ARCHIVES OF ACOUSTICS
36, 2, 225–238 (2011)
DOI: 10.2478/v10168-011-0018-8

Perceptually Correlated Parameters of Musical Instrument Tones[(∗)]

James W. BEAUCHAMP

University of Illinois at Urbana-Champaign
School of Music and Dept. of Electrical & Computer Engineering
Urbana, Illinois USA
e-mail: jwbeauch@illinois.edu

(*received January 3, 2011; accepted March 11, 2011*)

In Western music culture instruments have been developed according to unique instrument acoustical features based on types of excitation, resonance, and radiation. These include the woodwind, brass, bowed and plucked string, and percussion families of instruments. On the other hand, instrument performance depends on musical training, and music listening depends on perception of instrument output. Since musical signals are easier to understand in the frequency domain than the time domain, much effort has been made to perform spectral analysis and extract salient parameters, such as spectral centroids, in order to create simplified synthesis models for musical instrument sound synthesis. Moreover, perceptual tests have been made to determine the relative importance of various parameters, such as spectral centroid variation, spectral incoherence, and spectral irregularity. It turns out that the importance of particular parameters depends on both their strengths within musical sounds as well as the robustness of their effect on perception. Methods that the author and his colleagues have used to explore timbre perception are: 1) discrimination of parameter reduction or elimination; 2) dissimilarity judgments together with multidimensional scaling; 3) informal listening to sound morphing examples. This paper discusses ramifications of this work for sound synthesis and timbre transposition.

Keywords: musical timbre, music synthesis, loudness, pitch, duration, attack, decay, spectral envelope, spectral centroid, spectral irregularity, spectral flux, vibrato, inharmonicity, discrimination, dissimilarity, multidimensional scaling (MDS), timbre transposition, rms amplitude, fundamental frequency, correspondence.

1. Introduction

The principal long-term goal of this study is to achieve a synthesis system where a minimal set of independent but perceptually meaningful parameters are

[(∗)] This paper is based on a talk of the same title given by the author at the Second Vienna Talk on Musical Acoustics, held in Vienna, Austria during September 19–21, 2010, and on the associated proceedings paper of the same title.

used to control and synthesize musically useful sounds, including sounds of traditional musical instruments. Basic steps for accomplishing this goal are a) using spectral analysis to obtain static and time-varying parameters; b) building synthesis models to utilize these parameters; c) conducting formal listening tests on single sounds to test the efficacy of the models; and d) conducting informal listening tests using synthesis of extended musical passages.

Although it might seem that this goal could be achieved in a few weeks or months, in practice, musical timbre has been studied for decades using a series of less ambitious steps. Typically the first step is to select a group of musical sounds to study. The parameters to be identified from the sounds are first of all the time-varying amplitudes and frequencies obtained from spectral analysis. Then, more detailed, possibly perceptually important parameters can be inferred, such as attack and decay times, spectral envelope features (such as spectral centroid, spectral irregularity, and spectral flux), vibrato characteristics, and inharmonicity (McAdams *et al.*, 1999; Beauchamp, Lakatos, 2002).

Conducting a formal listening test for timbre requires the following steps:
- Stimuli preparation.
- Psychoacoustic testing (the actual listening test).
- Data processing and presentation.
- Interpretation of results.

Either synthetic or recorded acoustic ("real") sounds can be used as stimuli, but in either case they should be normalized to eliminate sonic attributes that are not part of timbre, namely, loudness, pitch, and duration. The latter two are not a problem for synthetic sounds (sounds consisting solely of harmonically related frequencies), but for either sound type loudness equalization through gain factor adjustment must be achieved by additional loudness testing, by using a loudness prediction program (Moore, 1997), or by randomizing the levels of the stimuli (Dai, 2008). For pitch normalization of harmonic sounds it is generally acceptable to make certain that the fundamental frequencies are the same. Duration is a bit more complicated, but a method is given in (McAdams *et al.*, 1999), where attack and decay structures are retained and the total duration is reduced to a standard 2 s.

If a test is designed to only compare different acoustic sounds, as in the case of a dissimilarity test, no further stimuli modifications may be necessary. Physical (spectral) differences between sounds can be measured in terms of specific parameters and correlated with the measured perceptual differences. However, for a discrimination test the experimenter will often want to modify specific acoustic parameters of the sounds and then examine how discrimination ability varies with each parameter that is changed or, in more detail, the amount of change of each parameter.

Important questions are: 1) What specific parameters should be varied? 2) Why do we choose these particular parameters? 3) How do we measure them? 4) How do we vary them? For the studies reviewed in this paper, the specific parameters are spectral irregularity, spectral flux, spectral centroid variation, amplitude and frequency microvariations, and inharmonicity. Reasons for choosing these parameters are discussed in the timbre literature (see Hajda (2007) and Donnadieu (2007) for reviews.) Methods for measuring and varying them are given in (McAdams *et al.*, 1999) and (Beauchamp, Lakatos, 2002).

