to as *place analysis*, due to the variation of the place of maximum response on the membrane with frequency. The variation of critical bandwidth, or ERB, with

https://doi.org/10.1017/S1355771897009011 Published online by Cambridge University Press

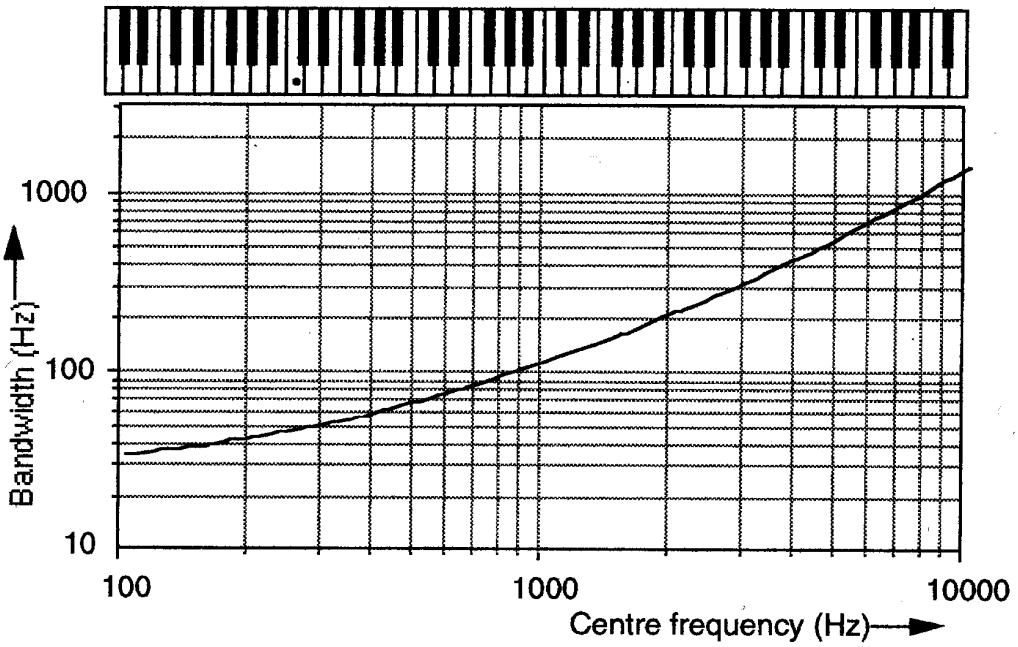


Figure 5. The variation of equivalent rectangular bandwidth ERB, just noticeable difference JND for frequency, and bandwidths of 1, 2, 4 and 7 semitones plotted against filter centre frequency (adapted from Howard and Angus 1996).

filter centre frequency is illustrated in figure 5. In this representation, filter centre frequency is equivalent to the frequency of components of the input sound, since any signal present at the output of a filter is due to that frequency being present in the input sound. A keyboard is presented on the frequency axis to set the plot in a musical context, with middle C being indicated by a black spot.

The figure clearly shows that the ERB increases as a function of filter centre frequency; thus the effective bandwidth of the auditory filters becomes wider with increasing frequency. The nature of the ERB scale is such that no harmonic above the seventh is resolved no matter what the value of f_0, and harmonics below the fifth are resolved separately. (A moment considering a few f_0 values with respect to the plot confirms this.) This is a vital conclusion, and one which underpins discussions relating to contemporary theories of timbre, as well as loudness and pitch perception (see, for example, Howard and Angus 1996).

Judgements of the timbre of a sound are highly subjective and prone to individual differences. Unlike pitch or loudness judgements, where listeners might be asked to rate sounds on scales of low to high or soft to loud, respectively, there is no scale on which timbre can be ordered. Timbre judgements are often made on the basis of giving a position between two extremes, such as bright and dark, or brilliant and dull. For any particular sound, a number of such positions on such descriptive scales might be rated, such as between 'bright' (1) to 'dark' (10), or 'pure' (1) to 'rich'(10). However, timbre judgements such as these are much more highly subjective and prone to individual differences than, for example, loudness judgements on a scale from 'soft' to 'loud'.

