Proceedings of The Institute of Acoustics

DETERMINANTS OF THE TONAL QUALITY OF MUSICAL INSTRUMENTS

P.J. SIMPSON

DEPARTMENT OF PSYCHOLOGY, UNIVERSITY OF SURREY

The musical instrument families differ in their musical compass, dynamic range and agility. They are also perceived as having distinct tonal qualities or timbre. All these properties are utilized by musicians in the composition and performance of music. The tonal qualities of instruments have been related to the particular physical processes which govern the production and radiation of sound from instruments. Modern analytic procedures have been applied to the processes involved in the excitation of tuned columns in organ pipes, the action of reeds or the lips on the airflow in the woodwind and brass, and the excitation of strings in the violin and piano. Work in the field of musical acoustics has attempted to define the acoustic features associated with perceived tonal quality.

ACOUSTIC DETERMINANTS OF TIMBRE

Work on the acoustic determinants of timbre poses problems of definition, measurement and the sheer complexity of the phenomena. The American Standards Association [1] seeks to define timbre by a process of exclusion. It proposes that timbre is that attribute of auditory sensation in terms of which a listener can judge that two sounds, similarly presented and having the same loudness and pitch, are dissimilar. Investigation of the determinants of perceived timbre in musical instruments has brought out the importance of both transient and steady state acoustic features. The way in which a complex sound grows and decays is of crucial importance. However, during these transient phases the relative loudness of individual harmonic components can covary. In the case of brass instruments, Rissett and Wessel [2] report that a key feature of tonal quality or timbre is an increase in spectral content with amplitude. Moreover there is a non linear relationship between the different harmonics present in the sound although the individual harmonics are all functions of the fundamental regardless of the general level. Thus it is difficult to achieve the independence of factors envisaged by the American Standards Association.

SIGNIFICANCE OF THE INITIAL ATTACK TRANSIENT

In general terms a musical tone may be considered to consist of a starting transient, a more or less steady position, and then a decay section whose pattern may be subject to interference between the instrument vibration modes and the room modes. In some instruments, for example, the piano, there is no steady state. A piano notes starts with an impulse of broad spectrum noise associated with the contact between the hammer and strings. There follows a initially fast and then slower decay of string vibrations which are accompanied by complex interactions with other strings within the instrument, Weinreich [3]. In a number of instruments the duration of the starting transient, steady state portion and decay period is under the control of the player and governed by considerations of musical style and interpretation.

In practice the significance of each stage for the listener depends on the context in which it occurs. Thus Rissett and Wessel found that the details of the starting transient are more audible in long sustained tones than brief tones. This effect is easily demonstrated with inexpensive electronic keyboards which offer a range of synthetic instrument tones. Short notes produce a reasonably sounding copy. But sustained notes reveal the lack of the appropriate starting transient when that exists in the original instrument. The initial attack transient of instruments has been subject to detailed analysis. This feature could uniquely define an instrument since it reflects the primary sound generating process characteristic of an instrument. Benade [4] has discussed the problem players of brass instruments have in stabilizing lip tension in order to inject pulses of air into the instrument and establish a mode of oscillation at a chosen pitch. While the lips have a high mass relative to air within the instrument and thus have a major role in establishing a mode of vibration, they are responsive to the state of the air column. Thus initially there will be some instability and hunting before the intended mode of vibration is achieved. A parallel case can be made for other instruments. In the violin family, the tone of the instrument may include an initial random frequency component which corresponds to the erratic vibration which occurs when the string is first set in motion by the bow.

ANALYSIS OF TRUMPET TONES

Rissett and Matthews [5] reported work based on the pitch-synchronous analysis of trumpet tones which provides a detailed cycle by cycle picture of the amplitude of a large number of harmonics as they change over time. In the case of the data actually published for a 200 ms D4 trumpet tone it appears that the 1st and 2nd harmonic reach peak amplitude after 20 ms and then abruptly diminish. Higher harmonics now increase in amplitude so that the 5th, 6th and 7th harmonic peak at 40 ms but thereafter decay at different rates. This note has no effective steady state component and the component harmonics grow and decay in a complex fashion. Rissett and Matthews do not represent these results as typical. They remark that the notes within an instrument's compass may all show a different pattern of development. Luce and Clarke [6] undertook a much more extensive investigation and sampled the performance of different players and instruments within the brass family and a number of notes on each instrument.