Also, it should be remarked that two specific parameters, average spectral centroid and attack time, have proved to be so salient that they are sometimes factored out (equalized or normalized) from the stimuli. Such was the case with (Beauchamp, Lakatos, 2002) and (Horner *et al.*, 2006) and is a method utilized in the second study covered in this paper.

Preparation of a psychoacoustic test requires the selection of listener subjects and the design of the test. Chosen listeners are generally young people with good hearing and are divided between those with extensive and those with meager musical experience. The two formal tests described in this review paper use either timbre discrimination or dissimilarity judgments. With discrimination, the subjects are generally asked to judge whether pairs of sounds are same or different. With a dissimilarity judgment task, they are asked to estimate the amount of timbral difference between tone pairs on a scale of say 0 to 10.

Once a test is completed and the data is collected, the data can be processed in various ways. Discrimination averages can be simply presented, or graphs of discrimination vs. a particular parameter can be illustrated. In the case of dissimilarity judgments, the method of multidimensional scaling (MDS) is commonly used to display the positions of the sound stimuli in a two- or three-dimensional space (Miller, Carterette, 1975; Grey, 1977). To show trends within the space the dimensions can be correlated with measures of various spectrotemporal parameters of the stimuli.

Finally, an interpretation of the data is usually given. One of the biggest problems is estimating the scope of validity of the results. The scope is necessarily limited because it is difficult to design tests that cover a wide range of cases and can still be conducted over a reasonable time period.

2. Three timbre studies

Three projects are reviewed, a timbre discrimination study (McAdams *et al.*, 1999), a timbre dissimilarity judgment study with MDS solution (Beauchamp *et al.*, 2006; Hall *et al.*, 2010), and a timbre transposition study (Beauchamp, Bay, 2008).

2.1. Timbre discrimination study

The objective of this study (McAdams *et al.*, 1999), originally published in 1999, was to investigate the relative importance of some different spectrotemporal parameters by simplifying musical sounds with respect to these parameters. The stimuli prototypes (reference sounds) consisted of tones performed on seven different instruments: clarinet, flute, oboe, trumpet, violin, harpsichord, and marimba at pitch E_4^b (311 Hz). Loudnesses were equalized using a brief test, and durations were equalized to 2 s using a method described in (McAdams *et al.*, 1999). The sound signals were analyzed using a pitch-synchronous short-time Fourier transform program (Beauchamp, 2007), and the resulting partial amplitude and frequency data were simplified as follows:

1) partial amplitude-vs.-time envelopes smoothed,
2) spectral envelope smoothed (irregularity reduced),
3) spectral flux (otherwise referred to as *incoherence*) eliminated,
4) partial frequency-vs.-time envelopes smoothed,
5) partial frequencies locked to harmonics of a time-varying fundamental,
6) partial frequencies flattened to harmonics of a fixed fundamental.

The sounds were then resynthesized to the time domain by additive synthesis. Note that the word *partial* is used here instead of *harmonic* because even though the frequencies of these tones are close to harmonic, departures from harmonicity are possible.

Although there was considerable variation with instrument (see McAdams *et al.*, (1999) for details), the discrimination results averaged over the seven instruments were:

a) spectral envelope smoothed – 96%,
b) spectral flux eliminated – 91%,
c) frequencies flattened – 71%,
d) frequency envelopes smoothed – 70%,
e) frequencies locked harmonically – 69%,
f) amplitude envelopes smoothed – 66%.

An interpretation of these results is that the spectral parameters *spectral irregularity* (i.e., spectral jaggedness) and *spectral flux* (change of spectrum shape over time) are, for this set of instruments, most salient. Smoothing the amplitude and frequency envelopes (using a 10 Hz cutoff low-pass filter) eliminates fine-grained temporal detail, but this elimination is relatively unnoticeable. So is locking the frequencies harmonically or removing any trace departure from fixed harmonics.

However, when an error metric (similar to those discussed in (Horner *et al.*, 2006) was constructed based on differences between reference and modified partial amplitudes and a regression line was constructed to fit discrimination (given in terms of d') against the log of this error, it was found that the regression straight line explained 77% of the discrimination variance (88% if one outlier point was removed). Since modifications a), b), and f) could also change the spectral centroid, d' was also plotted against the log of *normalized spectral centroid difference* between the reference and modified sounds. In this case, a regression straight line explained only 54% of variance, but when both the spectral centroid difference and the partial amplitude error metric were combined together, 83% of variance was explained (with no outliers removed).

A final interpretation from these results is that, yes, spectral irregularity and flux are important specific parameters, but discrimination is also strongly correlated with a total metric difference between the time-varying spectra of two similar sounds.

2.2. Timbre dissimilarity study

With this study, originally presented as a talk in 2006 (Beauchamp *et al.*, 2006), subjects were presented with the task of judging dissimilarity between musical sounds. The original stimuli consisted of tones performed on ten sustained-tone instruments: bassoon, cello, clarinet, flute, horn, oboe, recorder, alto saxophone, trumpet, and violin. Two types of tones were constructed from these: *dynamic* (with flux) and *static* (without flux). The tones were also equalized with respect to pitch ($F_0 = 311$ Hz), attack time (.05 s), decay time (.05 s static, .15 s dynamic), total duration (0.5 s static, 2.0 s dynamic), loudness (Moore *et al.*, 1997), and average normalized spectral centroid (3.7). Average centroids were equalized by applying a fixed multiplier k^p to each harmonic k's amplitude, where p was varied to achieve the desired centroid value, as described in (Beauchamp, Lakatos, 2002).