Timbre relates to any perceptible difference that occurs when acoustic variations are made to a given note that do not alter the perceived pitch, loudness or duration. Such acoustic variations might include the detailed manner in which the frequencies and amplitudes of the spectral components present change over the duration of the note. In particular, a note can be considered as having three phases: the 'attack' or 'onset' as it builds from silence to the 'steady state' where there is relatively little variation, to the 'release' or 'offset' when it returns to silence. The attack and release tend to occur over a few tens of milliseconds, and these have a very significant effect on the perceived timbre of notes. The importance of the attack phase was clearly demonstrated by Grey (1977) in his experiment to place instrumental sounds in a multidimensional space of three dimensions based on similarity ratings between the sounds given by his listeners. He identified the following acoustic factors with respect to each of the three axes: (i) increasing high-frequency components in the spectrum, (ii) synchronicity of attack between harmonics, and (iii) increasing high-frequency energy during the attack phase. Axes (ii) and (iii) focus particularly on the note onset phase, whilst axis (i) relates to the spectral energy distribution throughout the sound.

The 'tristimulus diagram' of Pollard and Jansson (1982) plots the timbral course of individual pitched notes on a triangular graph where the *X* axis relates

https://doi.org/10.1017/S1355771897009011 Published online by Cambridge University Press

to the proportion of energy in the fifth harmonic and above, and the *Y* axis to the proportion of energy in the second, third and fourth harmonics. The area of the triangle towards the origin (0,0) relates to the proportion of energy in the fundamental. The dynamic nature of the attack and release phases of a note would be represented by a line with an uneven time course which represents the relative energy weighting between the f_0 component, the second to fourth harmonics and the harmonics above and including the fifth, and during the steady state the line would remain basically stationary.

The tristimulus representation essentially gives the relative weighting between the f_0 component, those harmonics other than the f_0 component that are resolved, and those that are not resolved. Helmholtz (1877) deduced the following four general rules which follow much the same reasoning (but without the benefit of our sophisticated understanding of psychoacoustics) 'to show the dependence of quality of tone from the mode in which a musical tone is compounded', which provide an excellent point from which to consider the objective nature of timbre:

1. *Simple tones* like those of tuning forks applied to resonance chambers and wide stopped organ pipes, have a very soft, pleasant sound, free from all roughness, but wanting in power, and dull at low pitches.
2. *Musical tones*, which are accompanied by a moderately loud series of the lower partial tones, up to about the sixth partial are more harmonious and musical. Compared with simple tones they are rich and splendid, while they are at the same time perfectly sweet and soft if the higher upper partials are absent. To these belong the musical tones produced by the pianoforte, open organ pipes, the softer piano tones of the human voice and of the French horn.
3. If only the unevenly numbered partials are present (as in narrow stopped organ pipes, pianoforte strings struck in their middle points, and clarinets), the quality of the tone is *hollow*, and, when a large number of such partials are present, *nasal*. When the prime tone predominates the quality of the tone is *rich*; but when the prime tone is not sufficiently superior in strength to the upper partials, the quality of tone is *poor*.
4. When partials higher than the sixth or seventh are very distinct, the quality of the tone is *cutting or rough*. . . . The most important musical tones of this description are those of the bowed instruments and of most reed pipes, oboe, bassoon, harmonium and the human voice. The rough, braying tones of brass instruments are extremely penetrating, and hence are better adapted to give the impression of great power than similar tones of a softer quality. (*1954 translation of Helmholtz 1877: 118–19*).

These general rules provide a basis for considering timbral variations with respect to frequency domain representations of sounds, since they have a direct correlation with the nature of the critical bandwidth. His second and fourth general rules make a clear distinction between sounds in which partials higher than the sixth or seventh are distinct or not, and his first rule gives a particular importance to the f_0 component. The similarity here with the tristimulus representation as well as axis (i) resulting from Grey's multidimensional experiment is striking.

