The Vocoder

Thomas Carney 311107435

Digital Audio Systems, DESC9115, Semester 1 2012
Graduate Program in Audio and Acoustics
Faculty of Architecture, Design and Planning, The University of Sydney

A vocoder (short for voice encoder) is a synthesis system, which was initially developed to reproduce human speech. Vocoding is the cross synthesis of a musical instrument with voice. It was called the vocoder because it involved encoding the voice (speech analysis) and then reconstructing the voice in accordance with a code written to replicate the speech (speech synthesis).

A Brief History

The vocoder was initially conceived and developed as a communication tool in the 1930s as a means of coding speech for delivery over telephone wires. Homer Dudley [1] is recognised as one of the father's of the vocoder for his work over forty years for Bell Laboratories' in speech and telecommunications coding. The vocoder was built in an attempt to save early telephone circuit bandwidth. Theorising that by replacing the natural carrier sound of human speech with a synthesized carrier sound at a higher frequency, speech could be reproduced more clearly over long distances and low volumes, since higher frequency sounds are heard more clearly than lower ones. The fidelity of the machine was limited; the machine was intended as a research tool for compression schemes to transmit voice over copper phone lines; as such the vocoder had a prosaic speech compression goal.

Dudley's breakthrough device analysed wideband speech, converted it into slowly varying control signals, sent those over a low-band phone line, and finally transformed those signals back into the original speech, or at least a close approximation of it. The vocoder was also useful in the study of human speech, as a laboratory tool.

A key to this process was the development of a parallel band pass filter, which allowed sounds to be filtered down to a fairly specific portion of the audio spectrum by attenuating the sounds that fall above or below a certain band. By separating the signal into many bands, it could be transmitted easier and allowed for a more accurate resynthesis.

It wasn't until the late 1960s [2] and early 1970s that the vocoder was reinterpreted and used by engineers, musicians and producers in a creative way as a form of digital signal processing. The technique was used to make musical instruments 'sing'. It became widely popular in electronic and more traditional 'acoustic' forms of music, and created a signature sound all of its own[3].

How It Works

Similar to Dudley's early inventions, the modern "musical" vocoder has two input signals. One of these signals usually comes from a microphone (*the speech/modulator signal*) and another signal often

comes from a synthesized tone, which is often a square, sawtooth or perhaps pulse wave (*the synthesis signal*)[4].
The vocoder takes the two signals, and using their spectral information, creates a third signal. The goal of the vocoder is to imprint the amplitude and frequency characteristics of the speech signal onto the timbre of the synthesis signal, whilst maintaining the pitch of the speech signal.

This is achieved by implementing Dudley's [5] ground breaking (in the 1930s) band pass filter technology. The two signals travel through multiple band pass filters, each one 'tuned' to a frequency range. The band pass filter splits the incoming speech signal into a number of separate signals, each of which contains only the sound energy within the narrow pass-band of that particular filter. For instance, each pass-band might be an octave wide. Voltage (amplitude) information is also captured and this is used to create a third signal, which mirrors the previous signals. Typical vocoder designs have 8, 16, 24, or 32 bands. With fewer than 8 bands, the speech input won't be replicated accurately enough for us to understand the output. Conversely, using too many bands can reduce the personality of a vocoder by glossing over its characteristic distortion.

The vocoder works on the principle of formants [6]. When the human voice speaks, formants are the distinguishable or meaningful components of speech. These formants are produced by resonances in the vocal tract and allow the human brain to distinguish sounds and thus speech from these tones. The vocoder's ability to distinguish these formants gives it the ability to reproduce speech-like sound.

Types of Vocoders

The channel vocoder [7] is the most similar to Dudley's initial vocoder. It is a two stage process. Source A (speech signal) travels through fixed frequency band pass filters. The output of each filter is connected to an envelope detector, which determines the amplitude of the signal. Source B's signal travels through band pass filters identical to Source A, and then onto voltage controlled amplifiers (VCA). These VCAs are controlled by Source A's envelope detectors (Source B's signals will have the same amplitude as their corresponding Source A signal). The signals from Source B's filters are then combined to form the output signal *[see Figure 1].*

Linear Predictive Coding (LPC)[8] has been extensively used in speech and music applications. The basic idea behind LPC analysis is that a speech sample can be approximated as a linear combination of past speech samples. LPC is a process whereby future values are estimated by the system.

