ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 92 (2006) 820 – 825

Does Timbral Brightness Scale with Frequency and Spectral Centroid?

Emery Schubert
School of Music and Music Education, University of New South Wales, Sydney, Sydney, 2052, Australia. E.Schubert@unsw.edu.au

Joe Wolfe
School of Physics, University of New South Wales, Sydney, Sydney, 2052, Australia. J.Wolfe@unsw.edu.au

Summary
Two models that predict perceived timbral brightness in terms of the centroid of the frequency spectrum were investigated. One model simply uses the centroid of the frequency spectrum, the other divides this same value by the fundamental frequency: the latter scales the centroid of the frequency spectrum with the fundamental frequency. Different single tone and pitch combinations, presented sequentially, were compared. Participants were not asked to ignore pitch differences and intervals of greater than an octave were compared. The results indicate that brightness is much better correlated with frequency spectrum centroid ($r = 0.513$, $p < 0.01$) than with the ratio of the centroid of the frequency spectrum to the fundamental frequency ($r = 0.030$, $p = 0.441$).

PACS no. 43.66.Jh, 43.66.Hg, 43.66.Ki, 43.66.Lj

1. Introduction

In a steady tone, the timbre depends, among other things, upon the power spectrum (the distribution of power as a function of frequency) [1, 2, 3]. Many researchers believe that the timbral quality of brightness correlates with increased power at high frequencies. For example, the vowel sound 'ee' sounds brighter than 'oo' (see also [4]). One simple quantification of the distribution in the power spectrum is the spectral centroid, f_c. For a power spectrum with components $P_i(f_i)$, f_c is defined as $\sum f_i P_i / \sum P_i$, where f_c is a frequency. It is sometimes described by analogy with the centre of mass: f_c represents the power distribution over frequency in the way that centre of mass represents the mass distribution over position.

What happens to the brightness of a musical note when the fundamental frequency (F_0) is shifted? In most cases, a higher pitched note has a somewhat higher f_c than does a lower pitch played on the same instrument[1]. Does the higher f_c of a high note produce greater perceived brightness? Or does human audition make some compensation for the F_0 shift?

Some research [5, 6] suggests that brightness might be better correlated with what Kendall and Carterette call the 'unitless centroid' [5], the ratio of the centroid to the fundamental frequency (i.e., f_c/F_0 hereafter the F_0AC, for F_0 Adjusted Centroid).

However, the use of the F_0AC to quantify brightness would lead to some paradoxical results. For instance, consider what happens when a recording of a musical instrument or a voice is played back at lower speed. In this case, both the F_0 and the spectral envelope are scaled by the same factor, so the F_0AC is unchanged. (This situation is sketched in Figure 1.) It seems intuitive that the half-speed playback will sound less bright, even though its F_0 adjusted centroid remains unchanged. Yet the F_0AC model predicts the same brightness for the original and the slowed replay.

In a recent study by Marozeau *et al* [7], which investigated the similarity between different sounds at different F_0, participants were asked to try to ignore the pitch difference while rating the perceptual dissimilarity of tones with different F_0 but matching spectral structure and loudness. This study found that differences in timbre depended little on pitch when the pitch difference was either 2 semitones or 11 semitones. In their Experiment II, they found that "*timbre* itself was stable for some instruments (eight for 2 semitones, four for 11 semitones, out of nine instruments)" (p. 2953, c.1, italics in original). Marozeau *et al.* point out that the concept of timbre includes the aspects of identity

Received 22 April 2005, revised 15 June 2006,
accepted 23 June 2006.

* Portions of this work were presented in: E. Schubert, J., Wolfe, A. Tarnopolsky: Problems in the relationship between perception of brightness and spectral centroid in complex, multivoiced timbres. Proceedings of the 8th International Conference of Music Perception and Cognition. Evanston, IL, August 3-7, 2004.

