Musical Effects of Cross-Vocoding, Software Implemented

Erez Weinman, Uri Shimony, Zohar Poupko, Dalit Caspi & Moty Gerner
Laboratory of Computer Music Engineering
Department of Electrical Engineering, Technion IIT,
Technion City, Haifa, Israel
weinman@piano.technion.ac.il

Abstract

This work presents a technique of changing timbres of sounds by manipulating harmonic envelopes in spectral bands, called Cross-Vocoding. The spectrum of the source sound is divided into fairly narrow bands, and the spectral envelopes are analyzed and recorded. A new sound is then re-built using new spectral envelopes which are obtained by exchanging the original envelopes among spectral bands, according to a pre-defined exchange table. The output sound is similar to the input sound, but contains different spectral inter-relationships, altering the perceived timbre and "character" of the sound. The cross-vocoder process was tested on various sounds with various settings. Generally speaking, the output sounds were musically interesting and useful.

1 Introduction

This article describes experiments of changing timbres of recorded sounds using the new method of Cross-Vocoding, which is an extension of what is known as Channel Vocoder. The aim is to enrich the sound palette available to the musician and make it more variable and interesting. It opens a new expressive dimension while preserving the "feel" of the performer.

1.1 The Channel Vocoder

The channel vocoder is a process which gets two input sound signals, called here signal A and signal B. Spectra of both signals are then divided into n bands by using two identical filter banks. Typically, n is around 30 and the width of the bands is a third of an octave. Modulating spectral envelopes are then calculated from signal A, and a carrier is extracted from signal B, by eliminating its original modulation. The output of the channel vocoder is then obtained by modulating the carrier signal (of signal B) by the modulating envelopes (of signal A).

The Channel Vocoder is traditionally used as a speaking instrument: where a speaking vocal (signal A) is modulating a relatively static sound, with a rich spectrum (signal B). The result is a "singing voice" whose pitch behaviour is derived from the carrier input. Some additional, more interesting ideas for creative uses of channel vocoders can be found in [Anerton 1995].

2 Cross-Vocoding

2.1 Single-Input Cross Vocoding

One variant of cross vocoding has only a single-input audio signal. Spectrum of the input signal is divided into frequency bands. Modulating spectral envelopes are then calculated, and a carrier is extracted, both from the input signal. A modulating signal is then obtained by shuffling the spectral envelopes of the bands, according to a pre-defined exchange table. The output signal is built by modulation of the carrier signal by the modulating signal.

The output sound is similar to the original input sound, but contains different spectral inter-relationships, altering the perceived timbre and "character" of the original input sound.

2.2 Dual-Input Cross Vocoding

This is another variation of cross vocoding, which has dual-input audio signals, as in the channel vocoder (i.e. signal A and signal B). However, processing here is different from that of the channel vocoder as the modulating signal is created by shuffling signal A's spectral envelopes.

Details of cross-vocoding and channel-vocoding algorithms are conveyed in much greater precision in the Appendix below, using mathematical description.
3 The Experiment

Cross Vocoding processes were tested by evaluating the effects on various high quality sound sources, with various spectral divisions and various exchange-tables, like different cyclic shifting and random shuffling. Each of the filter banks had a cut-off of at least 7db.

Evaluation of the output sounds was done subjectively and in most of the cases the sounds were then tried in mosaic context. The questions which were asked for each experiment were: How does the cross-vocoded timbre sound? What type of sound sources give interesting results? How does the specific choice of filter-rank and exchange-table influence the results?

The Cross Vocoding algorithms were software - implemented and operated off-line. Fast hardware makes it possible to build real-time versions. The experiments focused only on the single-input cross vocoder. Future work is expected also on dual-input cross-vocoding.