PROBLEMS WITH THE ANALYSIS OF ATTACK TRANSIENTS

Both Luce and Clarke, and Matthews and Rissett employed an analysis procedure which provided a Fourier analysis of the notes on a cycle by cycle basis. This technique allows the spectral and temporal resolution of the transient features and high harmonics which a bandpass filtering technique precludes. A cycle by cycle Fourier analysis assumes that the sound to be analysed is quasi-periodic. However, problems may arise in the initial attack stage of a note which may be aharmonic or accompanied by variations in the pitch. A Fourier analysis programme can present an erroneous picutre of the spectral components of the initial states of a note if the cycle length is changing as pitch varies. Both groups of workers acknowledge this problem. Luce and Clarke supplied the fundamental frequency of the note presented for analysis by proceding it with a sinusoid of the same frequency to label it and indicate the beginning. Thus the computer programme was primed for the frequency and starting point. Matthews and Rissett processed their notes backwards becasue, as they noted, their pitch synchronous programme gave more information when proceeding backwards in time. In practice Rissett and Matthews' technique may prove efficient because the programme is effectively primed by processing the pitch stable section of the note first. Any disruption arising from pitch changes or the initial periodic section of the sound is delayed until the very last moment of the note. If the note is periodic but the pitch and therefore the cycle length is varying at the start then a backwards processed note will produce a different resolution of the harmonic components to that obtained when the note if forward processed. This hypothesis was tested on a sample of trumpet notes which, after recording and conversion to a digital representation, were processed in either the original temporal direction or in the reversed order, i.e. backwards. The results obtained from pitch synchronous analysis of forward and backward processed notes revealed differences in the time course of development of the component harmonics. These differences were only evident to a marked degree in the initial stages of the notes. The steady state and decay section of the notes corresponded except for small scale discrepancies in the amplitude of the harmonics. These results cast doubts on the adequacy of earlier published accounts of the initial attack phase in the case of trumpet notes.

PSYCHOLOGICALLY VALID ANALYSES OF MUSICAL NOTES

Rissett and Matthews did not claim that their findings were generally representative of trumpet tones. However, they reported that the incorportion of these initial attack transients features into a synthesised trumpet tone led that tone to be judged indistinguishable from real trumpet tones. In practice, as Rissett and Matthews note, this could be because the inclusion of the initial transient components increased the perceived roughness or complexity of the synthesised note rather than because of the inclusion of an essential component. They found that they were able to employ a simplified model of the attack transient phase and still achieve good synthesis. This observation suggests that the very detailed temporal resolution provided by pitch synchronous analysis may be unrepresentative of the resolution, spectral and temporal, achieved by the ear. Pollard and Jansson [7] have argued that acoustic analyses should employ procedures which match the performance of the ear in so far as the critical parameters are known. In their investigation, they analysed musical tones using ⅓ octave filters whose output were measured, stored and then subsequently averaged to produce values taken at 5 ms intervals and then converted into loudness values. This approach is modelled on the assumption that the ear behaves as a set of 24 bandpass filters with an overall averaging time for responses to fast changes in loudness of approximately 5 ms. This form of analysis may be held to produce an accurate and representative picture of an evolving note provided that the pitch of the note coincides with the central frequency of the bandpass filters. Such an analysis procedure will be free of the artefacts associated with the pitch synchronous procedure. However, the bandpass filter method does not yield an analysis which can of itself differentiate between pitch variation, noise or aperiodic components, and periodic tones as the human listener can. The human ability depends on a capacity for pattern analysis and the use of knowledge of the sound structure of instruments within a pattern recognition process which operates in a temporal, spatial and musical context.

TIMBRE PERCEPTION FROM THE LISTENER'S PERSPECTIVE

Investigations of the transient component of musical tones have been justified on the grounds that this component provides crucial information for the identification of instrument type. Luce and Clarke argue that the relative amplitudes of the harmonic components during the steady state region of a note are of secondary importance. Identification of instrument type is reportedly maintained under circumstances in which the relative amplitude of the harmonic components can be drastically modified. This can occur when sound is generated by electro-acoustic devices, for example a loudspeaker, as a result of reinforcement and cancellation effects in the auditorium, or with variations in the angular distribution of the sound radiating from an instrument with frequency. From the listener's point of view some or all of these factors may obscure not only the steady state but also information about the initial attack transient and the decay portion.