[1] A tone with a low F_0 can have a higher spectral centroid than a second tone with a higher F_0 if, for instance, the low tone contains strong higher partials and the high tone contains strong lower partials. For example, a sine wave at the note A4 has a lower spectral centroid (440 Hz) than does a trumpet playing *forte* in the octave below A4.

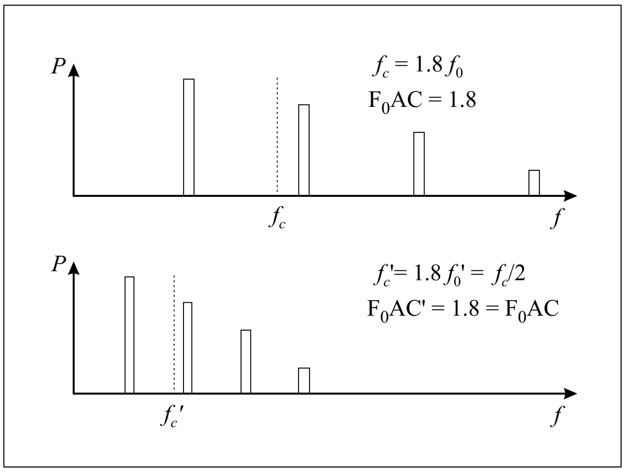

Figure 1. An illustration of the effect on spectral parameters of replaying a recording (original spectrum at top) at half speed (spectrum at bottom). The spectral centroid f_c is halved, but the F_0AC is unchanged.

(which instrument produces the sound) and quality (how the sound differs from another of the same loudness and pitch). In the cases when no difference (no dissimilarity) was reported for different-F_0—same-instrument combinations, it may have been that the participants were observing that the identity was unchanged, and not noticing (or not reporting) that the quality was different. For example, although the spectra of notes played on the bassoon vary greatly across its pitch range, a musician would be very likely to identify it as a bassoon in all cases.

Marozeau *et al.* define a quantity closely related to f_c (which gave results strongly correlated with those obtained using f_c. They found that this quantity correlates with one of the perceptual dimensions they report). Marozeau *et al.* do not specifically examine F_0 adjustment as a model for brightness perception, the issue which concerns us in the present study. Further, it is conceivable that asking participants to ignore pitch difference when listening to familiar acoustic instruments may encourage a top-down (identity) strategy which competes with the bottom-up (tone-quality assessment) strategy.

We decided to investigate further the relationship between f_c and perceived brightness by examining brightness responses (1) without specifically requiring the participant to ignore pitch differences, and (2) when tones were greater than an octave apart. Both of these variations, we argue, provide realistic listening experiences. It is unusual to ask a musician comparing two tone colours to try consciously to disregard pitch. Also, it is common to find music of Western traditions that span well over one octave, and in piano and orchestral ensembles spans of two to five octaves are fairly common. A limitation of recent research on timbre is that "almost all timbre studies hover around A4, Eb4, Bb4 or C4, limiting the generalizability of results" [6, p. 597]. We are interested in determining which of two models is a better predictor of perceived timbral brightness when no instruction is provided regarding the ignoring of pitch differences.

2. Hypotheses

Two hypotheses for the brightness of tones were tested.

- The *frequency centroid* (FC) hypothesis is that the perceived timbral brightness of a tone is simply correlated with f_c.
- The F_0-*adjusted-centroid* (F_0AC) hypothesis is that perceived timbral brightness correlates with f_c/F_0.