4 Results and Discussion

4.1 Single-Input Cross Vocoding and “Noisy” Sources

Various “noisy” sound sources such as drums, cymbals, other percussion and white-noise, were tried to test the effects of single-input cross vocoding. The results were found to be very interesting and musically useful. For example, the cross-vocoded signal of a snare drum sounded like a totally different snare drum. Surprisingly, in many cases the result sounded quite “natural”, although different. A sound from a ride cymbal which was hit at its edge could be cross-vocoded into a sound which is perceived like a ride cymbal hit at its center. Of course, “unnatural” sounds could be produced too, like new inauthentic cymbals, but these still kept the performance nuances and expressiveness.

In several cases it was necessary to equalize (i.e. static chart of the spectral balance) the cross-vocoded result to achieve “best” sound.

It was found that filter banks with small number of bands, i.e. less than 5, usually yielded less interesting results, while with a number of bands higher than 15, good results could be obtained.

To sum up, the palette of tone colors and timbres which could be produced by single-input cross-vocoding, from a very limited number of noise sound sources, is huge.

4.2 Single-Input Cross Vocoding and “Tonal” Sources

The effect of single-input cross-vocoding on “tonal” sound sources was generally less impressive. In many cases, the result was dull and of lesser quality. This is probable due to the fact that quite a few bands were originally empty of spectral content and the shuffling and modulation unusual even more bands. However, in sources which were relatively rich harmonically, a suitable setting of exchange-tables could be found to give interesting new sounds.

In this case of “tonal” sources, filter banks with a small number of bands gave sometimes better results than those with larger number of bands.

4.3 Additional Potential Applications

It is possible to foresee several possible applications of cross-vocoding. Static timbre change is perhaps the most obvious application and is obtained by changing a sound source using cross-vocoding with a fixed pre-defined filter bank and exchange table.

Another application is to use real-time modulation possibilities of cross-vocoding. A performer can continuously control parameters of the cross-vocoder. For example, a drummer can use a continuous pedal to control the mix between the original sound and the cross-vocoded sound. In this way, the timbre change can be an aesthetic and expressive property.

Playing drum recorded samples, which may sound dull and static, can be made more natural. For example, in real snare drum playing, each hit sounds slightly different and this is a quality that is missing when a drum machine is used. By adding a small random amount of cross-vocoded sound to the recorded sound, the sound can be made more live and real.

Cross-vocoding can be useful also for transferring a mono sound to stereo. The original mono source can become the hard right channel, whereas the vocoded sound can become the hard left channel. Unlike some other such methods, this one is stereo compatible.

Since the attack part of a sound usually contains much spectral information, cross-vocoding can be applied to only the attack part of “tonal” sounds with good results. Changes in the attack part give the feeling of a timbre change, because the attack is of high importance in the perception of sound.
5 Appendix

5.1 Some Basic Definitions

Let $AS$ be the set of all sampled audio signals, given a sample-rate $\text{rate}$ and sampling-resolution (i.e. number of bits per sample).

$$\{0(\text{bit})\} \text{ is a sampled audio signal}

\text{Let } EN \text{ be the set of all envelope signals. An envelope signal is used to describe the amplitude envelope of an audio signal.}

\text{def}\quad EN = \{en(t)\}

\text{en(t) is in (Min\_energy; 0) \subseteq R.}

\text{Min\_energy is the energy of the quantization noise level (for example, when using 16 bit sampling).}

\text{def}\quad EN = \{en(t)\}

\text{Let } \text{env}: AS \to EN \text{ be a function that estimates the amplitude envelope of the input audio signal, by using a pre-defined method (such as RMS with square window). For convenience, env's result is measured in dB.}

\text{Let } BPF \text{ be the set of all band-pass filters:}

\text{def}\quad BPF = \{bpf | bpf is a band-pass filter\}

\text{def}\quad \text{FBF}_n = \left\{\text{FBF}_n \in AS^{\text{rate}} \text{ s.t. } \sum_{i=0}^{n-1} bpf_i(\text{as}) = \text{as} \right\}