The frequency-domain manifestation of the four general rules is illustrated in figure 6 by means of spectrograms plotted from the real-time human peripheral hearing model. The frequency axis of each of the four spectrograms is based on the ERB scale itself, since the output from each filter is given equal distance on the frequency axis of the output spectrogram.

It was concluded above, in regard to the nature of the variation of ERB with centre frequency (equation (2)), that harmonics below the fifth are resolved by the peripheral hearing system whilst those above the seventh are not. The spectrogram of the sinewave exhibits a single horizontal band of energy which rises in frequency during the course of the sound. This is the f_0 component or first harmonic ($1*f_0$) rising from 128 to 160 Hz, and it is worth noting the presence of the f_0 component in all four spectrograms, since the f_0 component is present and its variation is the same in each case. In the second sound, where just the first five harmonics are present, it can be seen that they are isolated separately as five horizontal bands of energy. These harmonic bands of energy become closer together with increasing frequency on the spectrogram due to the nature of the ERB frequency scale. In the spectrogram of the fourth sound, where all harmonics are present up to the twentieth, the lowest six or seven harmonics appear as horizontal bands of energy and are therefore resolved, but the energy in the frequency region above the seventh harmonic is plotted as vertical lines, known as *striations*, which occur once per cycle. This is because all filters with centre frequencies above the seventh harmonic have bandwidths that are greater than the f_0 of the sound, and therefore at least two adjacent harmonics are captured by each of these filters. When two or more adjacent high harmonics are combined together, the result is a beat waveform whose beat period is equal to f_0, which is the origin of the striations on the spectrogram. In the spectrogram of the third sound, where only the odd harmonics up to the nineteenth

https://doi.org/10.1017/S1355771897009011 Published online by Cambridge University Press

