The effect of fundamental frequency on the brightness dimension of timbre

Jeremy Marozeau[a)]
Institute for Hearing, Speech & Language and Department of Speech-Language Pathology & Audiology (106A FR), Northeastern University, Boston, Massachusetts 02115

Alain de Cheveigné
Equipe Audition, FRE 2929 LPP, CNRS-Université Paris5, Ecole Normale Supérieure, 29 rue d'Ulm 75230, Paris, France

(Received 10 May 2006; revised 2 October 2006; accepted 9 October 2006)

The dependency of the brightness dimension of timbre on fundamental frequency (*F0*) was examined experimentally. Subjects compared the timbres of 24 synthetic stimuli, produced by the combination of six values of spectral centroid to obtain different values of expected brightness, and four *F0*'s, ranging over 18 semitones. Subjects were instructed to ignore pitch differences. Dissimilarity scores were analyzed by both ANOVA and multidimensional scaling (MDS). Results show that timbres can be compared between stimuli with different *F0*'s over the range tested, and that differences in *F0* affect timbre dissimilarity in two ways. First, dissimilarity scores reveal a term proportional to *F0* difference that shows up in the MDS solution as a dimension correlated with *F0* and orthogonal to other timbre dimensions. Second, *F0* affects systematically the timbre dimension (brightness) correlated with spectral centroid. Interestingly, both terms covaried with differences in *F0* rather than chroma or consonance. The first term probably corresponds to pitch. The second can be eliminated if the formula for spectral centroid is modified by introducing a corrective factor dependent on *F0*. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2384910]

PACS number(s): 43.66.Jh, 43.66.Hg, 43.66.Lj, 43.75.Cd, 43.75.Bc [DD] Pages: 1–XXXX

I. INTRODUCTION

In a previous study (Marozeau *et al.*, 2003) we looked at the change with fundamental frequency (*F0*) of the timbre of sounds produced by a set of 12 musical instruments. The timbre of most instruments was found to be stable as a function of the note played. Nevertheless, small *F0* dependencies were observed that appeared to be specific to particular instruments. The choice of natural instrument sounds rather than synthetic stimuli in that study ensured musical relevance, but the large interinstrument variations of timbre within the stimulus set made the investigation of small *F0*-dependent effects difficult. Because distance has a quadratic dependency on the difference along each dimension, large interinstrument differences along one dimension can "swamp" smaller differences along other dimensions.

This study used a set of synthetic instrumental sounds designed to differ along the single physical dimension of spectral centroid (in addition to *F0*). Variations along other dimensions known to affect timbre (temporal envelope shape and spectral spread) were minimized so as to maximize sensitivity along the dimension of interest. The perceptual correlate of spectral centroid is termed "brightness." The centroid is defined from the spectral envelope that determines amplitudes of all partials, and thus does not depend directly on *F0*, but there are several reasons to expect a dependency of the perceptual correlate, brightness, on *F0*. One is that *F0* determines pitch, which is known to have a complex multidimensional nature, involving a cyclic chroma dimension related to position within the octave (*F0* modulo a ratio power of 2), a "tone height" dimension related to *F0*, and possibly a "spectral pitch" dimension determined by the overall spectral distribution (Shepard, 1999). The latter can certainly be expected to interact with brightness. Another reason is that *F0* dependencies of *vowel timbre* have been documented in several studies. Specifically, it appears that in order to maintain constant vowel identity over a one-octave increase of *F0*, formant frequencies must be shifted upward by about 10% (Slawson, 1968; Nearey, 1989). One might think that such dependencies are specific to speech sounds, but Slawson (1968) found similar effects when subjects were instructed to treat the stimuli as musical sounds rather than vowels.

The standard methodology for timbre studies is to apply multidimensional scaling (MDS) analysis to dissimilarity matrices obtained by asking subjects to rate the dissimilarity between pairs of stimuli. Our previous study (Marozeau *et al.*, 2003) extended it to allow timbre comparisons between sounds that differed in *F0*. Subjects were instructed to ignore the resulting difference in pitch. They were quite successful in doing so, but there was nevertheless some evidence of an effect of *F0* difference—or the pitch difference that it induces—on timbre. The present study aimed at understanding the nature of this interaction.