\text{FBF}_n \text{ is a subset of the } n\text{-band filter-bank.}

5.2 Spectral Bands and Envelopes

Given a filter-bank $FB$, an audio signal $as$ and an envelope function $env$:

$$\text{Spectral bands of } as = \begin{align*}
\text{env}(as) = \frac{fb^{(n)}(as)}{\text{bpf}_n(as)}
\end{align*}

\text{Spectral envelopes of } as = \begin{align*}
\text{env}(as) = \frac{env^{(n)}(as)}{\text{bpf}_n(as)}
\end{align*}

5.3 Manipulation and Exchange Functions

$$\text{MANIP}_n \text{ def } = \begin{array}{ccc}
\text{EN} & \cdots & \text{EN} \\
\vdots & \ddots & \vdots \\
\text{EN} & \cdots & \text{EN}
\end{array}

\text{EXCHNG}_n \text{ def } = \begin{array}{ccc}
\text{en}_1 & \cdots & \text{en}_n \\
\text{en}_n & \cdots & \text{en}_1
\end{array}

\text{EXCHNG}_n \text{ is a subset of the manipulation function set. A function } \text{ex} \text{ is actually an exchange-table for spectral envelopes. Given a series of } n \text{ spectral-envelopes, } \text{ex} \text{ returns a series of } n \text{ spectral-envelopes built from the original spectral-envelope series, in a different order. The } \text{ex} \text{ function defined below is called the trivial exchange-function:}

$$\text{ex}\_\text{id} \in \text{EXCHNG}_n \text{ def } = \begin{array}{ccc}
\text{en}_1 & \cdots & \text{en}_n \\
\text{en}_n & \cdots & \text{en}_1
\end{array}
5.4 Spectral Envelope Enforcement

Let the amplify with noise-gate function \( \text{amp} \) be:

\[
\text{amp}(\text{av}(s), \text{val}) = \begin{cases} \text{0} & \text{env}(\text{av}(s)) < \text{threshold} \\ \text{av}(s) \text{ amplified by val dB} & \text{else} \end{cases}
\]

The \( \text{amp} \) function amplifies the signal by val dB, except for periods at which the signal is too-quiet. This process is called noise-gating and its goal is to prevent the signal’s noise to be amplified.

Let the spectral envelope enforcement function \( \text{sef} \) be:

\[
\text{sef}: \text{AS} \times \ldots \rightarrow \text{AS}
\]

\[
\text{sef}(s_1, \ldots, s_n) = \sum_{i=1}^{n} \text{amp}(f^{10}(a), \text{en}_{i} - \text{env}^{10}(a))
\]

The \( \text{sef} \) function “enforces” the given spectral envelopes on the given audio signal \( a \). If, however, a spectral band of \( a \) is too-quiet (probably containing only noise) at a specific period, its envelope would not be changed in that period.

5.5 Cross-Vocoder and Channel-Vocoder Definitions

Given a number of bands \( n \), a \( n \)-band filter-bank \( f \), and a manipulation function \( m \), usually an exchange-function \( e \), a dual-input Cross-Vocoder is a function \( \text{XV}: \text{AS} \times \text{AS} \rightarrow \text{AS} \) which gets two signals: signal-A \( M(s) \) and signal-B \( C(s) \), and returns a single output audio signal:

\[
\text{XV}_{n, \text{av}, \text{m}}(C(s), M(s)) = \text{sef}(C(s), \text{en}(M(s)))
\]

A Single-Input Cross-Vocoder \( \text{XV} \) is a cross-vocoder which gets only one input signal \( C(s) \):

\[
\text{XV}_{n, \text{av}, \text{m}}(C(s)) = \text{XV}_{n, \text{av}, \text{m}}(C(s), C(s))
\]

The traditional channel vocoder can be presented in the following way:

\[
\text{Ch}_{n, f}(C(s), M(s)) = \text{XV}_{n, \text{av}, \text{m}}(C(s), M(s))
\]

A schematic description of the single-input cross-vocoder is shown in figure 1.

References