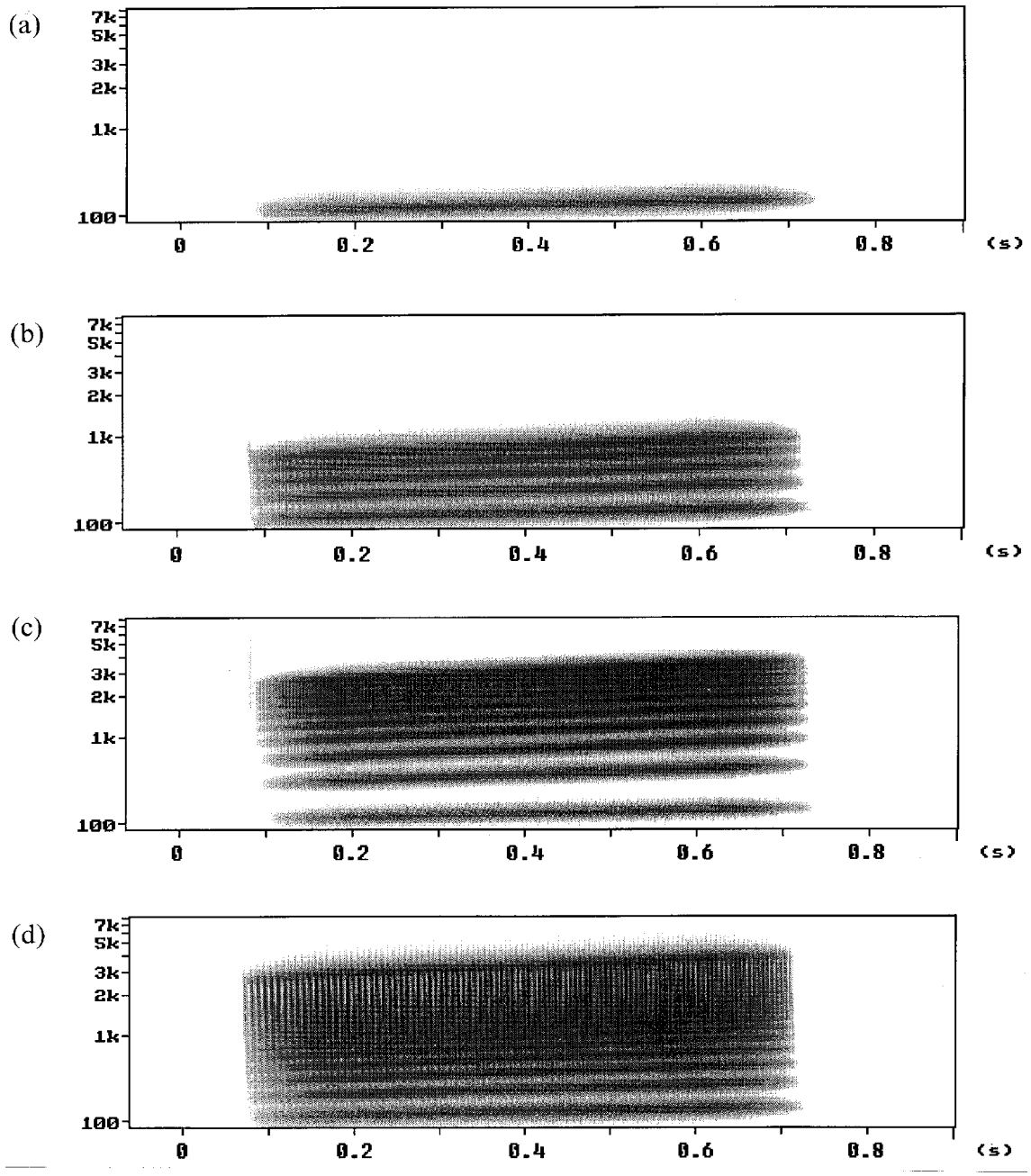


Figure 6. GammaTone spectrograms of synthesised sounds with f_0 varying between 128 and 160 Hz: (a) a sinewave, (b) a complex periodic tone consisting of the first five harmonics, (c) a complex periodic tone consisting of the odd harmonics only from the first to the nineteenth inclusive, and (d) a complex periodic tone consisting of the first twenty (odd and even) harmonics.

are present, there are approximately seven resolved horizontal energy bands, but in this case these are the lowest seven odd harmonics, or the first to the thirteenth harmonic inclusive. The point at which the odd harmonics cease to be resolved occurs where the spacing between them ($2*f_0$) is less than the ERB, and this will occur at about the position of the fifth to seventh harmonic of a sound with double the f_0, or an octave higher. This should occur somewhere between the tenth to the fourteenth harmonic of the sound with only odd harmonics, which occurs with the spectrogram.

The four examples illustrated in figure 6 would be described by the four general rules of Helmholtz in the order in which they are plotted as follows: the sinewave by rule 1, the complex periodic tone consisting of the first five harmonics by rule 2, the complex periodic tone consisting of the odd harmonics only from the first to the nineteenth inclusive by rule 3, and the complex periodic tone consisting of the

https://doi.org/10.1017/S1355771897009011 Published online by Cambridge University Press

Table. Summary of the frequency-domain properties as exemplified by the GammaTone spectrograms, example timbre descriptions and example instruments which fall under each of the four rules introduced by Helmholtz (1877).

Helmholtz rule	Frequency-domain GammaTone spectrogram	Example timbre descriptions	Example instruments
1	f_0 dominates (e.g. figure 6(a))	Pure Soft Pleasant Dull at low pitch Free from roughness	Tuning fork Wide stopped organ flues
2	Harmonics dominate (e.g. figure 6(b))	Sweet and soft Rich Splendid Dark Dull Less shrill Bland	French horn Flute Tuba Open organ flues Soft sung sounds
3	Odd harmonics dominate (e.g. figure 6(c))	Hollow Nasal	Clarinet Narrow stopped organ flues
4	Striations dominate (e.g. figure 6(d))	Cutting Rough Bright Brilliant Shrill	Other reed instruments Other brass instruments Other sung sounds Bowed instruments Harmonium Organ reeds