Supposing that *F0*-induced *pitch* differences affect timbre, one can speculate on the form of the dependency. Do timbre differences vary monotonically with the *frequency* difference between notes (on a linear or log scale) [Fig.

[a)]Electronic mail: marozeau@neu.edu

PROOF COPY 021701JAS

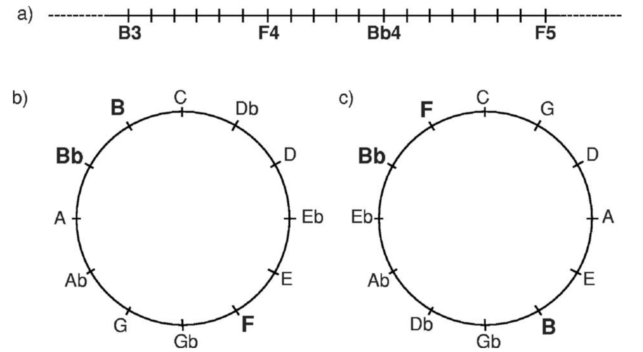

FIG. 1. Hypothetical geometrical structures to predict dissimilarity between sounds that differ in *F0*: (a) logarithmic frequency axis, (b) chroma circle, and (c) circle of fifths.

1(a)]? Are they instead a function of the difference in chroma? If so, they should vary as distance along a chroma circle [Fig. 1(b)]. Because the stimuli are presented pairwise they might also be function of the *consonance* between notes. If so, they should vary as distance along a circle of fifths [Fig. 1(c)] (Shepard, 1982). Stimulus frequencies in this study were chosen so as to test each of these hypotheses.

II. METHOD

A. Stimuli

Stimuli were produced according to a simple model of additive synthesis. Each had a spectrotemporal envelope shaped as the outer product of a frequency-independent temporal envelope multiplied by a time-independent spectral envelope function. The temporal envelope, common to all stimuli, comprised a linear 50-ms onset (attack) followed by a 22.5-ms linear decay of 20% of the maximum amplitude, a 382.5-ms constant amplitude part (sustain), and a linear offset (release) of 45 ms (so-called ADSR or attack decay sustain release profile). The spectral envelope was shaped as a Gaussian when expressed as "partial loudness" (intensity per critical band raised to the power 0.3) as a function of frequency on an equivalent rectangular bandwidth (ERB) (Moore, 2003) scale. The width of the Gaussian was 5 ERB and the same for all stimuli. The centroid of the Gaussian envelope took on six values equally spaced on an ERB-rate scale from 17 to 22 ERB-rate (1196, 1358, 1539, 1739, 1963, and 2212 Hz, respectively). Figure 2 illustrates the waveform and spectra of two stimuli with the same *F0* (247 Hz) but different centroids (17 and 22 ERB-rate).

For each of these envelopes, stimuli were produced at four *F0*'s: 247, 349, 466, and 698 Hz (notes *B3*, *F4*, *Bb4*, and *F5*), for a total of 24 stimuli. This *F0* set was designed to produce intervals both small and large in terms of frequency, chroma, and consonance.

Stimuli were sampled at 44.1 kHz with a resolution of 16 bits. They were presented diotically over earphones at approximately 75 dBA. The term "instrument" will be used in the following to designate the set of stimuli with the same centroid.

B. Listeners

Fourteen subjects (including seven women and seven musicians) participated in the experiment. Musicians were defined as having played an instrument for at least 3 years.

