THE MODFM SYNTHESIS VOCODER

Victor Lazzarini, Joseph Timoney

Sound and Digital Music Technology Group
National University of Ireland, Maynooth, Ireland
victor.lazzarini@nuim.ie
jtimoneyi@cs.nuim.ie

ABSTRACT

Vocoders have been practical tools for synthesis and processing since the original voice encoder by Homer Dudley in the 1930s . They enable the transformation and reproduction of an input sound by the manipulation of its spectral envelope. This paper presents a new variation on the principle utilising the technique of Modified FM (ModFM) synthesis for sound generation. The general design of the vocoder is introduced, followed by a non-mathematical introduction to the ModFM synthesis algorithm. The article is completed with an outline of applications and a discussion of some basic examples of the vocoder operation.

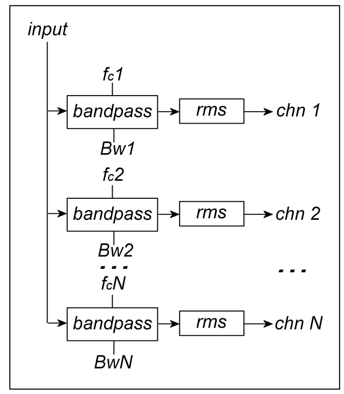

Figure 1. The channel vocoder analysis stage.

1. INTRODUCTION

As is the case with many of today's electronic musical instruments, the vocoder was originally developed by the telecommunications industry as means of coding speech signals [6]. It is seen as the first breakthrough in the effort to reproduce speech using electronic means. Various designs of vocoders have been proposed along the years, including the use of linear prediction [1] and spectral analysis (e.g. the Phase Vocoder [7], which has become a more general-purpose analysis-synthesis method).

In this work, we will be concentrating on the *channel* vocoder. Since the 1970s, little research work has been published on the channel vocoder with only a few new papers appearing, such as [8] and [4]. The work in [8] suggested some new speech processing applications that would be possible using a digital version of the vocoder, while [4] looked at implementing the analysis/synthesis procedure of the vocoder algorithm using wavelets. This is curious in that although it is a widely used device very little innovation has occurred since its original design over 70 years ago.

For speech processing, the vocoder operates by first carrying out a sub-band analysis of signal power or amplitude (fig.1). The data obtained by this analysis can then be used to resynthesise the signal by its application to matched sub-band synthesis channels (fig.2). The excitation source for the synthesis component is generally based on broad bandwidth oscillators. As far as a communication system is concerned, this arrangement performs reasonably well for vocal signals of pitched nature (the so-called *voiced* signals), mostly represented by vowels and semivowels. The use of an added noise generator for non-pitched sounds, represented by some of the consonants, can also improve the overall quality of the reproduced sound. The development of alternative methods of speech spectrum analysis led to the channel vocoder being superseded as an engineering tool, however it did find a distinct place in electronic music. A well-known early example of its use in popular music is found in Wendy Carlos' soundtrack to the 1971 film *A Clockwork Orange.* The vocoder still features as an instrument in various types of electronic and pop music. Its ability to blend vocal and instrumental sounds (which take the place of the excitation oscillators) into something of an other-worldly quality has meant that instrument is still used in various types of electronic and pop music.

This indicates us that, in terms of a musical system, the vocoder has a lot to offer. As we are not always concerned with the ability to communicate a given speech signal, but with opportunities to manipulate and transform a source sound in a creative way, there is now plenty of scope for work. Depending on how we store the analysis signal, it is possible to apply timescale and pitch changes to the sound, to modify vocal formants, to create transitions between distinct sounds and to cross-synthesise two or more input signals. The design of the channel vocoder is simple and open enough to allow for variations and improvements. In this paper, we will describe how we married this basic design with a method of formant synthesis based on the Modified Frequency Modulation (ModFM) algorithm [10] to give rise to yet another variation with different-sounding characteristics.