The listening test employed ten musically experienced subjects to judge dissimilarity between tone pairs using a method of triadic comparison (Plomp, 1970). While average dissimilarity scores could theoretically vary from 0 to 17, actual scores varied from about 4 to 13. The scores were placed in 10×10 dissimilarity matrices (see Table 1).

Two different classical multidimensional scaling (MDS) programs, SPSS and Matlab, were used to process the dissimilarity matrices. For the static tones, only 2D solutions were made, whereas both 2D and 3D solutions were made for the dynamic tones. Stresses (average normalized difference between inter-timbre distances given by a dissimilarity matrix and those given by a corresponding MDS solution) for the 2D SPSS and Matlab solutions were both 0.12 for the static case and 0.15–0.17 for the dynamic case; for both 3D solutions for the dynamic case, stresses were reduced to 0.095.

(It was somewhat of a surprise for this author to discover the degree to which the distances between pairs of timbres in an MDS solution do not exactly match the values given by the dissimilarity matrix and that *stress* is commonly given by MDS programs as an important measure of their average agreement. Stress generally decreases as the number of dimensions increases, but for visualization 2 or 3 dimensional solutions are preferred. Stress is useful for estimating the accuracy of an MDS solution. Unfortunately, in reading several papers on musical timbre employing the MDS method, this author could not find a single mention of the word *stress*, even though it is a very basic concept in the theory of MDS.)

Table 1. Dissimilarity matrices for static (upper) and dynamic (lower) tones (Bs = bassoon, Ce = cello, Cl = clarinet, Fl = flute, Hn = horn, Ob = oboe, Rc = recorder, Sx = saxophone, Tp = trumpet, Vn = violin). (Reproduced by permission from Table 1 of (Hall *et al.*, 2010).)

Static tones:

	Bs	Ce	Cl	Fl	Hn	Ob	Rc	Sx	Tp	Vn
Bs	0	10.6	11.3	5.3	8.3	8.4	11.6	7.5	7.4	8.4
Ce	10.6	0	3.8	7.6	11.4	11.4	4.9	10.4	9.4	9.8
Cl	11.3	3.8	0	8.2	12.7	8.8	3.6	11.5	10.8	9.7
Fl	5.3	7.6	8.2	0	9.9	8.5	9.3	6.7	9.2	9.5
Hn	8.3	11.4	12.7	9.9	0	9.0	11.9	7.9	9.8	9.5
Ob	8.4	11.4	8.8	8.5	9.0	0	10.0	10.3	5.8	7.3
Rc	11.6	4.9	3.6	9.3	11.9	10.0	0	11.8	9.6	10.0
Sx	7.5	10.4	11.5	6.7	7.9	10.3	11.8	0	9.3	8.4
Tp	7.4	9.4	10.8	9.2	9.8	5.8	9.6	9.3	0	8.5
Vn	8.4	9.8	9.7	9.5	9.5	7.3	10.0	8.4	8.5	0

Dynamic tones:

	Bs	Ce	Cl	Fl	Hn	Ob	Rc	Sx	Tp	Vn
Bs	0	11.0	9.7	7.3	5.6	6.6	10.2	6.2	7.9	9.4
Ce	11.0	0	7.3	9.7	13.3	11.1	8.0	9.4	9.5	7.5
Cl	9.7	7.3	0	10.1	11.6	6.4	4.7	9.6	9.9	11.9
Fl	7.3	9.7	10.1	0	9.7	9.5	6.7	9.8	8.0	10.3
Hn	5.6	13.3	11.6	9.7	0	6.3	11.1	7.9	9.6	9.1
Ob	6.6	11.1	6.4	9.5	6.3	0	9.4	9.4	7.3	8.8
Rc	10.2	8.0	4.7	6.7	11.1	9.4	0	10.9	9.6	11.1
Sx	6.2	9.4	9.6	9.8	7.9	9.4	10.9	0	8.7	9.8
Tp	7.9	9.5	9.9	8.0	9.6	7.3	9.6	8.7	0	8.1
Vn	9.4	7.5	11.9	10.3	9.1	8.8	11.1	9.8	8.1	0

Meanwhile, static tone solutions were correlated with two parameters measured from the sound signals, *even/odd harmonic ratio* (ratio of the average rms amplitude of the even harmonic amplitudes to that of the odd harmonics) and *spectral irregularity*. The dynamic tone solutions were correlated with those parameters plus two others: *spectral flux* (also referred to as *incoherence*) and *normalized spectral centroid variation* (spectral centroid standard deviation divided by its average value). All MDS solutions were rotated so that the best possible even/odd correlation aligned with the horizontal axis. For the other parameters, best-fit straight lines of highest correlation to the various parameters were computed.

