Spectral Envelope Correction for Real-Time Transposition: Proposal of a "Floating-Formant" Method

Richard Dudas

Cycling '74, 379A Clementina Street, San Francisco, CA 94103 USA
email: dudas@cycling74.com

Abstract

A common problem of sound transposition is that the spectra of the resulting transpositions are often very unconvincing to the human ear in terms of their relation to the original sound. This paper examines the problem, takes a look at some existing solutions and their drawbacks and proposes a generalized method to correct this "unconvincingness" which can be used on a variety of instrumental and non-instrumental sounds.

1 Introduction

At some point in our electro-acoustic and computer-music training, most (if not all) of us will have made the realization that transposing recorded sounds by increasing or decreasing their playback speed will result in sounds which, though evidently related to the original, seem quite unnatural and generally unconvincing to our timbre-sensitive ears. The same problem also plagues the various spectral sound transposition methods, whereby either tracks of sinusoidal partials derived from analysis or frequencies defined by peak amplitude in a spectrum obtained via STFT are transposed by simple frequency scaling. Often sounds transposed in this manner either sound too bright when transposed up, or too dull when transposed down. This is seemingly the inverse of the natural timbral behaviour of many acoustic instruments (including the singing voice), whose low registers produce tones generally rich in higher harmonics, and whose high registers usually produce sounds whose waveforms begin to approach pure sinusoidal tones.

In the past half century, there has been an important body of research into our perception of musical timbre and the various characteristics which help the brain associate related sounds. It has been shown that a sound's spectral envelope is the basic defining factor for its timbre (Grey 1975). Therefore, it seems logical that by adjusting the spectral envelope of a crudely transposed sound, we should be able to correct to some degree the unnatural quality of simple linear transposition. Indeed, this has been shown to be true, but the question remains as to how.

1.1 The Fixed-Formant Model

A sure-fire way of keeping a sound's timbre constant is to keep its spectral envelope constant by applying the original sound's spectral envelope to the transposed sound (Slawson 1968, 1985). Using such a fixed-formant structure for deriving amplitude values for a transposed spectrum can indeed yield sounds whose spectral invariance helps the brain more closely associate the original and transposed sounds. This technique is especially useful for transposing the singing and speaking voice, as the source-filter nature of the human vocal tract produces fixed formants in the spectrum for specific vowel sounds. Such a technique for spectral correction of transposed sounds has already been implemented in many systems and is more or less in common usage. Examples of this technique used successfully to correct vocal sounds can be found in the work of Schwarz (1998) and Sprenger (1999). Although this method works particularly well to create convincing transpositions of the human voice, few musical instruments have spectral envelopes with such a fixed-formant structure, due to the strong coupling of source and filter (Slawson 1985) in these instruments.

1.2 Spectral Variance and the Centroid

In addition to the fact that sounds with similar spectral envelopes will have a similar "sound color" (Slawson's redefined term for timbre), it has also

been recognized that we naturally tend to associate timbres whose spectral centroids are close to one another in frequency (Grey and Gordon 1978). If we graph the spectral centroid as a function of fundamental frequency across the playing range of several orchestral instruments, we see that the centroid remains relatively constant over a large part of the instrument's playing range. (Of course, some instruments may have several registers each with a distinct tone color, or be able to vary their timbre, and consequently the spectral centroid, via different playing techniques.)

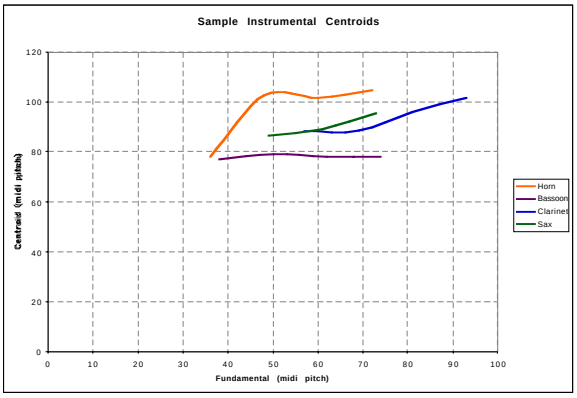

Figure 1. The spectral centroid as a function of fundamental frequency for a selection of orchestral instruments.