In fact the conditions under which reliable instrument identification can be maintained has not been systematically explored. Should we consider the identification of individual notes or phrases, with the instrument on its own or in an ensemble, playing familiar or unfamiliar music? Prior experience listening to music in the room used for the investigation would be an important factor following Plenge's [8] demonstration of the effects of familiarity with room acoustics on sound location. The problem of defining a representative condition is further complicated by the observation that the timbre of an instrument varies with the position of a note within the instrument's compass. Thus the clarinet has two major regions of tone colour. The lowest octave of its compass has a pecularly dark tone and is designated the chalumeau register. The upper registers are perceptually clear and allow expressive variation. Performance in the two regions involves changes in the technique used to control the operation of the reed by the player which may be reflected in both the initial attack transient, steady state harmonics and dynamics of the notes.

Research into the determinants of instrumental timbre has followed the path of experimentation, analysis and synthesis. The analysis of the sound structure associated with an instrument has often been limited to only a few notes, of limited dynamic range and separated from any musical context. Similary investigations of the perceptual determinants of timbre have employed tasks in which the listener classifies tones from a restricted set of instrumental notes presented in isolation or even in truncated form, Shepherd and Simpson [9]. There are good practical, methodological and, sometimes, theoretical reasons for using this approach to investigation. However, the search for a reliable psychophysical relationship between acoustic features and perceived timbre has foundered in the face of uncertainly about the acoustic features associated with an instrument's performance and the information actually available to the listener. A parallel situation is found in the context of recent attempts to understand speech perception.

DETERMINANTS OF THE TONAL QUALITY OF MUSICAL INSTRUMENTS

AN INTERPRETIVE APPROACH TO THE PERCEPTION OF ACOUSTIC FEATURES

It has long been accepted that acoustically speech has a formant structure in the sense that energy is concentrated in various parts of the spectrum. This arises from resonances in the vocal tract resulting from the shape of the vocal tract and relative position of the articulatory apparatus. Vowel sounds, for example, can be characterised in terms of their formant frequences and the amount of energy at each of the formants. Using a speech spectrogram to analyse continuous speech has revealed that there are often rapid changes in the frequency of a given set of formants. These changes are known as formant transitions and have been found to be important acoustic cues for the perception of consonants.

The task facing any speech recognition system whether embodied in man or, more recently, computer is to translate from acoustic features to words. One major problem in achieving speech recognition with the computer arises from the variability of the sound spectrum associated with any particular word. This is true both for the speech produced by a single speaker and when we compare different speakers. Levinson and Liberman [10] suggest that this arises because the components of the vocal tract do not take up fixed configuration. Rather they may move through a sequence of target positions defined by consonants and vowels. The vocal tract is in complex and continuous motion and the sequence of movements will reflect the target positions demanded by the particular word which is being pronounced. In practice the movements corresponding to a particular linguistic unit will overlap and combine in a individual fashion. Similarly the position of any unit within a phrase, degree of emphasis and rate of pronouciation will be reflected in the vocal tract movements. These changes in turn will be apparent in the sound spectrogram. Thus it is hard to define the elements of the acoustic message which can be mapped onto the linguistic domain in a direct unequivocal fashion. The situation is further complicated by the effects of room acoustics, background noise and the transmission channel. Computer speech recognition systems can achieve significant levels of recognition using pattern recognition procedures. These procedures make use of nromalisation and transformation operations which reflect the range of variation arising from normal speech dynamics. However, Levinson and Liberman report that recognition can be substantially improved if the computer program can utilize syntactic knowledge which constrains the sequence of words within an utterance. Adding knowledge of meaning and topic adds another major constraint which can be utilized in the interpretive recognition process. Defining and utilizing that knowledge in a coordinated interpretive process presents major problems of design and evaluation. However, it provides a more productive model of the human capacity to interpret uncertain acoustic information.

ELEMENTS OF A MODEL FOR TIMBRE PERCEPTION

As we have seen the uncertain nature of the acoustic features involved in speech recognition has a parallel in the variability of the acoustic features associated with the tonal qualities of musical instruments. In the case of speech, this variability arises from differences between speakers with regard to voice pitch, size and shape of vocal tract, articulatory habits and dialect. But within an individual speaker, interactions between individual elements of the articulatory movements and other dynamic factors will also influence the acoustic features present in continuous speech. On balance the marked 'organic' variations in size of vocal tract element found in man will not have a parallel in manufactured instruments taken from an instrument family. However, within a family of instruments there are members differentiated by pitch and compass. These differences are reflected in their construction, for example, size of mouthpiece, length of strings, size of instrument body and playing technique used to achieve a musical tone. The tuba and the trumpet, double bass and violin, and the bass saxophone and soprano saxophone present different tasks for the player and have different sounds. In identifying an instrument on the basis of acoustic information alone, the perceiver can use the range of pitch of the notes produced to limit the number of possibilities. But the way in which a player can execute a sequence of notes within an instrument's compass is a significant cue to instrument type. The agility that can be achieved both with regard to speed of executing notes and variation in dynamic range and timbre will be reflected in the type of music which can be played on an instrument. In fact recognising a familiar theme known to be carried by the horn, oboe or trumpet, for example, can lead to the identification of an instrument on that basis alone. When trying to identify instruments which share a common pitch and compass, and have similar tonal qualities, for example the oboe and soprano saxophone, the listener may have to adopt a more analytic approach. The listener may be able to discriminate between the instruments on style of performance as it reflects the demands of reed control, the mechanical operation of the instrument and variations of tonal quality with position in the instrument's compass and output level.