C. Procedure

The subjects were asked to rate the dissimilarity of the 276 possible pairs of the 24 stimuli. They were instructed to base their judgments only on the timbre of each stimulus and to ignore the pitch difference. Previous experiments showed

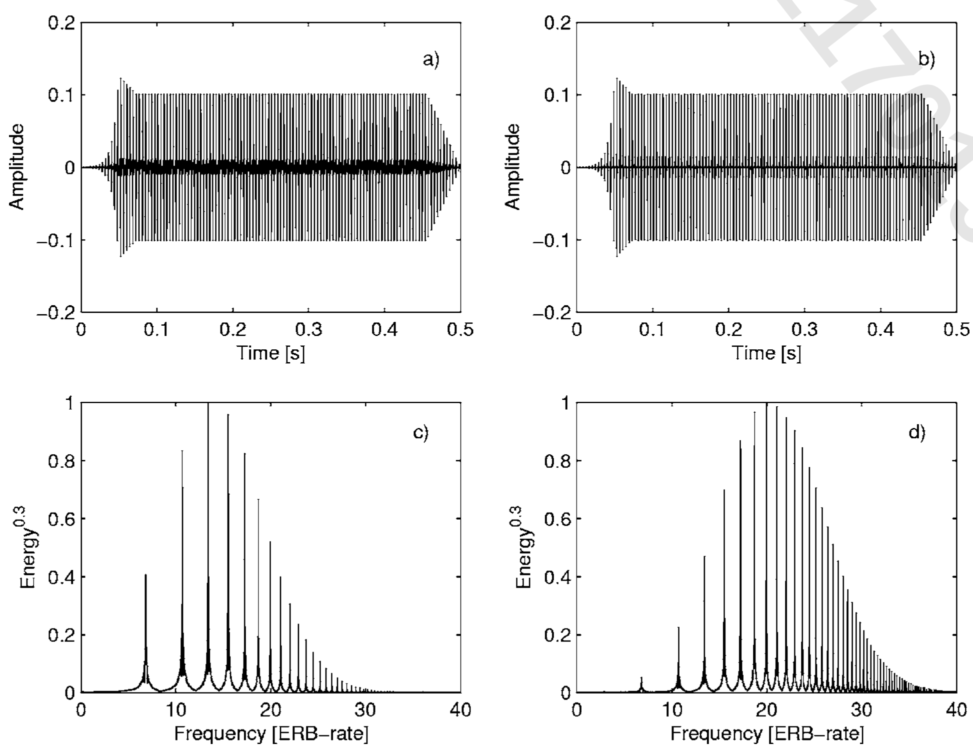

FIG. 2. Waveforms (top) and spectra (bottom) of two stimuli used in our experiments. Both had the same F0 (247 Hz) but different spectral centroids: 17 ERB rate [(a) and (c)] and 22 ERB rate [(b) and (d)].

PROOF COPY 021701JAS

that this task was possible, at least for *F0* differences below one octave. The order within pairs and the order of pairs were presented were random (a different randomization was used for each session and subject). The experiment was run inside an audiometric booth, and stimuli were presented diotically over Sennheiser 520 II headphones. Further procedural details can be found in Marozeau *et al.* (2003).

III. RESULTS

A. Outliers, effect of musical experience

Correlation coefficients between dissimilarity scores were calculated for all pairs of subjects. These scores were submitted to a hierarchical cluster analysis to identify eventual outlier subjects. No outlier was found. An ANOVA was performed with between-subjects factor musical experience (2) and within-subjects factor instrument pair (276), to reveal an eventual effect of musical experience. No effect of musical experience was found, either as a main effect or as an interaction. Results were therefore averaged across all subjects.

B. MDS analysis

The data were first transformed with a hyperbolic arctangent function to attempt to correct for the effect of the bounded response scale (Schonemann, 1983; Marozeau, 2004). MDS produces a spatial configuration such that distances between points fit observed dissimilarities between instruments. Distances being unbounded whereas dissimilarities are bounded, the solution is necessarily distorted. The transformation reduces this distortion. The hyperbolic tangent has a slope that is close to 1 for arguments between 0 and 0.5, and that increases exponentially to infinity as its argument approaches 1. It thus leaves unchanged the dissimilarities smaller than 0.5 and expands dissimilarities close to 1.