As can be seen in Figs. 1–3, the corresponding SPSS and Matlab solutions can be quite different. However, for the 2D cases, instrument groupings appear to be very similar (see Figs. 1 and 2). The most obvious 2D static case groupings are {recorder, clarinet, cello} and {trumpet, oboe, violin}. R^2 correspondences with

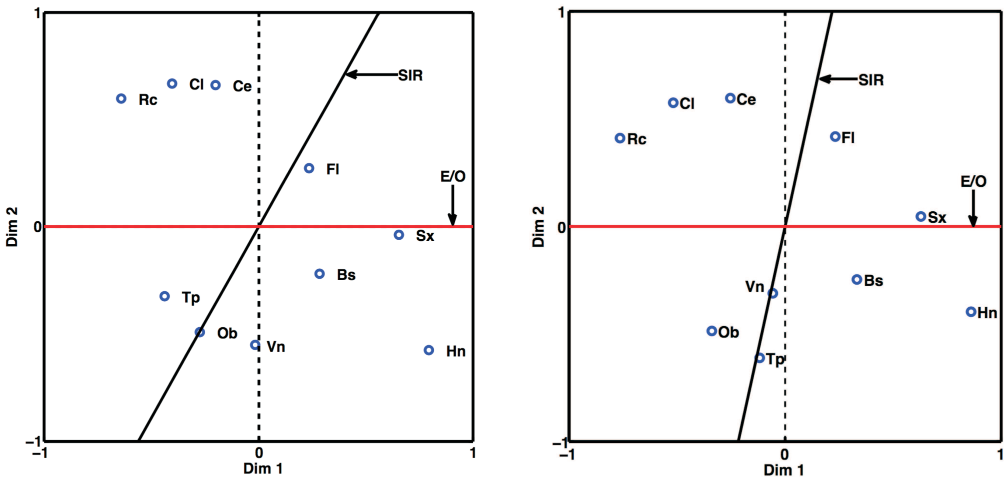

Fig. 1. Two-dimensional MDS solutions for the static tones: SPSS (left), Matlab (right). (Reproduced and enhanced from Fig. 1 of (Hall *et al.*, 2010) by permission).

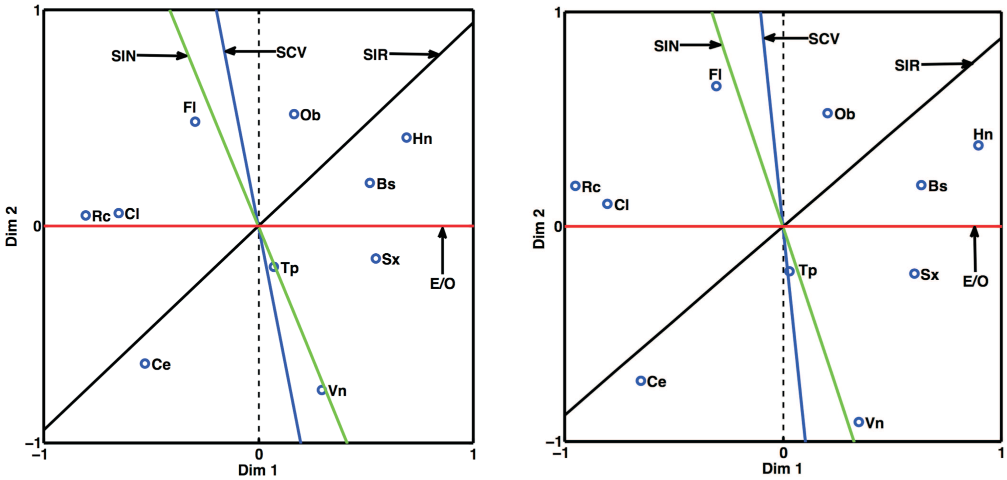

Fig. 2. Two-dimensional MDS solutions for the dynamic tones: SPSS (left), Matlab (right). (Reproduced and enhanced from Fig. 2 of (Hall *et al.*, 2010) by permission).

the even/odd (E/O) and spectral irregularity (SIR) parameters were measured at 78–79% and 69–75%, respectively, for the two solutions. For the 2D dynamic case (see Fig. 2), the correspondences were 71–69% for even/odd, 68–68% for spectral centroid variation, 56–53% for spectral incoherence, and 39–40% for spectral irregularity. Also, the spectral centroid variation (SCV) and spectral incoherence (SIN) straight lines of greatest correspondence appear close together, suggesting that these variables are tightly correlated in the stimuli.