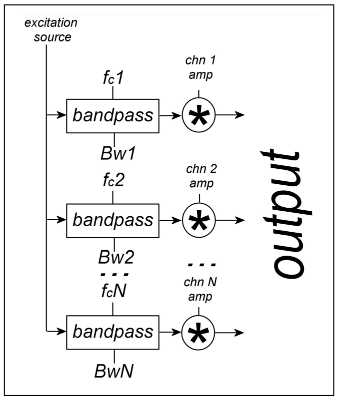

Figure 2. The channel vocoder synthesis stage.

1.1. ModFM synthesis

The technique of ModFM synthesis is a non-linear distortion method based on frequency (or phase) modulation [5]. It uses a complex index of modulation (which is purely imaginary), generating sidebands that are scaled by modified Bessel coefficients. The following discussion is intentionally non-mathematical, but the interested reader is referred to our paper on the ModFM theory[10], where an extensive discussion of its technical principles is found. The phase-synchronous variation of this technique has been shown to be useful in the synthesis of formants in [9]. Based on these ideas, a ModFM operator can be designed to emulate a region of resonance in the spectrum. Its flowchart is shown on fig.3.

As it can be seen, these operators are based on four table lookups and a few multiplications. Each one takes as parameters the centre frequency f_c, fundamental f_0 and bandwidth Bw (or alternatively, a Q factor as the ratio $f_c : Bw$). From these, the internal variables seen in the fig.3 flowchart are calculated (for details see [10]). They can synthesise a number of harmonics within the given bandwidth around f_c (fig.4), with a fundamental frequency based on a common input phase signal (obtained from a simple modulo phase source). As all the harmonics generated are always in cosine phase, there are no phase interaction issues regarding the mixing of various parallel ModFM operators.

2. THE MODFM VOCODER

The design of the ModFM vocoder is shown on fig.5. It is based on a sub-band analysis of an input signal using a bank of parallel filters, from which the RMS amplitudes of each channel are extracted on a sample-by-sample basis (via a standard RMS detection method). Channels are logarithmically spaced, covering a user-defined range and with user-defined Q factors. In addition, we obtain the fundamental frequency and overall amplitude of the input sound (pitch and envelope detection).

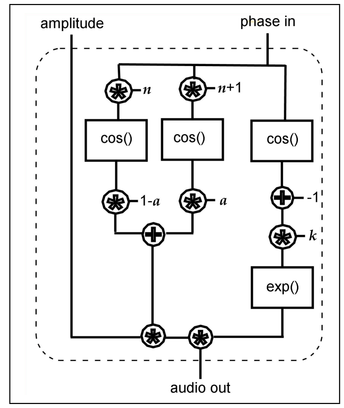

Figure 3. A ModFM synthesis operator. It takes an amplitude and a modulus phase as input, producing an audio signal as output. Its parameters a, n and k are dependent on the centre frequency f_c, fundamental f_0 and bandwidth Bw [10].

These elements comprise the analysis data that is then applied to a bank of ModFM operators, one for each subband or channel. ModFM bandwidths are set to those of the filter bank and each operator centre frequency is tuned to a matching filter. The analysed pitch and overall amplitude are used, respectivelly, to control the fundamental frequency of all operators and to impose an envelope to the synthesis mix.

In general, the arrangement of channels in the ModFM vocoder follows a basic principle whereby N centre frequencies are spaced logarithmically over a given range. The number of channels N, and the difference between the highest and lowest frequencies in that range define the actual centre-frequency spacing and the filter bandwidth then can be either automatically defined by this spacing or independently set according to a given Q factor. The latter option is probably the one that offers a better flexibility and greater range of musical effects. In general, depending on the application, the typical number of channels will range between 12 and 30, with bigger filterbanks used for finer spectral modelling.

The use of ModFM synthesis here gives the vocoder a distinct sound, which provides an interesting variation to the usual filterbank synthesis of fig.2. The characteristics of ModFM synthesis (as discussed, for instance, in [9]) will impart a distinct tone colour to the instrument output.