Transformed scores were analyzed using the MDSCAL procedure, implemented according to the SMACOFF algorithm (Borg and Groenen, 1997). A two-dimensional solution was selected because higher-dimensional solutions did not decrease significantly the stress of the model (Borg and Groenen, 1997). As the MDSCAL solution is rotationally undetermined, the solution was rotated with a procrustean procedure in order to maximize the correlation between spectral centroid and position along the first MDS dimension for stimuli with a 247 Hz (*B3*) *F0*. Figure 3 shows this solution.

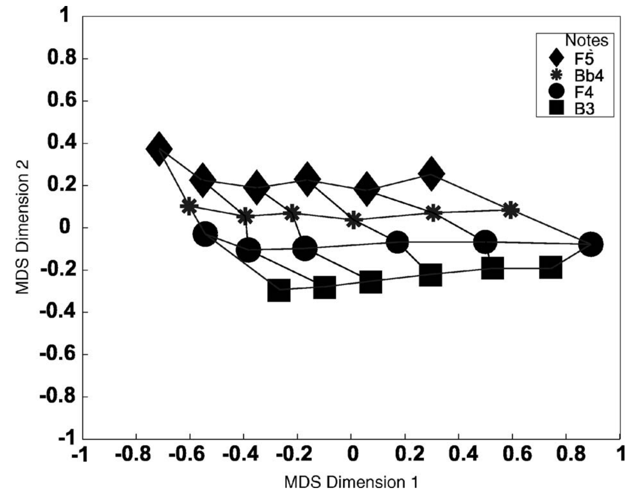

FIG. 3. Two-dimension multidimensional scaling (MDS) solution. The solution was rotated to maximize correlation between dimension 1 and spectral centroid.

Stimuli at *B3* are represented by squares, those at *F4* by circles, those at *Bb4* by stars, and those at *F5* by diamonds. Each stimulus is connected by a segment to the two stimuli with the closest spectral centroid and the same *F0* and to the two stimuli with the closest *F0* and the same spectral centroid. Roughly speaking, stimuli are distributed along the first dimension in order of their spectral centroid, and along the second dimension in order of *F0*. Stimuli with a given *F0* tend to follow a horizontal line, indicating that the dimension related to *F0* is not affected by spectral centroid. In contrast, stimuli with a given centroid follow a line slanted to the left, indicating that the dimension related to the centroid is affected by *F0*. In other words, an *increase* in *F0* has an effect similar to a *decrease* in the centroid.

C. Horizontal shift

To test if this shift is significant, an ANOVA was performed on a restricted set of the data. To better understand this analysis let us consider only two instruments (*X* and *Y*) differing along a timbre dimension. The positions of these two instruments along this dimension are represented at two different *F0*'s by X_1, Y_1 and X_2, Y_2, respectively, in Fig. 4. Two cases may be considered. In the first, represented in the left panel, the position of the instruments remains invariant

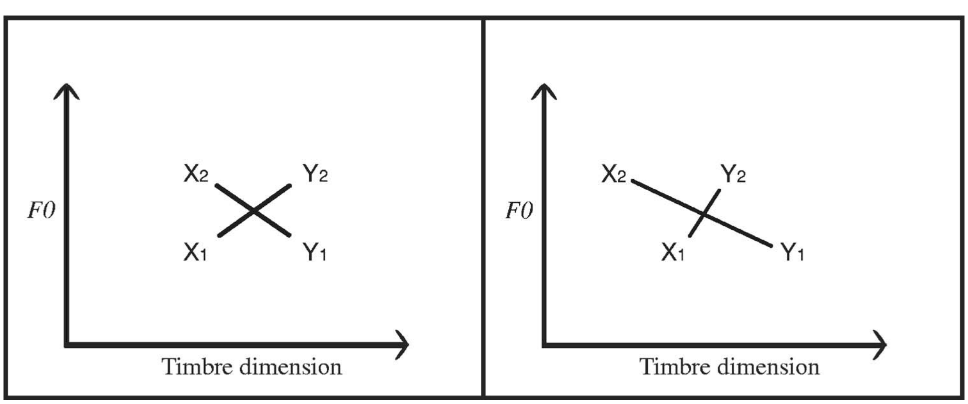

FIG. 4. Hypothetical geometrical structures describing the dependency on *F0* of the timbres of two instruments. Left panel represents the case where the timbres do not change (along the horizontal dimension) with *F0*. Right panel represents the case where the timbres do change with *F0*.