INFERENCE AND REASONING IN TIMBRE PERCEPTION

The use of these more subtle cues in the recognition process requires, of course, that the perceiver is familiar with the detailed characteristics of instruments and has integrated that knowledge into the recognition process. It is also requires that the perceiver is able to discriminate the pitch, loudness and timbre of individual notes and changes in these parameters both within and between notes. Yet this appears to depend on the reception and classification of acoustic features which earlier were described as uncertain since they varied across an instrument's dynamic range and compass and were open to interference before reaching the listener. This objection can be contained if we allow that under normal listening conditions perceptual classification is a process which, over time, samples a number of information sources to achieve a classification. As in the case of speech perception, uncertain or ambiguous information derived from primary acoustic features can be evaluated in the context of higher order information about performing style or musical theme to achieve classification. The process is akin to the operation of inference originally proposed by Helmholtz [11] and subsequently by Oatley [12] to account for the maintenance of visual perception under conditions of degraded or ambiguous perceptual information. This idea has recently taken on a tangible form in the shape of a number of computer programmes which can process grey scale representations of visual scenes to identify objects. These objects are represented in the form of higher order descriptions within the programme and control the procedure which maps between the grey scale representation and the object description domain to achieve classification. This approach to the perceptual process entails a reasoning operation in which fragmentary or incomplete information is reconciled with structural knowledge. While the process may contain a number of individual elements of pattern description and classification, the resulting information must be effectively integrated to achieve structural recognition and identification.

CONCLUSION

One way to understand the phenomena of the perception of instruments tonal quality is to adopt the view that timbre perception is an interpretive activity combining pattern recognition and classification based on uncertain information and consideration of higher order factors like performance style and musical form. This approach is consistent with Bregman and Pinker's [13] view that timbre seems to be a perceptual description of a stream, not of an acoustic wave form. Abandoning the requirement to seek out a strict psychological relationship between acoustic features and perceived tonal qualities may enable us to understand better the relation of representation underlying perceived qualities like the golden sound of the brass or the silvery sound of the flute.

REFERENCES

1. AMERICAN STANDARD ASSOCIATION 1960. American Standard Acoustic Terminology, New York.

2. J.C. RISSETT and D.L. WESSEL 1982 in The Psychology of Music Academic Press. Exploration of Timbre by Analysis and Synthesis.

3. G. WEINREICH 1979 Scientific American 240, 118-127. The Coupled Motions of Piano Strings.

4. A.H. BENADE 1976. Fundamentals of Musical Acoustics London and New York: Oxford University Press.

5. J.C. RISSETT and M.V. MATTHEWS 1969 Physics Today 22, 23-30. Analysis of musical instrument tones.

6. D. LUCE and M. CLARKE 1967 Journal of the Acoustical Society of America 42, 1232-1243. Physical Corrlates of Brass Instrument Tones.

7. H.F. POLLARD and F.V. JANSSON 1982 Acustica 51, 249-262. Analysis and Assessment of Musical Starting Transients.

8. G. PLENGE 1972 Acustica 26, 213-221. Uber das problen der im-kopf-lokalisation.

9. R. SHEPHERD and P.J. SIMPSON 1981. Institute of Acoustics Spring Conference University of Newcastle, April 1981. Perception of the timbre of musical instruments.

10. S.E. LEVINSON and M.Y. LIBERMAN 1981 Scientific American 244, 56-68. Speech Recognition by Computer.

11. H. von HELMHOLTZ 1966 Treatise on physiological optics, Vol. 111. Edited by J.P.C. Southall, 1962. New York, Dover.

12. K. OATLEY 1978. Perceptions and Representations: Methuen and Co. Ltd.

13. A.S. BREGMAN AND S. PINKER 1978 Canadian Journal of Psychology 32, 20-31. Auditory streaming and the building of timbre.