Validation of a Multidimensional Distance Model for Perceptual Dissimilarities among Musical Timbres

Nicolas Misdariis[1], Bennett K. Smith[1], Daniel Pressnitzer[1], Patrick Susini[1] and Stephen McAdams[1,2]

[1]Institut de Recherche et de Coordination Acoustique/Musique (IRCAM),
1 place Igor Stravinsky, F-75004 Paris, France.
[2]Laboratoire de Psychologie Expérimentale (CNRS),
Université René Descartes, EPHE, 28 rue Serpente, F-75006 Paris, France.

Abstract: Several studies dealing with the perception of musical timbre have found significant correlations between acoustical parameters of sounds and their subjective dimensions. Using the conclusions of some of these studies, a calculation method of the perceptual distance between two sounds has been developed. Initially, four parameters are considered: spectral centroid, irregularity of the spectral envelope, attack time, and degree of variation of the spectral envelope over time. For each of these, a transformation factor between the physical axis and the corresponding subjective dimension is obtained by linear regression. After a normalization of the data, the final distance values between sounds is given by a linear combination weighted by the four transformation factors. Since this model is based on numerical results derived from experiments that mostly used synthesized sounds, the application to a database of recorded musical instrument sounds needs a strong validation procedure. This procedure involves the adjustment of the coefficients of the first four parameters as well as the eventual introduction of new ones to attain a perceptually relevant distance between two musical sounds.

THEORETICAL BASIS

Musical timbre is a complex auditory attribute for which multidimensional scaling techniques (MDS) have usually been employed to reveal the main underlying perceptual dimensions: from dissimilarity judgments on pairs of sounds, an MDS analysis builds a low-dimensional space where sounds are located with regards to their mutual timbral differences. In this context, the study of Krumhansl et al. [1,2] on a set of 21 synthesized sounds of musical instruments and the derived study of McAdams et al. [3] — with a more elaborated MDS model — on 18 stimuli of the same set are used as the basis for the present work. In both cases, the analysis leads to a timbre space with three common dimensions ; these will be referred to as ‘Space 1’ and ‘Space 2’, respectively.

Krimphoff et al. [4,5] attempted to quantify the perceptual axes of Space 1 in terms of acoustic parameters and the parameters derived were tested on Space 2. Both spaces gave similar results for the two of the dimensions, which fit with Spectral Centroid (average over the sound duration of the instantaneous spectral centroid within a running time window of 12 ms) and Log Attack-Time (rise time measured from the time the amplitude envelope reaches a threshold of 2% of the maximum amplitude to the time it attains this maximum amplitude). As for the third dimension, the most significant correlation occurs, for Space 1, with Spectral Irregularity (log of the spectral deviation of component amplitudes from a global spectral envelope derived from a running mean of the amplitudes of three adjacent harmonics) and, for Space 2, with Spectral Flux (average of the correlations between amplitude spectra in adjacent time windows) . The values of the correlation coefficients are shown below in Table 1.

TABLE 1. Correlation between perceptual dimensions and physical parameters (Krimphoff et al. [4,5])

	Spectral Centroid	Log Attack-Time	Spectral Irregularity	Spectral Flux
Space 1	+ 0.93	– 0.94	– 0.87	
Space 2	– 0.94	– 0.94		+ 0.54

CONSTRUCTION OF THE DISTANCE MODEL

The main objective of this study was to build a multidimensional distance model from the relations among the perceptual criteria obtained by MDS analysis and the calculated physical parameters. First, the subjective data, i.e. the coordinates of the sounds along each perceptual axis, were normalized to make the results of the two initial studies comparable: a multiplicative factor of 1.93 was applied to Space 1. Afterwards, a mathematical relation between these subjective data and the values of the corresponding parameter was obtained by a linear regression method: the slopes of the resulting six straight lines could then be considered as proportionality coefficients between the perceptual and the physical dimensions. Finally, these coefficients were used as weighting factors for the four acoustic correlates in a Euclidean distance formula representing the perceptual distance, as described in equation 1:

$$\mathrm{DIST} = \sqrt{3.5385 \cdot 10^{-5} \cdot \mathrm{SC}^{2} + 15.5236 \cdot \mathrm{LT}^{2} + 0.011881 \cdot \mathrm{SI}^{2} + 2728.7 \cdot \mathrm{SF}^{2}}\ , \qquad (1)$$

where SC, LT, SI, and SF are the differences in Spectral Centroid (Hz) ,Log Attack-Time ($\log_{10}$(s)), Spectral Irregularity (dB) and Spectral Flux (unitless), respectively, and DIST is the resulting distance between two sounds in this acoustic parameter space. Note that the coefficients for SC and LT are derived from the mean of the values from Space 1 and Space 2.