TABLE I. Amount of variance (R^2) of timbre dissimilarity accounted for by each factor manipulated in the experiment. Only the values of effects significant at the $p=0.005$ level are included, as determined by an ANOVA. Factors are instrument pair (IP, df=15) and *F0* order (*F0*, df=2).

	B3			F4			Bb4		
	IP	*F0*	IP×*F0*	IP	*F0*	IP×*F0*	IP	*F0*	IP×*F0*
F4	42.87	8.53	5.11						
Bb4	28.49	9.5	n.s.	43.41	11.39	3.92			
F5	8.31	12.47	n.s.	21.17	27.45	3.85	29.18	17.45	n.s.

along the timbre dimension with a change of *F0*. The distance $\overline{X_1Y_2}$ is then the same as the distance $\overline{X_2Y_1}$. In the second case, represented in the right panel, the position of the instruments moves uniformly along the timbre dimension as function of *F0*; the distances $\overline{X_1Y_2}$ and $\overline{X_2Y_1}$ are in this case different.

To test if an instrument shifts along the first timbre dimension with respect to another instrument, an ANOVA was performed with the dissimilarities of all the pairs containing both instruments at different *F0*'s. The two main factors were the instrument pair (X_1Y_3 vs. X_2Y_4,) and the *F0* order inside the pair (X_1Y_3 vs. X_3Y_1). A lack of change with *F0* along the timbre dimension would imply a lack of significant effect for the latter factor. Conversely, a significant effect would imply a timbre change with *F0*.

Table I shows the result of this analysis for each instrument pair.

The results show that the factor of the *F0* order is always significant. Therefore the shift can be considered as significant for every instrument.

IV. DISCUSSION

A. The dependency of timbre on spectral centroid

Stimuli fall along the first dimension of the MDS solution in order of their spectral centroid, consistent with the pattern found for natural instruments (Marozeau *et al.* 2003). This is not surprising given that they were designed to vary according to that dimension. Their approximately equal spacing along the psychological dimension agrees with their equal spacing along the physical dimension. It is, however, unlikely that our methods could reveal a discrepancy in this respect, should it exist, because of experimental noise. The leftward shift with increasing *F0* suggests that this perceptual dimension is also slightly dependent on *F0*.

B. The effect of *F0*

Stimuli fall along the second dimension of the MDS solution in order of their *F0*. Evidence for a similar *F0*-dependent dimension was found for natural instruments by Marozeau *et al.* (2003) for a smaller range of *F0*'s. Scatter along this dimension is about 40% of the scatter along the first dimension, indicating that the effect of *F0* is moderate compared to the effect of the centroid. The question raised in the Introduction, concerning the dependency on *F0 difference*, chroma similarity, or consonance, is answered unambiguously: the component of dissimilarity induced by an *F0* difference is not a function of the difference in chroma, or the consonance between notes, but rather the size of the difference in *F0*.

A similar remark holds for the small variations observed along dimension 1: these appear to covary with a linear scale of *F0*, rather than a circular scale of chroma or a circular scale of consonance (Fig. 1).

C. An improved "spectral centroid" descriptor

The stimuli were created according to a definition of the spectral centroid as described by Marozeau *et al.* (2003). If this descriptor were accurate to predict the first dimension of timbre (brightness) independently of *F0*, we should have observed no shift along that dimension with *F0*. The significant effect that we did observe can be interpreted either as implying that perceptual dimension *does* depend slightly on *F0*, or that we should search for a better descriptor that ensures that it does not. There is no definite way to choose between these rival interpretations, but for practical applications it would be nice to have a signal-based descriptor of this dimension that is *not* sensitive to *F0*. The purpose of this paragraph is to present such an improved descriptor.