In it's most basic sense it analyses the two previous samples (and their slope of difference) and predicts what the following outcome may be. A system can be more accurate by using more samples for its predictions. It must be known that as a predictive system can inherently be errornous, a degree of error is accounted for in the initial calculations, hopefully to minimise the impact of any errors on the outcome.

The LPC is an all pole filter, which is a reasonable approximation to many sounds uttered by the human voice

and certain musical instruments (*however, it doesn't work on every instrument*). An all-pole filter has a frequency response function that goes infinite (poles) at specific frequencies, but there are no frequencies where the response function is zero. The inverse of an all pole filter is an all zero filter.

LPC analysis involves four directions:

1. Spectrum (formant) analysis.
2. Pitch analysis
3. Amplitude analysis
4. Whether voice or unvoiced

Voiced or unvoiced determines whether the sound will be pitched or not in resynthesis. A voiced sound is one in which the vocal cords vibrate. *Voiced* sounds such as a,e,I,o,u have a pitch which is determined by the vocoder whereas sibilance (s,z) or explosive (t,p) are *unvoiced* and don't excite the vocoder as much.

Being voiced determines that the signal will go through pitch detection and pulse generation. The unvoiced signals the system uses a noise generator to simulate resonance of the vocal tract.

The LPC then performs a synthesi*s [see Figure 2]*, using all the information gathered in the analysis. As the LPC has collected significant data on the input wave, it is possible to invert that data and resynthesize the sound. The LPC method works reasonably well in approximating speech on some instruments, but is not universal for all sounds. As a result it leaves artificiality in the synthesized sounds, which gives it the vocoder sound effect.

In the modern era, the vocoder can be implemented in the digital domain using computers. The vocoder is still implemented in a three stage system, almost identical to the analogue method: Analysis, Transformation and (Re)Synthesis.

Analysis

The analysis is generally performed by a Fourier transform, the entire process being carried out automatically and almost instantaneously. Fourier theory states that any wave form can be analysed into a collection of pure tones in a harmonic series (Fourier analysis). Conversely, that collection of pure tones can be added together to produce the original periodic wave (Fourier synthesis).

The Fourier transform analyses an input signal for amplitude, phase and frequency information over time and converts the analogue signal into a digital one, which can be manipulated by digital signal processors. The Fourier transform breaks the signal down into a series of sine and cosines which can each be individually 'manipulated'/analysed by a DSP. The Fourier transform allows a relatively difficult signal (analogue audio) to be simplified into a digital signal (1s and 0s), which allows much easier manipulation of the signal.

Transformation

The transformation occurs when the information from one signal is imposed on another signal. The most important aspect of the vocoder is it's ability to make a synthesized sound mimic a real world sound. It is through this 'transformation' that it is possible.

(Re)Synthesis

This step involves the inverse Fourier transfer mentioned earlier to convert the digital signal, which is now a composition of our two input signals,

back into an analogue signal, which can be played back and heard.

The vocoder has been used for voice synthesis for eighty years. It has found uses in music, other media industries such as film, television and games, telecommunications, defence, encryption and many other fields. The technology has been used in many other synthesiser applications and is quite important to the development of electronic music. Something Dudley surely could never have conceived eight years ago.

---

[1] Lawrence J. Raphael, Gloria J. Borden, Katherine S. Harris *Speech Science Primer: Physiology, Acoustics, And Perception of Speech* Lippincott Williams & Wilkins 2006 pp. 23

[2] Robert Moog "Voltage controlled electronic music modules" *J. of Audio Engineering Society July 1965 Volume 13 Number 3*

[3] Harald Bode (October 1984) "History of Electronic Sound Modification". *J. Of Audio Engineering Society* **32** (10): 730–739

[4] Jim Alkin March 29 2006 http://digitalmedia.oreilly.com/pub/a/oreilly/digitalmedia/2006/03/29/vocoder-tutorial-and-tips.html

[5] Patent US 2121142 Homer W. Dudley 1938 http://www.google.com/patents/US2121142

[6] Dowd, A., Smith, J.R. and Wolfe, J. (1997) *Learning to pronounce vowel sounds in a foreign language using acoustic measurements of the vocal tract as feedback in real time* Language and Speech, **41**, 1-20.

[7] William A. Sethares Channel Vocoder http://sethares.engr.wisc.edu/vocoders/channelvocoder.html

[8] Curtis Roads *The Computer Music Tutorial* February 1996 Short