If we take a look at what happens to the spectral centroid when we transpose sounds crudely (without subsequently correcting their spectral envelopes), we see that the centroid actually moves directly proportionally with the fundamental.

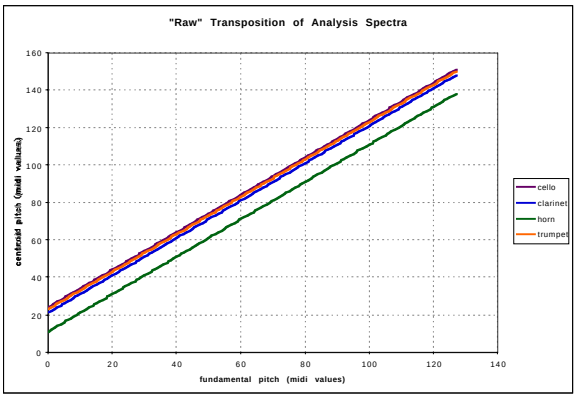

Figure 2. The spectral centroid as a function of fundamental frequency for several sets of transposed analysis data.

N.B.: For the purposes of experimentation and demonstration in this paper, Figure 2 and subsequent figures use a set of transposed analysis data derived from the spectra of several orchestral instruments (cello, clarinet, bassoon, and horn). The original (untransposed) note was D#3 (311 Hz., MIDI note 63) on all instruments, and the graph shows the data transposed (via simple scaling of partials) from C-2 (8.175 Hz., MIDI note 0) to G8 (12543 Hz., MIDI note 127).

1.3 Non-linear Frequency Scaling

By comparing the two graphs above, we notice immediately that the centroid curve is generally flatter in appearance on the graph derived from actual instrumental sounds. Our ears would also tell us the same story. If we presume that a flatter fundamental frequency to spectral centroid curve indicates a lesser degree of spectral variance, we could imagine scaling the partials or amplitude peaks non-linearly. This would result in a spectral envelope whose formant regions are not fixed, but which could be more or less invariant in different parts of the spectrum. This could potentially yield transposed sounds whose spectra or “sound color” would seem more closely related to the original sound. This technique, however, would be more analogous to frequency-shifting rather than transposing, and could result in the creation of highly inharmonic metallic-sounding variations or distortions of the original. If we use this non-linear transposition method on the same analysis data used in Figure 2, we will notice that a generally flatter centroid curve is obtained, especially for the range of extremely low transpositions.

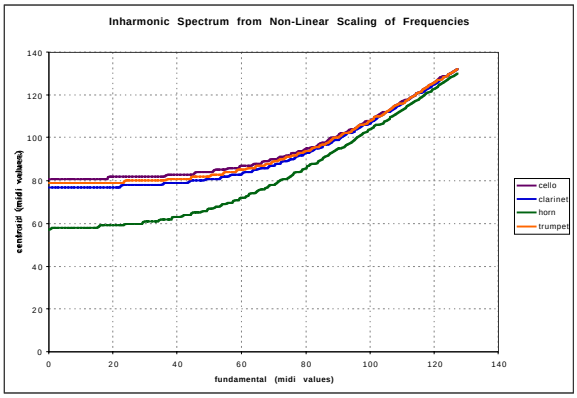

Figure 3. The spectral centroid as a function of fundamental frequency for several sets of inharmonically frequency-scaled analysis data.

Although this does indeed produce a flatter fundamental-to-centroid curve, the obvious drawback is that the original harmonic make-up of the sound has been sacrificed, or at least significantly altered.

2 A Hybrid Model

By using a hybrid transposition method which incorporates aspects of both fixed-formant transposition and non-linear frequency scaling, the author has arrived at a generalized procedure which allows for a more natural-sounding transposition of a wide variety of sounds. The goal of this method is not only to yield more realistic results than traditional spectral-domain transposition methods, but also to be simple enough to be used in a real-time context. This new transposition and spectral correction model can be referred to as “floating formant” transposition.