For the 3D dynamic case (see Fig. 3), the correspondences for even/odd, spectral centroid variation, spectral incoherence, and spectral irregularity were 82–68%, 83–82%, 53–83%, and 82–71%, respectively, indicating rather strong dis-

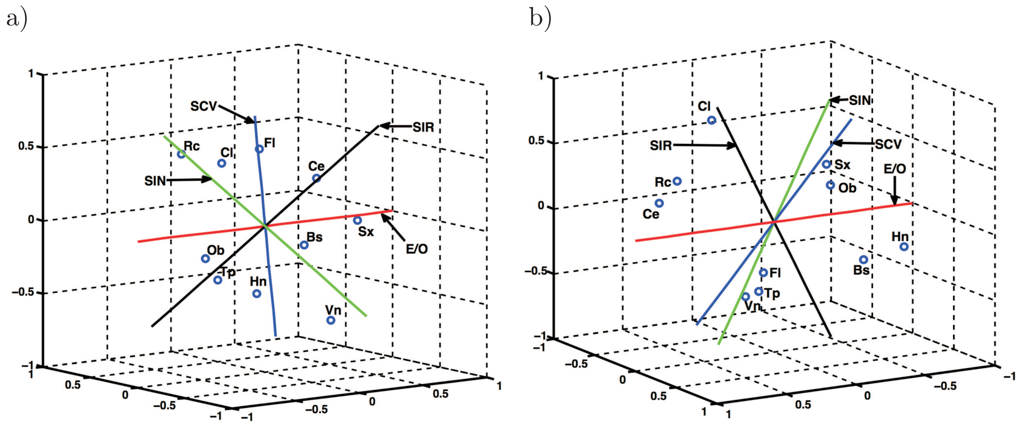

Fig. 3. Three-dimensional MDS solutions for the dynamic tones: SPSS (a), Matlab (b). (Reproduced and enhanced from Fig. 3 of (Hall *et al.*, 2010) by permission).

agreement between the SPSS and Matlab solutions as to the saliency of 3 out of 4 of the parameters. Averaging over the two solutions gives 82.5% for spectral centroid variation, 76.5% for spectral irregularity, 75% for even/odd, and 68% for spectral incoherence. Therefore, assuming that 3D solutions are best for the dynamic case because of their relatively low stress, it appears that for dynamic tones, spectral centroid variation is the parameter with the highest and most consistent saliency (beyond average centroid and attack/decay, which were equalized in the stimuli). On the other hand, any of the four parameters corresponds as well as the others for at least one of the two solutions. Also, it is curious that the average correspondence for the four parameters is about the same for the SPSS solution (75%) as for the Matlab solution (76%), which means it would be difficult to conclude that one solution is *better* than the other, even though they certainly are significantly different (average correspondence difference equals 14%).

After making all of these computations one might ask: What is the advantage of using MDS? Why not just correlate with the original dissimilarity data? Certainly MDS yields some attractive pictures, showing the relative positions of timbres relative to one another. But as the two 3D solutions for dynamic tones show, different solutions with the same stress can result in instruments in very different positions in "timbre space" and can yield quite different correspondences. At least with the original dissimilarity matrix, there is only one set of data to correlate with, and it has no stress.

2.3. Timbre transposition study

The point of this study, presented as a talk in 2008 (Beauchamp, Bay, 2008), was to explore synthesis using a small set of time-variable control parameters and a family of spectral envelopes (Beauchamp, 2007; Luce, Clark, 1967), which represent a particular instrument, but then switch the spectral envelope family to a different instrument and see what happens. Either the spectral envelopes or the temporal data will dominate, or a hybrid instrument that shares characteristics of both instruments will be produced. The instrument supplying the time-varying parameters is called the *source instrument*, and the one supplying the spectral envelope family is called the *target instrument*. The process is called *timbre transposition*.

In an earlier project we discovered that by using three time-varying parameters, rms amplitude, fundamental frequency (pitch), and spectral centroid (i.e., $A_{rms}(t)$, $f_0(t)$, and $f_c(t)$), combined with a spectral envelope family based on spectral centroid clustering, we could produce trumpet tones that were quite realistic (Beauchamp, 2007). The spectral envelope family was derived from a training set of trumpet tones that covered a wide gamut of pitches and dynamics (i.e., intensity levels). Every frame of every tone was analyzed (using the pitch-synchronous analyzer) and sorted into different "bins" based on ranges of centroid values, 0–200, 200–400, etc. The spectra in each bin were normalized and then sorted into critical bands, and finally the amplitudes within each band were averaged to give a single value that represents the band amplitude for a particular centroid range. These amplitudes as functions of the band center frequencies formed a collection of spectral envelopes each of which corresponded to a range of centroids. A demonstration of this method using a restricted parametric model for the temporal variations is given on the author's website (http://ems.music.uiuc.edu/beaucham/). This includes the addition of low-frequency noise microvariations to the pitch and amplitude controls to increase the realism of the trumpet synthesis.

This method can be applied to other instruments as well. Synthesis is done by first deriving representative time-varying parameters $A_{rms}(t)$, $f_0(t)$, and $f_c(t)$ from a solo recording. For each time instant, $f_c(t)$ is used to retrieve a spectral envelope from a spectral envelope library, and estimated harmonic amplitudes are obtained from the spectral envelope by sampling it at frequencies $kf_0(t)$, where k is the harmonic number. These amplitudes are then easily scaled to match the total amplitude $A_{rms}(t)$. The sound is synthesized using additive sinusoidal synthesis. Figure 4 shows block diagrams for the analysis/resynthesis procedures.

