PERCEPTION OF THE TIMBRE OF MUSICAL INSTRUMENTS

R. SHEPHERD AND P.J. SIMPSON

UNIVERSITY OF SURREY

Timbre has been described as the property which differentiates sounds of equal loudness, pitch and duration (e.g. American Standards Association), and this lack of a clear definition has probably led to less empirical work on the perception of timbre compared with either loudness or pitch. Plomp (1970) reviewing this literature concluded that little progress has been made in the previous hundred years. This may in part be due to the difficulty of analysing the physical signal,but the lack of methods of dealing with psychophysical relationships is probably more important. With the advent of multidimensional scaling procedures (Shepard, 1962) and easy access to computing facilities, work in this area has increased during the last ten years. Physical characteristics which are said to lead to differences in the perception of timbre include changes in the steady state spectral energy, the phase of the components, and temporal variation in the attack and the decay of the tone. Hence timbre might be expected to be a multidimensional attribute of the acoustic signal. However, as we will see the perception of timbre may also be related to the context of the sound and vary between individuals.

In this paper multidimensional scaling and verbal rating procedures will be described briefly, together with some of the results from studies using these methods. The advantages and disadvantages inhererent in each method will be discussed along with problems associated with the use of either method in the study of the perception of the timbre of musical instrument tones.

The use of multidimensional scaling techniques is based upon the assumption that stimuli can be represented by points in an n-dimensional space, where the dimensions relate to the dimensions underlying the perception by the subject. Stimuli are presented in pairs or triads and subjects rate the dissimilarity between these stimuli. The subjects are not usually told what factors to use in this rating. The dissimilarities are assumed to represent psychological differences between the stimuli, and these may then be represented by distances between points plotted in some form of space. If there are x points then a space of dimensionality (x-1) would be necessary in order to represent the interpoint differences exactly, but in general a much lower dimensionality will provide an adequate description. With lower dimensionality the relationships will be distorted, and the degree of distortion is expressed in a quantity called stress. The solution to the analysis is chosen by the investigator to give a low number of dimensions and not too great a stress, but the choice is usually somewhat arbitrary and may be related to the investigator's expectations concerning the type and number of dimensions. The dimensions in the chosen solution may be thought of as being real dimensions underlying perception, or they may be considered to be merely a convenient way of representing the data with no meaning attached to the actual dimensions.

Grey (1977) used this technique in a study with simplified tones derived from real instrument tones, and found three factors to be important. These may be characterised as (i) spectral energy distribution, (ii) synchrony in the attack and decay of upper harmonics, and (iii) the presence of high-frequency low-amplitude energy in the attack. An alternative interpretation of these latter two factors is that they relate to a grouping by instrument family. Grey and Gordon (1978) confirmed the interpretation of the spectral energy factor by exchanging the spectral envelopes of pairs of tones from different instruments which caused the stimuli to change position along this factor. Plomp and Steeneken (1969) investigated the importance of phase in the perception of timbre. They found that although phase could be detected it was far less important than spectral amplitude, and in a real environment multiple reflections would obscure the phase information.

Verbal rating procedures involve the presentation of the stimuli individually, and the subject is required to rate them on bipolar scales such as dull/bright or hard/soft. The scales are numbered from 1 for dull to 7 for bright and the rating gives a numerical value for how bright the tone is judged to be. Von Bismarck (1974) used this procedure in a study with 30 rating scales and 35 synthesised tones. Factor analysis of the results showed the 30 rating scales reduced to four factors, which may be described as sharp/dull, empty/full, compact/scattered, and colourless/colourful. The use of factor analysis makes the method similar to multidimensional scaling procedures in uncovering the factors in the perception but there are differences between the methods.

In the case of multidimensional scaling procedures it is not necessary to specify the dimensions, whereas with verbal ratings the choice of the scales is of crucial importance. With verbal rating scales different subjects may use the verbal labels differently, and in fact it may not be possible to devise verbal scales which necessarily characterise the dimensions of perception. On the other hand multidimensional scaling, which does not require prior assumptions about dimensions, is more time consuming and requires a representative sample of stimuli. However, since multidimensional scaling is usually used where the dimensions are not known, it is not possible to specify in advance the stimuli which will be representative of all the possible stimuli, and so it is usually necessary to use a large number. The results can be difficult to interpret, and sometimes the inclusion of one very salient factor will lead to that factor exerting a major influence on the subject's judgements. Miller and Carterette (1975), for example,found a less clear picture of timbre perception in an experiement including variation of the fundamental frequency of tones, than in a second experiemnt where frequency was held constant.