first twenty (odd and even) harmonics by rule 4. The GammaTone spectrograms derived from a model of human peripheral hearing provide a basis for a frequency-domain description of the timbre differences between these sounds, and it might be suggested that, in general, if f_0 dominates the GammaTone spectrogram the timbre is described by rule 1, if harmonics dominate the GammaTone spectrogram the timbre is described by rule 2, if odd harmonics dominate the GammaTone spectrogram the timbre is described by rule 3, and if striations are clearly present in the GammaTone spectrogram the timbre is described by rule 4. This is summarised in the table along with timbral terms which might be used, including those given by Helmholtz in his four rules, and acoustic instruments which tend to fall under each rule. As a caveat, it should be noted that the grouping of instruments into categories is somewhat generalised, since it may well be possible to produce sounds on these instruments where the acoustic output would fall into another timbral category based on the frequency-domain GammaTone description, for example by the use of an extended playing technique.

By way of examples to illustrate the relationship between the frequency-domain representation of the acoustic output from instruments in each respective Helmholtz rule category, GammaTone spectrograms are presented for the following notes held for approximately one second: middle C played on a stopped organ flue (figure 7), the G below middle C played on a tuba (figure 8), concert middle C played on a B clarinet (figure 9), and concert middle C played on a B trumpet (figure 11).

A stopped organ flue pipe supports odd harmonics only, since the pipe is open at the flue end and closed at the stopped end and therefore acts as a quarter-wave resonator (see Rossing 1990). In this example (figure 7), it can be seen that the f_0 component and the third harmonic are clearly present, with some evidence of energy in the region of the fifth harmonic. The f_0 is the dominant component. This stop is a 4' flute on an organ that is voiced in the baroque tradition with a fair degree of 'chiff' (see Hurford 1994) produced mainly on the musical twelfth by the third harmonic during the attack phase. This effect can be just observed at the start of the note where the third harmonic enters first, with a brief appearance of the fourth harmonic, before the f_0 component. This stop has a pure sound which is free from roughness. Hall draws attention to the fact, though subtle, that a pure tone from an electrical oscillator, in which only the f_0 component is present (see, for example, the upper plot of figure 6), has 'a timbre that changes from very dull at low frequencies to bright at high frequencies' (Hall 1980: 429). Helmholtz also notes this effect by describing simple tones (see above) as being 'dull at low frequencies'.

Figure 8 shows a GammaTone spectrogram for a tuba playing the G below middle C. The energy during the steady state portion of this note is mainly in its lowest five harmonics, where the majority is contained in the lowest three, with the fifth being weaker than each of the lower three and the fourth weaker still. This note clearly falls within the Helmholtz rule 2 category since the harmonics are resolved individually by the hearing system, and the timbre of such a note might be described as being

https://doi.org/10.1017/S1355771897009011 Published online by Cambridge University Press

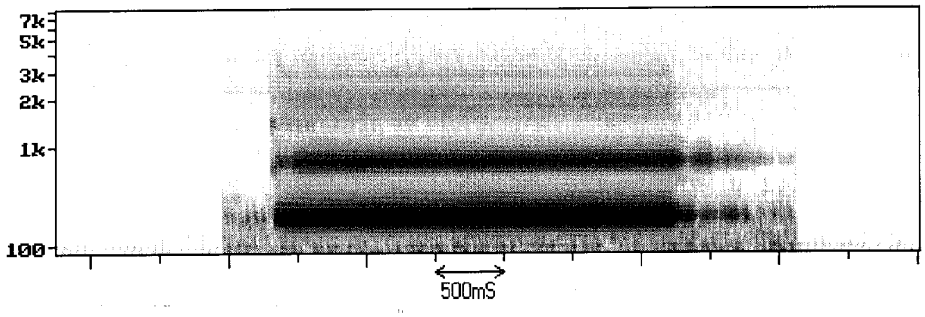


Figure 7. GammaTone spectrogram of middle C played on a stopped organ flue.