The first stages are the same as in Marozeau *et al.* (2003). Briefly, the waveform was first filtered to model the sensitivity of the outer and middle ear (Killion, 1978). Then it was filtered by a gammatone filterbank (Patterson *et al.*, 1992) with channels spaced at half-ERB intervals on an ERB-rate scale (z) between 25 Hz and 19 kHz (Hartmann, 1998). Instantaneous power was calculated within each channel and smoothed by delaying it by $1/4fc$ (where fc is the characteristic frequency of the channel), adding it to the undelayed power, and convolving the sum with an 8-ms window. Smoothed power was then raised to the power 0.3 to obtain a rough measure of "partial loudness" for each channel. The partial loudness-weighted average of ERB-rate was taken over channels, the result being an "instantaneous spectral centroid" function of time according to

$$\bar{Z}(t) = \sum_z z\psi_z(t) \Big/ \sum_z \psi_z(t), \qquad (1)$$

where $\psi(t)$ is the "partial loudness" of the channel z at instant t. Finally, the instantaneous centroid $\bar{Z}(t)$ was weighted by "instantaneous loudness" (sum over channels of partial loudness) and averaged over time to obtain a single descriptor value, $\bar{Z}$, to characterize the entire signal.

To better predict position along the first timbre dimension, a correction of the spectral centroid is now proposed. First, the descriptor is converted from ERB-rate to Hz according to the formula:

$$\bar{f} = (\exp(\bar{Z}/9.26) - 1)/0.004\,37, \qquad (2)$$

where $\bar{f}$ is the value of the spectral centroid in Hz (Hartmann, 1998). Then the value of the *F0* of the stimulus is subtracted from $\bar{f}$:

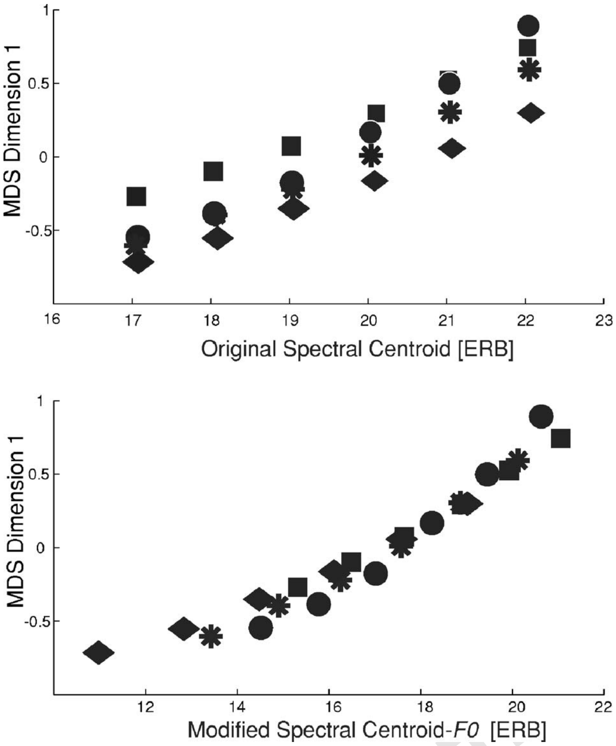

FIG. 5. Scatter plots showing the dependency of the first MDS dimension on signal-based predictors of brilliance. Upper panel: uncorrected spectral centroid. Lower panel: corrected spectral centroid.

$$\overline{f_{\text{corrected}}} = \bar{f} - F0, \tag{3}$$

where $\overline{f_{\text{corrected}}}$ corresponds to the corrected value of $\bar{f}$. Finally the value is converted back in ERB-rate:

$$\overline{Z_{\text{corrected}}} = 9.26 \ln(0.004\,37\overline{f_{\text{corrected}}} + 1), \tag{4b}$$

where $\overline{Z_{\text{corrected}}}$ corresponds to the corrected value of $\bar{Z}$ in ERB-rate. The upper panel of Fig. 5 shows a scatter plot of $\bar{Z}$ with the projection along the first timbre dimension. The lower panel shows the scatter plot of $\overline{Z_{\text{corrected}}}$ with the projection along the first dimension. The stimuli are better aligned along the regression line. This can be quantified by the coefficient of correlation of 0.97 (0.91 before correction), explaining more than 93% of the variance (82% before correction) ($dl=22$; $p<0.001$).