3. Design

Our design is different from that of Marozeau *et al.* in several respects. First, we did not attempt to produce stimuli with equal loudness. Instead, we produced each stimulus with two loudness levels – loud and soft. This approach meant that we did not need to produce equal loudness stimuli because if tone A had a different brightness from tone B, it would be the case for both loudness conditions. If not, one of the loudness levels might confound the response, and the result would be inconclusive. This would be clearly noticeable in a scatter plot. Therefore, having two loudness levels for each stimuli compensates for any differences incidentally produced by changes in loudness rather than changes in experimental conditions. We believe that, despite making the design more complex, this is a more robust approach than selecting stimuli of equal loudness—which is more susceptible to undesired within-subject variation. Second, we chose four test pitches that covered a wide F_0 range, with three high F_0s being at least two octaves above a low F_0 tone. The pitches used were E2, E4, A#5 and E5. The A#4 was chosen because of the multidimensional nature of pitch [8]: it ensured that some variation in chroma is made, rather than octave variation alone. This could in principle allow us to see if there were any effect on brightness response due to harmonic relations among the notes. For example, the harmonics of E4 and E5 are subsets of those of E2 whereas A#4 has few harmonics that fall very close to those of E2. While the hypotheses predict no difference in response due to lack of harmonic agreement, including this pitch ensures that the results could be generalized beyond simple octave relationships.

Before commencing the main experiment, we conducted a preliminary study (Experiment 1) with a panel of listeners to determine two tone-colours that were perceived as having significantly different timbral brightnesses. Having two stimuli with notably different brightness ratings would allow those two stimuli to be used to manipulate *a priori* the brightness levels for Experiment 2.

4. Experiment 1

The aim of Experiment 1 was to select two tones that were significantly different in perceived timbral brightness ratings. Our choice of stimuli was guided by the desire to allow easy reproduction of the experiment by other investigators at minimal cost. The internal, computer generated MIDI instrument sounds used in Finale

2004 for Macintosh OS 10.2.6 were used. A free version of Finale, called finale notepad, can be downloaded from http://www.finalemusic.com/. Finale uses a default internal playback setting, referred to as CoreMidi which is standard in the Macintosh OS X computer operating system.

Six participants with significant musical training (minimum of a Music degree completed) and self-reported normal hearing rated the brightness of 12 MIDI generated orchestral tones (which appear on the Finale 2004 large orchestral score template file) played for 2.6 s each. The participants ranged in age from 30 to 44 with a mean age of 39. The timbres were generated at E2 and the participants rated them one by one. The stimuli were then generated again at the pitch A#4 and again rated for their brightness. Participants were asked to rate each note on a scale of 1 (very lacking in brightness) to 10 (very bright timbre) twice. They performed the task once for familiarisation. Only the second ratings were analysed. All E2 comparisons were made in one session. All A#4 comparisons were made in a second session. That is, pitch was held constant across comparisons. Participants were reminded that timbre was multidimensional and that only one dimension, the brightness, was being rated. The sounds that produced the greatest mean difference in subjective brightness ratings for both E2 and A#4 were selected to generate the stimuli for Experiment 2.

The sounds used were ‘piccolo’ (which was rated by the panel as the *least* bright sound *at a given pitch* [coded in this paper as ‘−’]) and trumpet (for the bright sound [+]). Because this may seem surprising, it is important to note that the sampled and treated sound called ‘piccolo’ and that called ‘trumpet’ have the same musical range, which exceeds that of the acoustic instruments with these names. The choice of ‘piccolo’ and ‘trumpet’ was not made by the investigators. Rather, the subjects in this experiment chose the tone labeled as ‘piccolo’ as the least bright and that labelled as ‘trumpet’ as most bright, based in part upon their perception of the notes E2 and A#4, which are not in the piccolo range and only one of which is in the trumpet range.

5. Experiment 2

The aim of Experiment 2 was to examine individual tones with *a priori* differences in brightness and pitch. The data were also used to extend previous findings about the relationship between perceived brightness at different centroid and fundamental frequencies.

5.1. Participants

16 participants (10 males, 6 females, average age 38 years, youngest 23 years, oldest 52 years) with normal hearing participated in the experiment. Six participants reported less than 2 years of musical training. The other participants had significant musical experience, being enrolled in an undergraduate music degree, and/or having considerable experience as performing musicians with a mean of 8 years of formal training.

A is much	brighter	than *B* +2
A is	louder	than *B* +1
No difference	higher (in pitch)	0
B is	...	than *A* +1
B is much		than *B* +2

Figure 2. Scale used for dependent variables in Experiment 2.