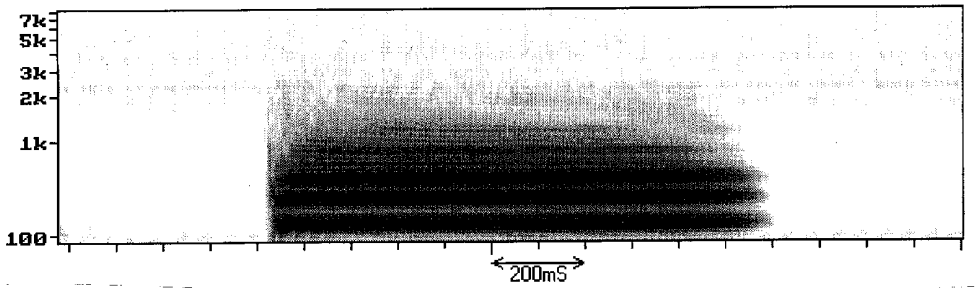


Figure 8. GammaTone spectrogram of the G below middle C played on a tuba.

rich but not cutting. The importance of the attack and release phases can be seen, even on the time scale presented in this spectrogram. During the attack, there appears to be a consonantal-like burst of energy with similar frequency domain characteristic to plosive bursts in speech (see Borden and Harris 1980, Baken and Danilof 1991, Kent and Read 1992) at around 1 kHz, which is followed by the entry of the third harmonic prior to the first and second, followed during the next 60 ms by the fourth and then the fifth harmonics. The release phase appears to begin approximately 150 ms before the sound ceases with the cessation of the fourth harmonic, followed by the fifth and then the lower three. (It should be noted that this is not intended as a detailed analysis of the attack and release phases which would require the use of a zoomed time scale such as presented in chapter 5 of Howard and Angus (1996).)

The spectrogram of middle C played on a B clarinet shown in figure 9 suggests that it is acoustically dominated by odd harmonics, but even harmonics are not totally absent. The first, third, fifth and seventh harmonics dominate the spectrum, but it can be seen that the even harmonics are increasingly in evidence between them with increasing frequency. The clarinet is closed acoustically at the reed end and open at the other. It therefore acts as a quarter-wave resonator, which supports odd-numbered harmonics and has an f_0 that is an octave below a similar length pipe open at both ends, such as a flute. The spectrum of the acoustic output from a single reed, such as that which excites the clarinet acoustically, will consist of all harmonics being present. There is therefore excitation energy at even harmonic frequencies, and the extent to which this is manifested in the acoustic output from the instrument depends on the relative

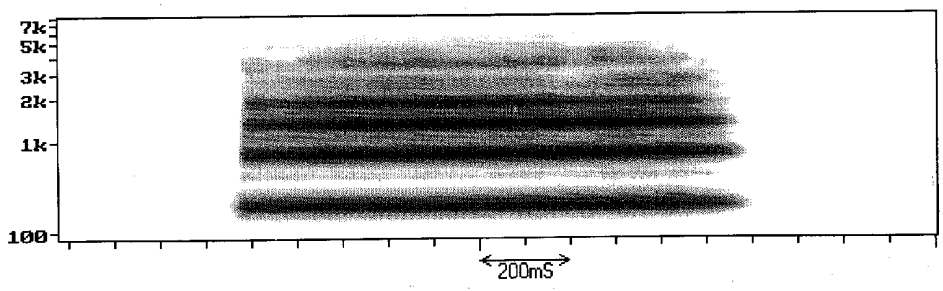


Figure 9. GammaTone spectrogram of concert middle C played on a B clarinet.

https://doi.org/10.1017/S1355771897009011 Published online by Cambridge University Press