A similar correction procedure has already been invoked in studies of vowel perception (Traunmuller, 1981; Hoemeke and Diehl, 1994). Specifically, it has been proposed to "correct" the values of vowel formants, in particular the first formant, by subtraction of the value of *F0* from that of the formant frequency.

V. CONCLUSIONS

We found that cross-*F0* comparisons of timbre were possible up to at least 18 semitones' *F0* difference, corresponding to one and a half octaves. However, dissimilarities appeared to contain a term that increased proportionally with the difference in *F0*. This showed up in MDS solutions as a dimension correlated with *F0* and orthogonal to that correlated with the spectral centroid. Along this dimension, the stimuli were ordered according to their *F0* and not according to their chroma or consonance. The MDS solution also showed a dependency on *F0* of the dimension that covaries with spectral centroid (brightness). That dependency can be reduced by modifying the definition of the spectral centroid. Although this correction seems to be well adapted to the data of this experiment, it needs to be tested and validated over a wider range of stimuli and a wider range of *F0*. If this correction is confirmed, it could be useful in applications that use timbre descriptors to discriminate, categorize, or generate instrumental sounds.

ACKNOWLEDGMENTS

This work was performed in partial fulfillment of the Ph.D. obligations of the first author. The project was funded in part by the Swiss National Science Foundation and by the European Union Project CUIDADO. It was conducted within the Music Perception and Cognition team at IRCAM. The authors wish to thank Diana Deutsch and two anonymous reviewers for their helpful comments.

Borg, I., and Groenen, P. J. F., (**1997**). *Modern Multidimensional Scaling: Theory and Applications* (Springer, New York).

Hartmann, W. M., (**1998**). *Signals, Sound, and Sensation* (Springer, New York).

Hoemeke, K. A., and Diehl, R. L. (**1994**). "Perception of vowel height: the role of *F1-F0* distance," J. Acoust. Soc. Am. **96**, 661–674.

Killion, M. C., (**1978**). "Revised estimate of minimum audible pressure: Where is the missing 6 dB?," J. Acoust. Soc. Am. **63**, 1501–1508.

Marozeau, J., (**2004**). "L'Effet de la Fréquence Fondamentale sur le Timbre," Ph.D dissertation. University Pierre et Marie Curie, Paris VI (Lulu.com, Paris).

Marozeau, J., de Cheveigné, A., McAdams, S., and Winsberg, S., (**2003**). "The dependency of timbre on fundamental frequency," J. Acoust. Soc. Am. **114**, 2946–2957.

Moore, B. C. J., (**2003**). *An Introduction to the Psychology of Hearing* (Academic, San Diego, CA).

Nearey, T. M., (**1989**). "Static, dynamic, and relational properties in vowel perception," J. Acoust. Soc. Am. **85**, 2088–2113.

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M., (**1992**). "Complex sounds and auditory images," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and L. Horner (Pergamon, Oxford).

Schonemann, P. H., (**1983**). "Some theory and results for metrics of bounded response scales," J. Math. Psychol. **27**, 311–324.

Shepard, R. N., (**1982**). "Geometrical approximations to the structure of musical pitch," Psychol. Rev. **89**, 305–333.

Shepard, R. N., (**1999**). "Structural representation of musical pitch," in *The Psychology of Music*, edited by D. Deutsch (Academic, New York).

Slawson, A. W., (**1968**). "Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency," J. Acoust. Soc. Am. **43**, 87–101.

Traunmuller, H., (**1981**). "Perceptual dimension of openness in vowels," J. Acoust. Soc. Am. **69**, 1465–1475.