5.2. Stimuli

Stimuli were generated from the two extremely different brightness-rated tones determined in Experiment 1 (‘piccolo’, rated as low in brightness at a given pitch [−] and ‘trumpet’, rated as high in brightness at a given pitch [+]). For each of these, versions were created at four pitches (E2, E4, A#4 and E5) and 2 loudness levels (using the default forte [f] and piano [p] setting in Finale notation software). A selection of these 2×4×2 stimuli were used for the experiment (combinations of tones used are shown in Table I.). Each stimulus pair was presented twice, with order of presentation reversed. Pairs were presented in random order.

For each of the stimuli, the centroids were calculated using PsySound [9]. For each pair (A/B) of stimuli the log ratio of centroids $\log(f_cA/f_cB)$ was calculated to predict the brightness rating according to the FC hypothesis (the A stimulus is brighter if $\log(f_cA/f_cB) > 0$). $\log([f_c/F_0]_A/[f_c/F_0]_B)$ was also calculated to predict the brightness response according to the F_0AC hypothesis (the A stimulus is brighter if $\log([f_c/F_0]_A/[f_c/F_0]_B) > 0$).

5.3. Procedure

Each participant sat at a computer and wore headphones for sound presentations. They were first given an interactive demonstration of the differences in timbre, pitch and loudness. For example, on the computer display for demonstrating brightness the participant was shown two buttons, and told that, when clicked, one sounds bright and the other sounds less bright. For this example the ‘trumpet’ sound and the ‘piccolo’ sound was used, for the bright and less bright examples respectively, each at the same pitch and dynamic marking. The participants heard the tones when they clicked on the corresponding button. When they completed this orientation phase, they began the main experiment. The presentations and data collection were conducted by an automated routine written by the first author.

In the main experiment, the participant clicked on a button marked ‘A’ and listened, then clicked on a button marked “B” and listened. The participant then rated the pitch, loudness and brightness differences within pairs on a five point scale, as shown in Figure 2. Participants rated other dimensions of the sound which may be related to timbre, namely vibrato [10], roughness, sharpness [11, 12]

Table I. Mean and standard error of brightness ratings of tone [A] with respect to tone [B]. For each tone pair, centroid, centroid-to-fundamental-frequency ratio, and loudness are shown, followed by the FC and F_0AC values used to plot Figure 3. In the stimulus pair column, f denotes loud, p denotes soft; '+' denotes bright timbre instrument ('trumpet'), '-' low brightness ('piccolo') as determined in Experiment 1. Results are shown in descending order of mean brightness ratings. Eq1: $\log(f_{cA}/f_{cB})$, Eq2: $\log([f_c/F_0]_A/[f_c/F_0]_B)$.

Stimulus pair	Tone A f_c [Hz]	Tone A f_c/F_0	Tone A Loudn. [Sone]	Tone B f_c [Hz]	Tone B f_c/F_0	Tone B Loudn. [Sone]	Eq1	Eq2	Perceptual Data: Brightness mean	Perceptual Data: Brightness SE
E2f+ / E2p-	704	9.17	67	179	2.14	53	0.595	0.632	1.47	0.235
E4f+ / E4f-	246	7.26	67	711	2.13	77	0.539	0.533	1.22	0.315
A#4f+ / A#4f-	276	6.26	68	991	2.20	79	0.445	0.454	1.13	0.325
E4p+ / E4f-	153	4.70	47	711	2.13	77	0.333	0.344	1.06	0.325
E5f+ / E5f-	246	7.26	67	1391	2.10	80	0.248	0.539	0.94	0.340
E2f+ / E2p+	704	9.17	67	639	8.03	49	0.042	0.058	0.53	0.240
A#4p+ / A#4f-	174	3.91	46	991	2.20	79	0.245	0.25	0.53	0.435
E2f+ / E4p-	704	9.17	67	695	2.08	53	0.006	0.644	0.06	0.415
E2f+ / E2f+	704	9.17	67	704	9.17	67	0	0	0.06	0.140
E5p+ / E5f-	143	2.16	50	1391	2.10	80	0.013	0.012	0.06	0.390
A#4f+ / A#4f+	276	6.26	68	2760	6.26	68	0	0	-0.06	0.060
E2f+ / E4f-	704	9.17	67	711	2.10	67	-0.004	0	-0.41	0.405
E2f+ / E4p+	704	9.17	67	1532	4.70	47	-0.338	0.29	-1.06	0.320
E2f+ / A#4p+	704	9.17	67	1744	3.91	46	-0.394	0.37	-1.06	0.325
E2f+ / E5f-	704	9.17	67	1391	2.10	80	-0.296	0.64	-1.38	0.265
E2f+ / E4f+	704	9.17	67	2394	7.26	67	-0.543	0.101	-1.47	0.225
E2f+ / A#4f+	704	9.17	67	2760	6.26	68	-0.593	0.166	-1.47	0.235
E2f+/E5f+	704	9.17	67	2222	3.49	69	-0.499	0.42	-1.59	0.215