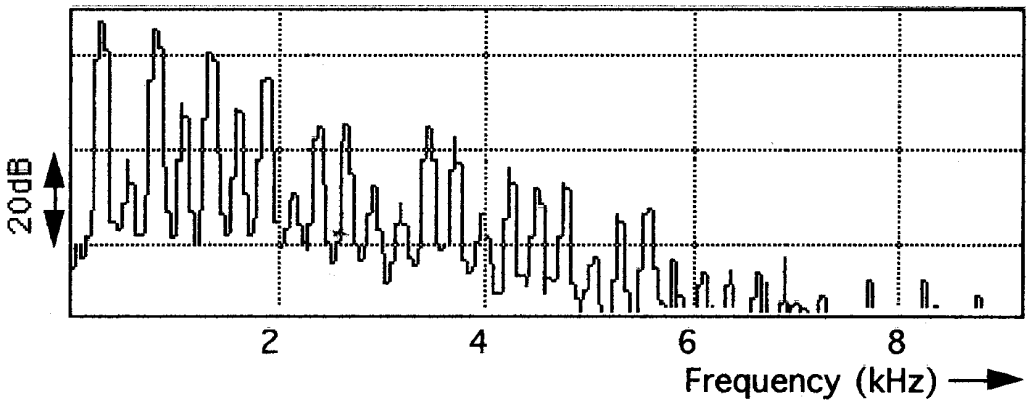


Figure 10. Long-term average spectrum during the steady-state portion of concert middle C played on a B clarinet (from Howard and Angus 1996).

response of the resonator to even compared to the odd harmonics. Sundberg gives this effect the following description:

> The resonator of the clarinet is basically a cylindrical open–closed tube. In other words, it has a quarter-wavelength resonator. This means that the even numbered modes are not welcome in the resonator, because the input impedance is low at their frequencies. A common misunderstanding is that these partials are all but missing in the spectrum. The truth is that the second partial may be about 40 dB below the fundamental, so it hardly contributes to the timbre. Higher up in the spectrum the difference between odd- and even-numbered neighbours are smaller. (*Sundberg 1989: 135*)

This is the case in this example, as can be observed in the long-term average FFT-based spectrum plotted in figure 10. The second harmonic is some 40 dB below the f_0 component and the differences between the odd and even harmonics become less pronounced with increased frequency.

The sound of the clarinet is often described as having a nasal timbre, and this has much to do with the fifth and seventh harmonics. The seventh is of particular importance since it is isolated in this case due to the low amplitude of the eighth harmonic which would otherwise be combined with it at the peripheral hearing analysis stage. Organists familiar with the use of the *septieme* (1 1/7') stop, which sounds the true seventh harmonic (a flat B 6 when a C4 is played), will appreciate its particular timbral significance, as well as the significance of the fifth harmonic by means of the more commonly found *tierce* (1 3/5') stop, which sounds the true fifth harmonic (a just tempered E6 when a C4 is played).

Figure 11 shows the GammaTone spectrogram for a concert middle C played on a B trumpet. The presence of harmonics above the seventh is clearly evident with their energy manifesting itself as striations which are very close together, due to the time scale required to view the complete note. Harmonics up to and including the sixth are clearly isolated, but those above the sixth are not. The timbre of such a note might be described as being rough and cutting. The release stage of this note is of interest, where the player reduces the energy in the high harmonics first over some 200 ms before the lower harmonics are gradually lowered in amplitude before the first and second cease into the following silence.

5. CONCLUSIONS

The human peripheral hearing system has been described in the frequency domain in terms of the

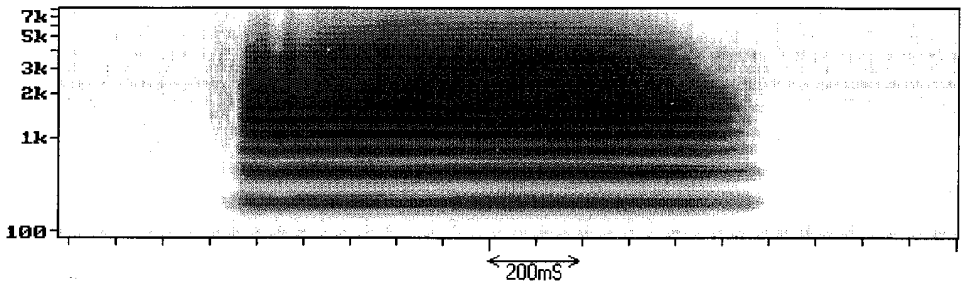


Figure 11. GammaTone spectrogram of concert middle C played on a B trumpet.