2.1 “Floating Formant” transposition

Floating formant transposition relies on either linear or polynomial curve interpolation between amplitude peaks in the spectral domain for generating spectral envelopes based on a given set of frequency-amplitude pairs. The curve is fitted not to peaks in the original sound's spectrum, but rather to peaks in a non-linearly frequency-shifted version of those peaks. This frequency-shifted spectrum would produce an inharmonic distortion of the original sound if it were used to resynthesize the sound, so it is used only to derive a new spectral envelope which is closely related to the original sound. This spectral envelope can be applied subsequently to the transposed sound in much the same way that a fixed-formant spectral envelope would be. Using this “floating formant” technique, we are able to create a convincing transposed spectrum whose spectral centroid is closer to that of the original sound.

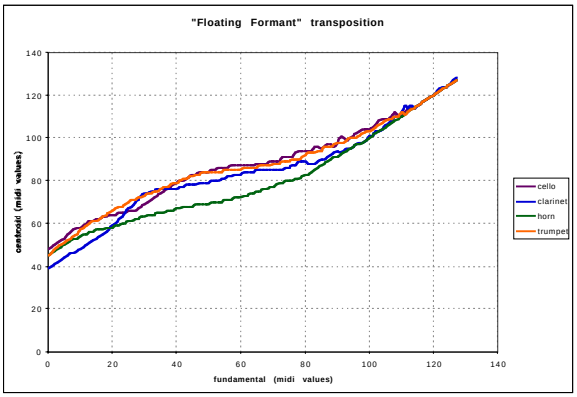

Figure 4. The spectral centroid as a function of fundamental frequency for several sets of analysis data transposed using the “floating formant” method.

2.2 Results

As we can see in Figure 4, by using the floating formant technique on the same set of analysis data as used in figures 2 and 3, although the spectral centroid varies somewhat, the transposed spectrum's centroid follows more closely that of an acoustic instrument across its pitch range. Although other transposition models might work particularly well for certain types of spectra (such as a fixed-formant model for the singing or spoken voice), the one presented here provides an reasonable generalized method for a more psychoacoustically convincing transposition of a wide variety of sounds (Dudas 1999).

3 A Real-Time Implementation

The “floating formant” transposition method is currently implemented in a real-time STFT-based pitch shifter for Max/MSP, provisionally named *gizmo~*. The STFT pitch-shifting method follows the Laroche and Dolson (1997, 1999) method, whereby ”regions of influence” around amplitude peaks in the spectrum are shifted together in order to preserve vertical phase coherence. This peak phase-locking method has been shown to yield cleaner, less ”phasy“ results for spectral-domain transposition effects.

The peaks for transposition are selected by comparison with the four neighbouring peaks (Laroche et Dolson 1999), however the spectral envelope is derived from peaks obtained from a slightly lowpass filtered version of the spectrum. This spectral-domain filtering’s purpose is twofold: 1) it helps to reduce the amount of data needed to create a spectral envelope shape by eliminating slight fluctuations in the spectrum (especially predominant in the high-frequency end of the STFT spectrum), and 2) it helps to eliminate predominant sidebands in the spectrum so they do not contribute to the shape of the spectral envelope.

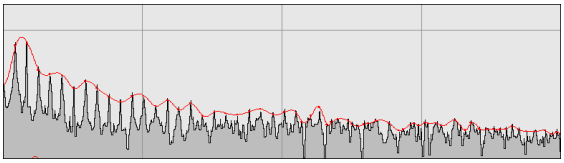

Figure 5. A GUI version of the gizmo~ object for Max/MSP which shows the peak and envelope detection, using spline interpolation to define the spectral envelope. Notice that the smaller, erroneous peaks in the spectrum are filtered out.