and warmth [13][2]. The reason for requesting these additional ratings was to make the participant cognisant of several of the dimensions that contribute to the timbre of sound apart from brightness. Apart from sharpness (which is thought to be a good descriptor of the same construct as brightness – see [11] and [12]), these additional data are not reported here.

6. Results

Subjective response was scored by adjusting all responses as though stimulus A were being compared with stimulus B (i.e., within-subject pair-randomization was removed). If A were rated as being more bright, loud, . . . etc. it was scored +2. If A were rated slightly more bright, loud. . . it was scored +1. If there were no rated difference, it was scored 0. If B were rated as more bright, loud etc, then A was scored −2. Finally, if B were rated as slightly more bright, loud . . . , then A was scored −1 (see Figure 2). The FC and F_0AC values are shown in Table I listed in order of descending mean brightness rating.

In Figure 3, the ordinate is used to plot the mean and standard error for the extent to which stimulus A was judged brighter than B on the score described above. In the upper, the abscissa is the ratio of spectral centroids f_{cA} and f_{cB} calculated for A and B, on a log scale ($\log(f_{cA}/f_{cB})$). In the power figure, it is the log of the F_0 adjusted centroid ratios (i.e., $\log([f_c/F_0]_A/[f_c/F_0]_B)$). This figure shows the clear relation between f_c and brightness, and the lack of such a clear relation between brightness and f_c/F_0.

Table II shows that the log ratio of FC correlates significantly with brightness ($r = 0.502$, $p < 0.001$), whereas the log F_0AC ratio has a correlation that is statistically insignificant ($r = 0.03$, $p = 0.441$). The results are similar for sharpness (for FC, $r = 0.511$ at $p < 0.001$, and for F_0AC, $r = 0.029$ at $p = 0.453$), suggesting that two of the commonly used adjectives to describe this same perceptual dimension of timbre can be better predicted using the unadjusted centroid calculation. Pitch correlates significantly (at $p = 0.01$) with each of the variables, however the correlation coefficient with the F_0AC adjustment is small and negative ($r = -0.135$).

7. Discussion and conclusions

The second experiment demonstrates little evidence to support use of the F_0 adjusted centroid (f_c/F_0) as a predictor or brightness. In contrast, the simple spectral centroid was a good predictor of perceived brightness. The scatter plot in Figure 3 also demonstrates that there is no obvious effect due to changes in loudness nor due to harmonic or non- harmonic overlap of tones: loud and soft version of the same comparison stimuli maintain the linear trend, as do pure octave (such as E2 and E4) and tritone (such as E2 and A#4) combinations. In addition to being a good predictor of perceived brightness, the centroid of the total spectrum (FC) is also easy to calculate from sound

[2] Schubert [13] argued that Benedini's sharpness model should include an additional module that processes a quality of sounds he reported as being warmth which related to attenuation of non-consecutive harmonics in a steady state tone.