There is a question of whether, for a given instrument, a library containing a single spectral envelope family is adequate for all fundamental frequencies (f_0's) or whether the library families need to change as a function of f_0. It seems to be the case (but is not proven) that single families are adequate for brass instruments (Luce, Clark, 1967), which have robust global spectral envelopes that work over a wide range of f_0, but not for some woodwinds or strings, where specific spectral envelopes for a variety of pitches may be needed. However, with the abundance of memory available in computers these days, it is entirely reasonable to compute and store a different family for many different values of f_0. Thus, spectral envelope can become a function of both spectral centroid (f_c) and pitch (f_0).

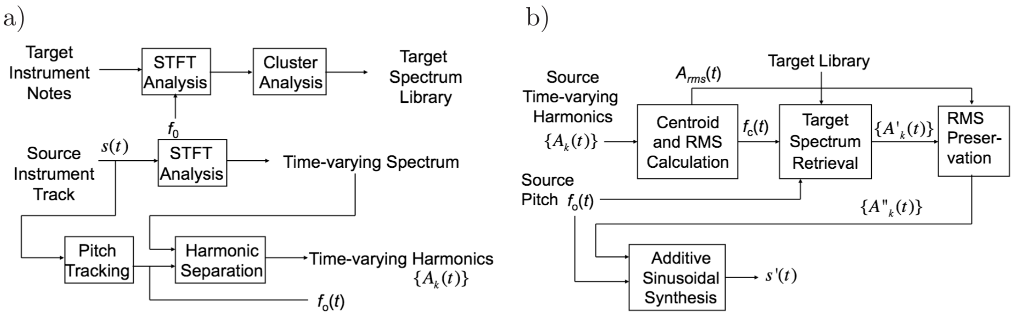

Fig. 4. Block diagrams for: a) target spectrum library computation (upper left) and time-varying pitch and harmonic amplitude computation (lower left); b) spectral centroid and rms amplitude computation, target spectrum retrieval, and harmonic amplitude scaling for rms amplitude preservation (upper right) and sinusoidal additive synthesis (lower right).

Figure 5 shows a family of spectral envelopes for an oboe, which is averaged over a range of f_0's ($262 \leq f_0 \leq 1187$ Hz), and for a clarinet, which is valid only for a specific value of f_0 (233 Hz). Note that if global spectral envelopes were attempted for the clarinet, the relative weaknesses of the first few even harmonics would be smeared out, severely compromising resynthesis.

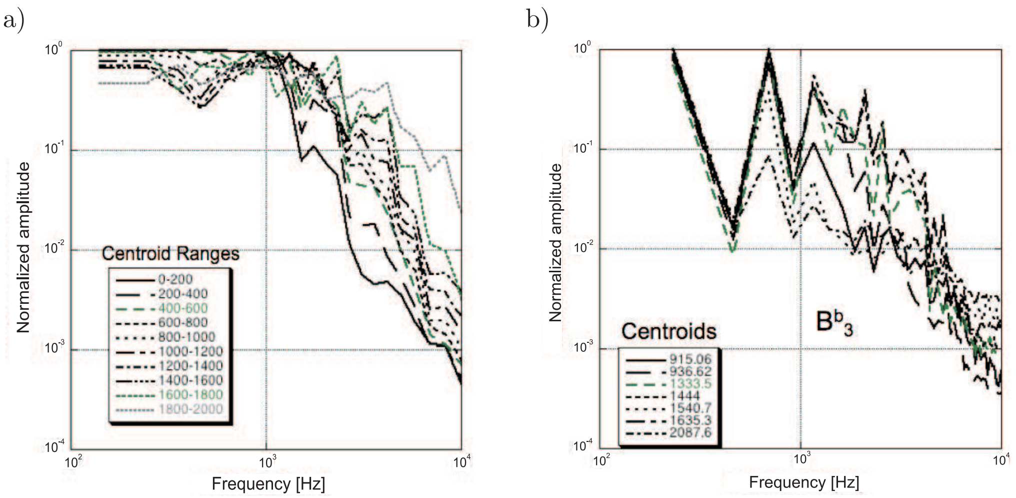

Fig. 5. Spectral envelope families for oboe (left) and clarinet (right). The oboe family is averaged over frequency and each spectral envelope represents a range of spectral centroids as given in the box. The clarinet family is specific to the pitch B^b_3 ($f_0 = 233$ Hz), and each envelope has its centroid given in the box. For both cases, maximum amplitudes of the spectral envelopes are normalized to 1.0.

Examples of $f_0(t)$, $A_{rms}(t)$, and $f_c(t)$ for a short musical phrase performed on a clarinet are given in Fig. 6. These parameters are used to control sound synthesis as depicted in Fig. 4.