In order to obtain our "floating formant" envelope, the spectrum is currently rendered inharmonic by finding the maximum peak frequency of the original sound, the minimum peak frequency of the transposed sound, and scaling the peaks in-between via an exponential curve.

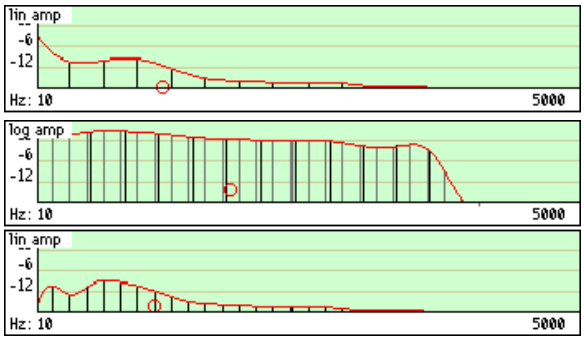

Figure 6. The original spectrum, the inharmonic spectral envelope, and the resulting transposition (one octave lower).

As we can see, once the transposed peaks are fitted to the new spectral envelope, we obtain a spectrally corrected transposition whose centroid (marked as a red circle on the graphs in Figure 6) has not been transposed directly proportionally to the amount of transposition. The shape of the spectral envelope in the transposed sound is similar to that of the original, although the formants have moved slightly with the transposition. This allows the relative intensities of the sound's partials to be kept relatively constant in the lower part of the spectrum for many sounds, which Slawson (1985) suggests may be appropriate for strongly-coupled source-filter systems. At the same time, the high end of the spectrum closely follows the fixed-formant spectral correction method, in order to retain a rich, full spectrum when transposed downward, and to attenuate frequencies when transposed upward.

4 Conclusion

Much of the aforementioned "unconvincingness" inherent in sound transposition can be corrected by modifying the spectral envelope of the resulting transposition. This has traditionally been achieved through fixed-formant transposition, which works particularly well on vocal sounds, but may not always be appropriate for other types of sounds. By taking a look at our perception musical timbre, we have formulated a way of creating a spectral envelope which can be used to correct instrumental sounds in a convincing way, without having to resort to modeling the behaviour of specific instruments. This proposal of a "floating formant" transposition method will hopefully serve as a springboard for other timbre-based spectral modifications.

5 Acknowledgments

The author would especially like to thank David Wessel, without whose Music Perception and Cognition course at UC Berkeley none of this would have come to fruition. Additional thanks also go to Cort Lippe, who complained about the horrible sound quality of my first attempt at at a STFT-based pitch-shifter, and to Jean Laroche who pointed me to the necessary papers and articles to help me get it right!

References

Dudas, R., "Generalized Transposition of Additive Analysis Data," Internal Report, CNMAT, University of California, Berkeley, 1999.

Grey, J., "An Exploration of Musical Timbre," Report STAN-M-2, Stanford University Dept. of Music, 1975.

Grey, J. and Gordon, J. W., "Perceptual Effects of Spectral Modifications on Musical Timbres," *Journal of the Acoustical Society of America*, 63(5), May 1978.

Laroche, J. and Dolson, M., "About this Phasiness Business," *Proceedings of the ICMC*, Thessaloniki, 1997.

Laroche, J. and Dolson, M., "New Phase-Vocoder Techniques for Real-Time Pitch-Shifting, Chorusing, Harmonizing and Other Exotic Audio Effects," *Journal of the Audio Engineering Society*, Vol.47 No.11, Nov 1999.

Schwarz, D., *Spectral Envelopes in Sound Analysis and Synthesis*, Diplomarbeit Nr.1622, Universität Stuttgart, Fakultät Informatik, 1998.

Slawson, W., *Sound Color*, University of California Press, Berkeley, California, 1985.

Slawson, W., "Vowel Quality and Musical Timbre as Functions of Spectrum Envelope and Fundamental Frequency," *Journal of the Acoustical Society of America*, 43, 1968.

Sprenger, S., "On the Importance of Formants in Pitch Scaling," online publication: www.dspdimension.com, 1999.