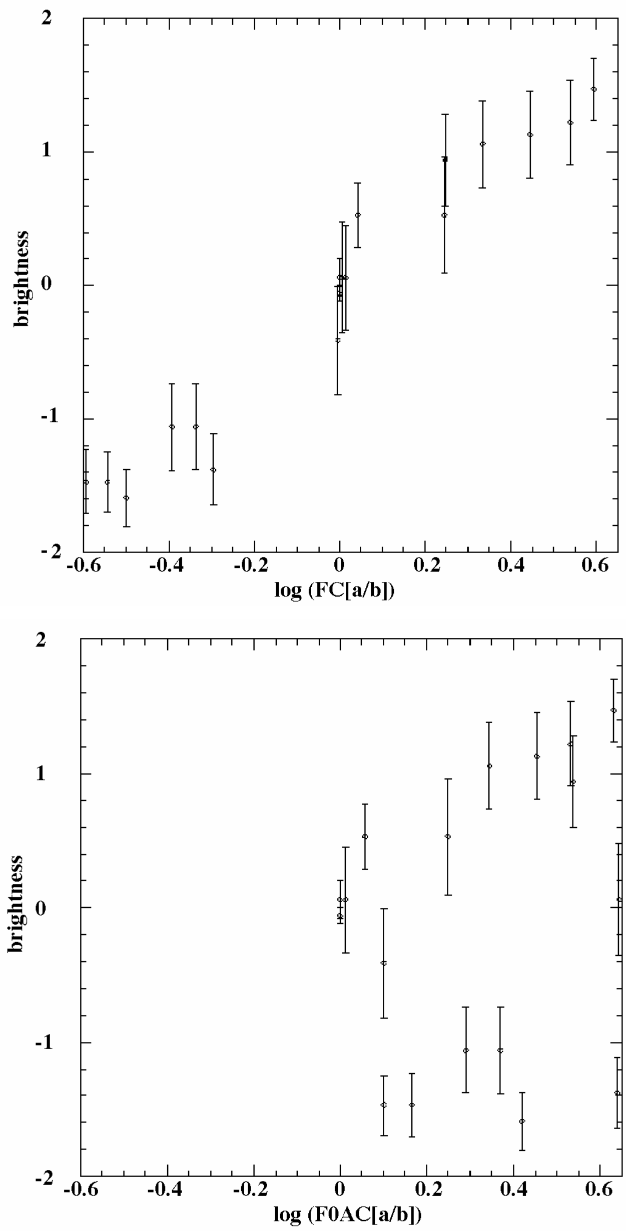

Figure 3. The reported differences in perceived brightness for single tone comparisons plotted as a function of the spectral centroid, FC [top] and the 'unitless centroid' F_0AC [bottom]. Further data are given in Table I.

recordings. This is a fortuitous result for researchers looking for acoustic models of brightness in rich and complex textures such as orchestras [14], bands and ensembles, in which many different F_0s are simultaneously present and in which the range of notes exceeds two octaves. However, further work is required to determine whether the FC model remains superior when more than one simultaneous tone is sounded.

The results of Experiment 1, which may seem superficially surprising, are now readily explained and provide a good example of how brightness is perceived. A (real) piccolo is regarded by musicians as having a bright timbre. So why did the musicians in our study find the MIDI 'piccolo' to have low brightness? The range of the piccolo is typically D5-C8, so its fundamental frequencies and all harmonics occur at frequencies above 580 Hz. Its f_c is inevitably higher than 580 Hz and may be much higher under the conditions where it is most noticeable: when played loudly in its top register (with a fundamental above 2 kHz). Compared with other instruments, members of the flute family, including the piccolo, have relatively weak higher harmonics. (Real) piccolos have high spectral centroids because they have high pitch, not because they have strong upper harmonics. In contrast, the MIDI 'piccolo' tone has the range of the MIDI keyboard. It has relatively weak higher harmonics, making it sound like a real piccolo when played in its high range. So, over the range of this experiment, it has a low spectral centroid. The MIDI 'trumpet' sound, on the other hand, has the strong higher harmonics that are characteristic of the sound of brass instruments. Consequently, at the same pitch, the 'trumpet' has a higher spectral centroid than does the 'piccolo'.