https://doi.org/10.1017/S1355771897009011 Published online by Cambridge University Press

frequency response of the outer and middle ears, followed by the frequency analysis characteristics of the inner ear in terms of the bandwidth variation with centre frequency, as well as the response curve shape of a bank of band-pass filters. This model has been implemented as a real-time spectrograph by means of one of the largest second-generation applications and one of the few that makes use of a mixture of first- and second-generation transputers. Some of the key issues that relate to the psychoacoustics of timbre have been explored, with particular reference to the unique insights demonstrated by the work of Helmholtz in the nineteenth century, and these have been illustrated with spectrograms from the transputer system for synthesised sounds and the outputs from a selection of acoustic musical instruments. These analyses demonstrate a clear relationship between the output from the human peripheral hearing system and those aspects of the frequency-domain representation of sound that appear to be relevant to commonly understood timbral descriptions of sounds.

REFERENCES

ANSI. 1960. *American Standard Acoustical Terminology*. New York: American National Standards Institute.

Baken, R. J., and Danilof, R. G. 1991. *Readings in Clinical Spectrography of Speech*. San Diego: Singular Publishing Group.

Borden, G. J., and Harris, K. S. 1980. *Speech Science Primer*. Baltimore: Williams and Wilkins.

Brookes, T. S., Tyrrell, A. M., and Howard, D. M. 1996. Musical analysis using a real-time model of peripheral hearing. *Proc. Int. Computer Music Conference* (ICMC-96), pp. 79–82.

Grey, J. 1977. Timbre discrimination in musical patterns. *Journal of the Acoustical Society of America* **64:** 467–72.

Hall, D. E. 1980. *Musical Acoustics, An Introduction*. Belmont, California: Wadsworth.

Helmholtz, H. L. F. von. 1877. *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, 4th edn, translated by A. J. Ellis. New York: Dover (1954).

Howard, D. M., Hirson, A., Brookes, T., and Tyrrell, A. M. 1995. Spectrography of disputed speech samples by peripheral human hearing modelling. *Forensic Linguistics* **2**(1): 28–38.

Howard, D. M., and Angus, J. A. S. 1996. *Music Technology: Acoustics and Psychoacoustics*. Oxford: Focal Press.

Hurford, P. 1994. *Making Music on the Organ*. Oxford: Oxford University Press.

Kent, R. D., and Read, C. 1992. *The Acoustic Analysis of Speech*. San Diego: Singular Publishing Group.

Moore, B. C. J. 1982. *An Introduction to the Psychology of Hearing*. London: Academic Press.

Moore, B. C. J., and Glasberg, B. P. 1983. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America* **74**(3): 750–3.

Patterson, R. D. 1976. Auditory filter shapes derived with noise stimuli. *Journal of the Acoustical Society of America* **59:** 640–54.

Pickles, J. O. 1982. *Introduction to the Psychology of Hearing*. London: Academic Press.

Pollard, H. F., and Jansson, E. V. 1982. A tristimulus method for the specification of musical timbre. *Acustica* **51:** 162–71.

Rossing, T. D. 1990. *The Science of Sound*. New York: Addison Wesley.

Scharf, B. 1970. Critical bands. In J. V. Tobias (ed.) *Foundations of Modern Auditory Theory*, Vol. 1. London: Academic Press.

Sundberg, J. 1989. *The Science of Musical Sounds*. London: Academic Press.

Swan, C., Tyrrell, A. M., and Howard, D. M. 1994. Real-time transputer simulation of the human peripheral hearing system. *Microprocessors and Microsystems* **18**(4): 215–21.

Tyrrell, A. M., Howard, D. M., and Beasley, N. A. 1992. Transputer model of the human peripheral hearing system. *Microprocessing and Microprogramming* **35:** 619–24.

https://doi.org/10.1017/S1355771897009011 Published online by Cambridge University Press