Whilst studies of this type involving judgements on individual selected stimuli might yield some insight into the perception of timbre, there are other factors influencing the perceptual process which are generally not taken into account. These may arise from the physical characteristics of the signal and from psychological factors. In almost all of the studies, while the range of stimuli has covered a number of different types of instrument such as piano, violin etc., it has been customary to present one 'typical' tone from each instrument. However, it is not easy to specify what constitutes a 'typical' tone for a type of instrument. Tones produced on violins in general, or pianos in general, will vary greatly depending on a number of factors. First there is the actual instrument used in the production of the tone. If we are to believe what players say about the differences between different makes or models of instruments, then we should expect there to be differences between instruments in the acoustic patterns produced. Similarly the fundamental frequency and dynamic level will effect the spectral content and transient features of the tone. When previous workers have compared different types of instruments the tones have generally been of the same pitch and loudness, since otherwise these two factors tend to mask timbre differences. This means that some of the instruments will be played in uncharacteristic parts of their register and the findings may be far from typical. The player employed to produce the tones might have a great effect on the tones produced, since the skill of playing a musical instrument is in part related to producing tones of a consistent and certain quality, and variations in individual playing style will lead to differences in the tones.

There are other factors in the perception of musical tones which are unrelated to the physical characteristics of the acoustic signal. Subjects can be influenced by expectations about the nature of the stimuli, which will depend on other stimuli presented and the exact experimental instructions. In speech perception, Warren (1970) showed that when part of a continuous string of speech is replaced by an extraneous sound, the subject will hear the replaced speech sound although the acoustic signal presented may bear little resemblance to this. Experience with real instrument sounds will lead to expectations about how they should sound, and these expectations are likely to influence the perception of stimuli in a given setting. In a real acoustic environment the signal arriving at the ear bears a complex relationship to that emanating from the instrument body due to the differential directional propagation of the components of the complex wave and to the multiple reflections from surfaces which differentially absorb these components. These factors do not appear to influence listeners unduly and a trumpet sounds not only like a trumpet but like the same trumpet when the acoustic signal is changed by moving the instrument position relative to the listener. An explanation of this effect might lie in the results of a study of localisation by von Plenge (1972) who found that subjects could localise sounds accurately only after experience of the acoustics of the room, and that when the room acoustics were varied between individual signal presentations the subjects failed to learn appropriate localisation. It would appear likely that room acoustics are taken into account in the perception of timbre, and that a rapid learning process in a particular environment allows the listener to account for variations in the signal caused by the room. Transient properties of the sounds, which will be relatively invariant under such transformations, may be of special importance in this type of perception. Little attention has been paid to the fact that music is a sequence of tones and the timbre of a tone is likely to be different when played in conjunction with other tones. Grey (1978), for example, found that the subject's ability to discriminate real instrument tones from simplified synthesised tones depended upon whether the tones were presented individually or in the form of a tune. Another important factor which has often been overlooked is the degree of musical sophistication of the subjects, and failure to control for this may account for some of the discrepancies between published studies.

One major problem with the use of multidimensional scaling is that the dimensions revealed are dependent upon the particular stimulus set used. Hence is only one tone from each instrument type is included this will only reveal a set of dimensions for the perception of differences between instrument types. However, this need not be the same as the dimensions underlying the perception of differences between instruments of the same type, for example between different violins or between different players on the same instrument, and the results in the two cases may reveal very different solutions. One possible way in which subjects perform the sorts of tasks presented to them is to classify stimuli as similar to some stereotype stimulus. Hence if they are presented with a set of tones from different instruments, subjects will classify one as a piano, one as a trumpet etc., whereas if they are presented with a set of trumpet tones subjects will have to use a different classification scheme, for example into dull and bright tones. There is, however, not necessarily any reason to suppose that these two classification schemes will vary along the same dimensions unless we hypothesise that there really is a concept which can be called timbre and can be perceived only in a single way. However, much evidence suggests that timbre is not a unitary concept, but rather a catch-all way of describing differences between complex tones. It is therefore likely that judgements will differ between different sets of stimuli, between different listeners and between different perceptual tasks set to the same listener.

REFERENCES

G. Von BISMARCK 1974 Acustica 30, 146-159. Timbre of steady sounds: A factorial investigation of its verbal attributes.

J.M. GREY 1977 Journal of the Acoustical Society of America 61, 1270-1277. Multidimensional perceptual scaling of musical timbre.

J.M. GREY 1978 Journal of the Acoustical Society of America 64, 467-472. Timbre discrimination in musical patterns.

J.M. GREY and J.W. GORDON 1978 Journal of the Acoustical Society of America 63, 1493-1500. Perceptual effects of spectral modifications on musical timbres.

J.R. MILLER and E.C. CARTERETTE 1975 Journal of the Acoustical Society of America 58, 711-720. Perceptual space for musical structures.

G. Von PLENGE 1972 Acustica 26, 241-252. Uber das Problem der Im-Kopf-Lokalisation.

R. PLOMP 1970 In R. Plomp and G.F. Smoorenburg (Editors) Frequency Analysis and Periodicity Detection in Hearing, pp. 397-414. Timbre as a multidimensional attribute of complex tones.

R. PLOMP and H.J.M. STEENEKEN 1969 Journal of the Acoustical Society of America 46, 409-421. Effect of phase on the timbre of complex tones.

R.N. SHEPARD 1962 Psychometrika 27, 125-140. The analysis of proximities: Multidimensional scaling with an unknown distance function. I.

R.M. WARREN 1970 Science 167, 392-393. Perceptual restoration of missing speech sounds.

ACKNOWLEDGEMENTS The authors would like to thank the SRC/SSRC Joint Committee for financial support.