Table II. Pearson correlations matrix between F_0AC model, FC model, and perceived brightness B, sharpness S and pitch P. *: Correlation is significant at the 0.01 level (2-tailed). Eq1: $\log(f_{cA}/f_{cB})$, Eq2: $\log([f_c/F_0]_A/[f_c/F_0]_B)$.

	Eq2	Eq2	B	S	P
Eq2	1	.481*	.030	.029	-.135*
Eq1	.481*	1	.513*	.511*	.507*
B	.030	.513*	1	.565	.443
S	.029	.511*	.565	1	.440
P	-.135*	.507*	.443	.440	1

While Marozeau *et al.* found increasing differences between sounds produced by like instruments as pitch increased, their study examined dissimilarity perception, rather than brightness *per se*. Further research will need to be conducted to determine whether timbral brightness is pitch dependent for smaller interval differences.

The aim of the present study was to examine brightness perception using stimuli that are easy to reproduce. The study differs from that of Marozeau *et al.* because we compared intervals greater than an octave and did not ask the listener to try to ignore pitch differences. We argue that these conditions closely approximate real music listening experiences. Our study supports the use of a simple model to calculate perceptual brightness, where the centroid of power spectrum correlates well with the brightness of a tone. The model is better than an alternate model which adjusts the centroid according to the fundamental frequency of the sounding tone (or tones), and so we conclude that brightness is dependent upon F_0 to the extent that increasing F_0 also increases spectral centroid.

Acknowledgement

The authors are grateful to Alex Tarnopolsky and Paul Evans for their assistance in this study. This research was supported by an Australian Research Council Grant ARC-DP0452290 held by the first author.

References

[1] J. M. Grey, J. W. Gordon: Perceptual effects of spectral modifications on musical timbres. Journal of the Acoustical Society of America **63** (1978) 1493–1500.

[2] S. McAdams et al.: Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. Psychological Research **58** (1995) 177–192.

[3] S. Handel, M. L. Erickson: Sound source identification: The possible role of timbre transformations. Music. Perception **21** (2004) 587–610.

[4] E. Zwicker, H. Fastl: Psychoacoustics: Facts and models. Springer-Verlag, New York, 1990.

[5] R. Kendall, E. Carterette: Difference thresholds for timbre related to spectral centroid. Proceedings of the 4th International Conference on Music Perception and Cognition, Montreal, Canada, 1996, 91–95.

[6] R. Kendall: Musical timbre beyond a single note, II: Interactions of pitch chroma and spectral centroid. Proceedings of the 7th International Conference on Music Perception and Cognition, Sydney: Causal Productions, Adelaide, 2002, 596–599.

[7] J. Marozeau et al.: The dependency of timbre on fundamental frequency. Journal of the Acoustical Society of America **114** (2003) 2946–2957.

[8] C. L. Krumhansl: Cognitive foundations of musical pitch. Oxford University Press, New York, 1990.

[9] D. Cabrera: Psysound: A computer program for the psychoacoustical analysis of music. Mikropolyphonie **5** (1999) http://www.mikropol.net/ accessed June 14, 2006.

[10] K. Jensen: The timbre model. Workshop on current research directions in computer music, Barcelona, Spain, 2001, 174–186.

[11] K. Benedini: A functional model of timbre differences. Biological Cybernetics **34** (1979) 111–117.

[12] M. Ilkowska, A. Miskiewicz: Sharpness versus brightness: A comparison of magnitude estimates. Acta Acustica united with Acustica **92** (2006) (this issue).

[13] E. Schubert: A model for timbre perception. Undergraduate thesis, School of Electrical Engineering. University of New South Wales, Sydney, 1985.

[14] E. Schubert: Modeling perceived emotion with continuous musical features. Music Perception **21** (2004) 561–585.