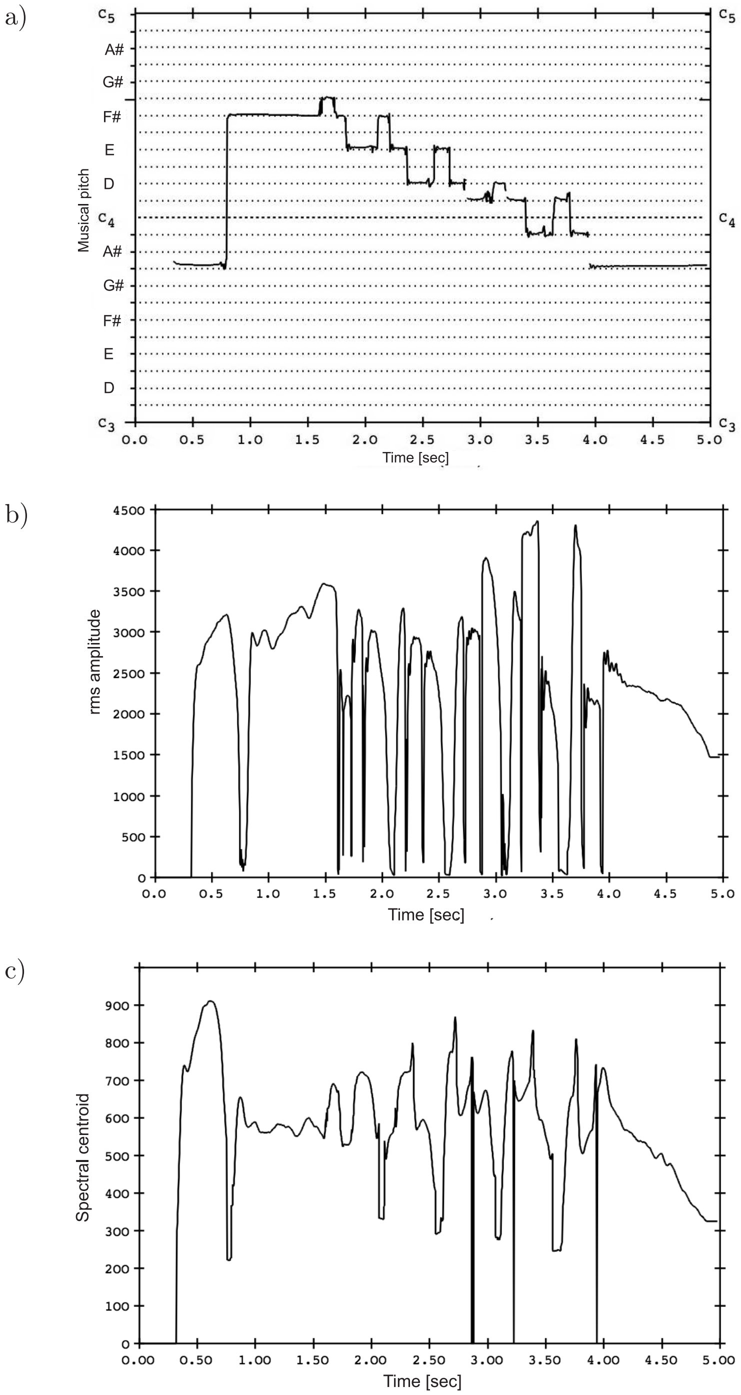

Fig. 6. Parameters to control synthesis of a clarinet phrase: a) pitch ($f_0(t)$), b) rms amplitude ($A_{rms}(t)$), and c) spectral centroid ($f_c(t)$).

It is fairly obvious that there needs to be a match between the pitch and centroid ranges of the source and target instruments. Thus, if the source and target pitch ranges and centroid ranges do not overlap sufficiently, timbre transposition will not work. However, the control ranges coming from the source can be easily mapped to correspond to the best ranges for the target using simple linear equations.

Findings for this study were informal. It appears that if the temporal data of the source instrument is similar to that of the target instrument, the result will likely be identified as the target instrument. On the other hand, if the target instrument's spectral envelope family is similar to that of the source, the temporal information of the source may dominate and the result may still be identified as the source instrument. We have found this to be true if the source is a bassoon playing a series of low-pitched short-duration notes, and a horn is the target. In between these cases are cases where neither the temporal information nor the spectral data are similar for the two instruments, so a true hybrid is produced. For example, if horn is the source and clarinet is the target, a situation ensues where the resulting hybrid sound can be recognized simultaneously as a horn in terms its temporal envelopes and as a clarinet in terms of its unique spectrum with emphasis on odd harmonics.

Gradual morphing or interpolation between the instruments could be produced by cross-fading the temporal controls or the spectral envelopes or both. We haven't actually tried this yet, but there is no reason why the method should not produce interesting results.

Another possibility to investigate is the addition of external controls designed to modify parameters such as those shown to be salient in the 1999 and 2006 timbre studies discussed above. Exactly how to accomplish this has yet to be determined.

3. Conclusions

The 1999 timbre study, which utilized parameter simplification and discrimination, indicated that spectral irregularity and spectral flux were more important than amplitude and frequency microvariations and inharmonicity. However, this author would take that result with a grain of salt because it is well known that temporal details and inharmonicity are important for instrument recognition and for warmth and realism.