APPLICATION OF THE MODEL TO A REAL DATABASE

The global aim of this work is the development of a search engine within a database of recorded musical sounds. However, the initial studies from which the model is derived were based on a restricted set of sounds controlled in pitch, duration, and loudness. To the contrary, the database upon which the model is expected to make predictions covers at present eleven instruments recorded over their entire pitch range, at all dynamics and using almost all playing techniques of each instrument (normal, flutter tonguing, staccato, multiphonics, etc.). Therefore, a validation step is necessary both in the definition of the perceptual dimensions and in the calculation method of the parameters.

For the time being, we have focused on the problems of parameter extraction when processing acoustically produced sounds rather than synthesized ones, such as a nonnegligible signal-to-noise ratio, natural or forced inharmonicity, high non-stationarity, etc.. To illustrate a typical problem often encountered, Figure 1 presents the Spectral Centroid as a function of the physical level (RMS) over the duration of a Bb0, staccato, trombone sound. The existence of areas where the Spectral Centroid information is quite false seems obvious: taking with reference the frame where RMS level is the highest (approximately between 0.4 and 0.6 seconds) and the theoritical pitch of a Bb0 in the tempered scale (55 Hz), we can notice an over-estimation of the parameter elsewhere.

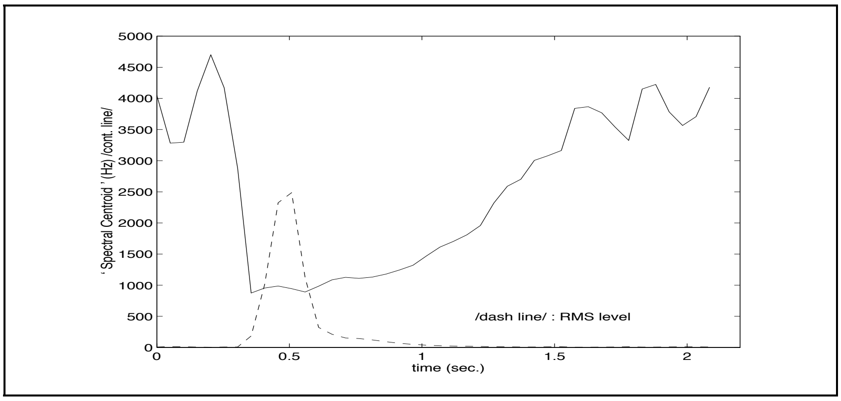

FIGURE 1. Spectral Centroid vs RMS level of Bb0, staccato trombone sound.

The adjustments suggested by this kind of investigation have been made directly within the parameter calculation procedures: different ways to calculate a given parameter have been tested and elements like noise-gate modules in the generation of spectral data have even been implemented. Since the modified model in its present state gives encouraging results, the next step of development will be to quantify empirically different perceptual scales, such as roughness, in order to establish more accurately the relation between the distances in perceptual and physical parameter spaces.

REFERENCES

1. Krumhansl C. L., in *Structure and Perception of Electroacoustic Sound and Music*, S. Nielzén & O. Olsson (Eds.), Elsevier, Amsterdam, 1989, pp 43-53.
2. Krumhansl C. L., Wessel D. L. & Winsberg S., unpublished data, IRCAM, Paris (reported in [1]).
3. McAdams S., Winsberg S., Donnadieu S., De Soete G. & Krimphoff J., *Psychological Research* **58**, 177-192 (1995).
4. Krimphoff J., McAdams S. & Winsberg S., *Journal de Physique* **4**(C5), 625-628 (1994).
5. Krimphoff J., *DEA thesis (unpublished)*, Université du Maine, Le Mans, France (1993).