Psychophysical quantification of individual differences in timbre perception

INTRODUCTION

Timbre is a word used to refer to a collection of auditory attributes that have been approached with many different experimental methods. Some involve deciding a priori what a given attribute is and then proceeding to explore it with unidimensional psychophysical scaling techniques. For example, one might be interested in roughness or sharpness and proceed to evaluate the relative roughness or sharpness of various sounds and then try to link the subjective judgments to physical quantities derived from the sound signals. However this approach presumes on the one hand that listeners know what is meant by the word presented to them, can focus on that attribute and ignore others possessed by the sound, and that they all make the same link between the word and a specific aspect of their perception. This approach also presumes that psychoacousticians are clever enough to imagine what all the attributes might be ahead

between them (e.g., the *vibrone* is a hybrid between vibraphone and trombone). All pairs of sounds were presented to 84 listeners who judged their relative dissimilarity on a numerical scale from 1 (very similar) to 9 (very dissimilar). In reanalyzing the data from the 24 professional musicians among those subjects, the CLASCAL analysis revealed a three-dimensional space without specificities and two latent subject classes. Figure 1 presents this timbre space. Note that while the timbres are distributed in a relatively homogeneous manner along Dimensions 2 and 3, they form two large clusters along Dimension 1.

low

$$SC = \frac{1}{T}\int_0^T B(t)\,dt \quad \text{with} \quad B(t) = \left[\frac{\sum_{k=1}^{N} kA_k(t)}{\sum_{k=1}^{N} A_k(t)} \right] \quad \text{for a given analysis window}$$

$$SS = \sum_{k=1}^{N} \left| 20\log(A_k) - \frac{20\log(A_{k-1}) + 20\log(A_k) + 20\log(A_{k+1})}{3} \right|$$

$$SF = \frac{1}{M}\sum_{p=1}^{M} \left| r_{p,p-1} \right| \quad \text{with} \qquad — \quad \text{and} \quad !$$

where t_{max} is the instant in time at which the rms amplitude envelope attains its maximum, $t_{threshold}$ is the time at which the envelope exceeds a threshold value ($0.02*t_{max}$ in our case), T is the total duration of the sound, t is the begin time of the sliding short-term Fourier analysis window, A_k is the amplitude of partial k, N is the total number of partials, $r_{p,p-1}$ is the Pearson product-

important special case where the spline has maximal continuity equal to the order of the splines at each junction point, the number of parameters required for each dimension is the order plus the number of interior junction points. The number of degrees of freedom in this model is equal to the sum of the number of parameters per dimension across all dimensions. Note that this model is extremely parsimonious compared to classical MDS models since one can add a lot of stimuli and subjects without increasing the number of model parameters, provided that the number of dimensions remains the same and the transformation remains as smooth.

bpM5-238.9

We applied this approach to the group data for the 24 professional musicians comparing the timbre set presented in Figure 1. We tested for the parameters LAT and SCG for dimensions 1 and 2 and tried various physical parameters for dimension 3 (SS, SF, and maxamp). Using Monte Carlo tests, this model was then compared to the CLASCAL model with specificities and latent classes. The CONSCAL model was rejected in favor of the CLASCAL model in all cases. Given that the individual analyses showed differences in dimensionality and in the underlying physical nature of the dimensions across amM5-238.9 (w.2 (n) a) -0.2 (T2.4 (o) 50 0 0 50 0 0 Tm /TQ (a) -0.2 (n) - ( ) 55(

phy(i) -0.2 (ca)-0.2 (w)50 0 0 vaiabwes ohnpncepuawdmnsions

dimnsions in thraindividuaws. urthn,. thn

FIGURE 2

attached to each dimension different (as would be estimated for individual subjects by INDSCAL or classes of subjects by CLASCAL), but that the forms of the psychophysical functions are different. To illustrate this point, the functions for three subjects have been highlighted in the figure. Listener L1 (open triangles) has the lowest values for attack time and the function is nearly linear. L1 has the second highest function for spectral centroid also with a nearly linear function. Listener L2 (open squares) has fairly high

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. Journal of the Acoustical Society of America, 61, 1270-1277.

Grey, J. M., & Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. Journal of the Acoustical Society of America, 63, 1493-1500.

Iverson, P., & Krumhansl, C. L. (1993). Isolating the dynamic attributes of musical timbre. Journal of the Acoustical Society of America, 94, 2595-2603.

Krimphoff, J., McAdams, S., & Winsberg, S. (1994). Caractérisation du timbre des sons complexes. II: Analyses acoustiques et quantification psychophysique. Journal de Physique, 4(C5), 625-628.

Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielzén & O. Olsson (Eds.), Structure and Perception of Electroacoustic Sound and Music, (pp. 43-53). Amsterdam: Excerpta Medica.

McAdams, S., Susini, P., Krimphoff, J., Misdariis, N., & Smith, B. K. (in preparation). A meta-analysis of timbre space. II: Acoustic correlates of common perceptual dimensions. .

McAdams, S., & Winsberg. (in preparation). A meta-analysis of timbre space. I: Multidimensional scaling of group data with common dimensions, specificities and latent subject classes. .

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. Psychological Research, 58, 177-192.