The 2006 dissimilarity study, which used multidimensional scaling to summarize relative perceptual distances between instrument timbres, yielded some interesting results and raised some nontrivial issues. One issue was the unexpected importance of the concept of stress (see above). Another was the usefulness of rotation for comparing solutions using different MDS programs (SPSS and Matlab). Yet another was that solutions from different MDS programs can be quite different, although for the same number of dimensions their stresses tend to be in approximate agreement. Still another was that straight lines aligned to maximize R^2 correspondence with particular spectral parameters do not normally correspond to dimensional axes, although rotation can be used force one of the lines to coincide with an axis. Finally, different MDS programs can yield different correspondences with same spectral parameters, making exact conclusions about the saliency of these parameters problematic. Nonetheless, our conclusions from the MDS solutions can be summarized as follows: For static tones (those without flux), two different 2D solutions, with stresses of 12%, had quite high correspondences with even/odd harmonic ratio (78–79%), better than spectral irregularity (69–75%). For dynamic tones (those with flux), for two different 2D solutions with stresses of 15 and 17%, even/odd corresponded best (69–71%), followed by spectral centroid variation, spectral flux, and spectral irregularity. With the 3D solutions for these tones, the stresses dropped to 9.5% and some correspondences increased to 82–83%, but there was very significant disagreement between the two solutions, except for spectral centroid variation, where both solutions yielded R^2 close to 82%.

The 2008 timbre transposition study showed that combining some time-variant parameters with fixed spectral envelopes not only provides a compact resynthesis model for a given instrument, but it can also serve as a method for merging the temporal characteristics of one instrument with the spectral characteristics of another. In some cases the resulting sounds demonstrate one of the two instruments dominating the other. However, when differences between corresponding temporal and spectral characteristics of the two instruments are both pronounced, a true hybrid is generally produced, where the temporal (articulatory) characteristic can be recognized as coming from one instrument and the spectral (tone color) characteristic is perceived to be coming from the other.

References

1. Beauchamp J.W. (2007), *Analysis and Synthesis of Musical Instrument Sounds*, [in:] Analysis, Synthesis, and Perception of Musical Sounds: The Sound of Music, J.W. Beauchamp [Ed.], pp. 1–89, Springer.

2. Beauchamp J.W., Bay M. (2008), *Timbre transposition based on time-varying spectral analysis of continuous monophonic audio and precomputed spectral libraries* (abstract), J. Acoust. Soc. Am., **123**, 5, 2, 3805.

3. Beauchamp J.W., Lakatos S. (2002), *New spectro-temporal measures of musical instrument sounds used for a study of timbral similarity of rise-time- and centroid-normalized musical sounds*, Proc. 7 Int. Conf. on Music Perception and Cognition (ICMPC 7), Sydney, Australia, pp. 592–595.

4. Beauchamp J.W., Horner A.B., Koehn H.-F., Bay M. (2006), *Multidimensional scaling analysis of centroid- and attack/decay-normalized musical instrument sounds (abstract)*, J. Acoust. Soc. Am., **120**, 5, 2, 3276.

5. Dai H. (2008), *On suppressing unwanted cues via randomization*, Perception & Psychophysics, **70**, 7, 1379–1382.

6. Donnadieu S. (2007), *Mental Representation of the Timbre of Complex Sounds*, [in:] Analysis, Synthesis, and Perception of Musical Sounds: The Sound of Music, J.W. Beauchamp [Ed.], pp. 272–319, Springer.

7. Grey J.M. (1977), *Multidimensional perceptual scaling of musical timbres*, J. Acoust. Soc. Am., **61**, 5, 1270–1277.

8. Hajda J.M. (2007), *The Effect of Dynamic Acoustical Features on Musical Timbre*, [in:] Analysis, Synthesis, and Perception of Musical Sounds: The Sound of Music, J.W. Beauchamp [Ed.], pp. 250–271, Springer.

9. Hall M.D., Beauchamp J.W., Horner A.B., Roche J.M. (2010), *Importance of Spectral Detail in Musical Instrument Timbre*, Proc. 11th Conf. on Music Perception and Cognition (ICMPC 11), Seattle, Washington, USA, pp. 69-74.

10. Horner A.B., Beauchamp J.W., So R.H.Y. (2006), *A Search for Best Error Metrics to Predict Discrimination of Original and Spectrally Altered Musical Instrument Sounds*, J. Audio Eng. Soc., **54**, 3, 140–156.

11. Luce D., Clark M. Jr. (1967), *Physical Correlates of Brass-Instrument Tones*, J. Acoust. Soc. Am., **42**, 6, 1232–1243.

12. McAdams S., Beauchamp J.W., Meneguzzi S. (1999), *Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters*, J. Acoust. Soc. Am., **105**, 2, 1, 882–897.

13. Miller J.R., Carterette E.C. (1975), *Perceptual space for musical structures*, J. Acoust. Soc. Am., **58**, 3, 711–720.

14. Moore B.C.J., Glasberg B.R., Baer T. (1997), *A Model for the Prediction of Thresholds, Loudness, and Partial Loudness*, J. Audio Eng. Soc., **45**, 4, 224–240.

15. Plomp R. (1970), *Timbre as a multidimensional attribute of complex tones*, [in:] Frequency Analysis and Periodicity Detection in Hearing, R. Plomp and G.F. Smoorenburg [Eds.], pp. 397–414, A.W. Sijtohoff, Leiden.

16. http://ems.music.uiuc.edu/beaucham/