2.1. Modes of operation

The vocoder can operate with realtime input signals or as a resynthesis tool. In the first scenario, the analysis data is streamed directly to the synthesis bank, with possible realtime modifiers inserted in between the two stages. This is the typical operation of the vocoder and a number of applications are possible, which will be discussed in the following sections of this paper. As a more general-purpose analysis-synthesis tool, we propose the use of the ModFM vocoder in a two-step set-up. The analysis filterbank and the pitch and amplitude detection, in this case, are per-

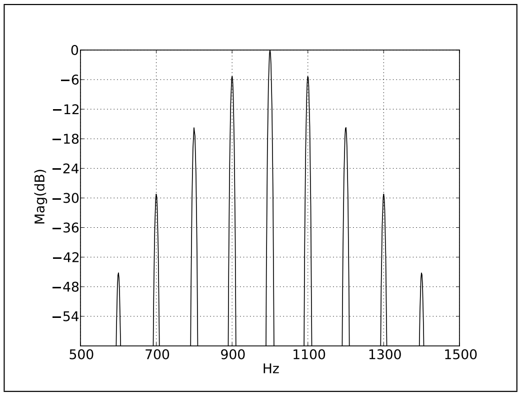

Figure 4. The output spectrum of a ModFM synthesis operator, centred at $1KHz$, with $Q = 5$ and $f_0 = 100Hz$

formed separately, offline, and the data stored for later re-synthesis using the ModFM synthesiser. This method produces similar results to the one described in [9], where a Kalman filter analysis is employed. However, as shown in fig.5, here we have only tracks of amplitudes for each channel, as well as pitch and amplitude envelope, whereas in that case, we would have tracked formant centre frequency and bandwidth as well (which are fixed in the ModFM vocoder). This allows for some additional transformation possibilities (e.g. timescale modifications) not possible in the realtime processing scenario.

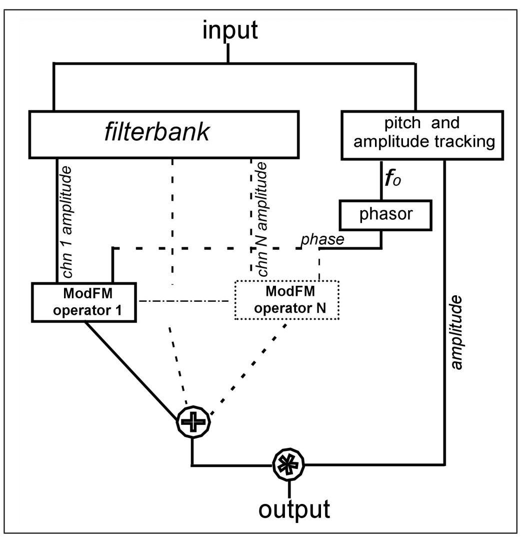

Figure 5. The ModFM vocoder flowchart.

3. APPLICATION EXAMPLES

3.1. Pitch effects

The most typical use of a channel vocoder is to reproduce vocal sounds with the substitution of the original pitch, which can lead to mechanical or robotic voice impressions. In fact, along these lines, a number of effects can be created.

One interesting application is in vocal tuning correction, where the detected pitch is modified to fit a given scale in 12-tone equal temperament (12TET; other tunings are also possible, but we use this one as an example here). A diagram describing this scheme is shown on fig.6. The output of the pitch tracker is first converted into a 12TET integral pitch value (by rounding), then into a mod12 pitch class. The pitch class number is compared to a table containing a scale. If the detected pitch matches a scale pitch class, it is taken as the target pitch of the note. If the pitch is not found on the scale, we check whether the rounding was done upwards or downwards. Depending on the direction of the rounding, we will increase or decrease the pitch by a semitone (in the case of the typical 7-note diatonic scale). The time response of the algorithm can be controlled by a simple first-order IIR lowpass filter [3] placed in the pitch control signal path. Various effects typical of electronic dance music can be performed with this design.

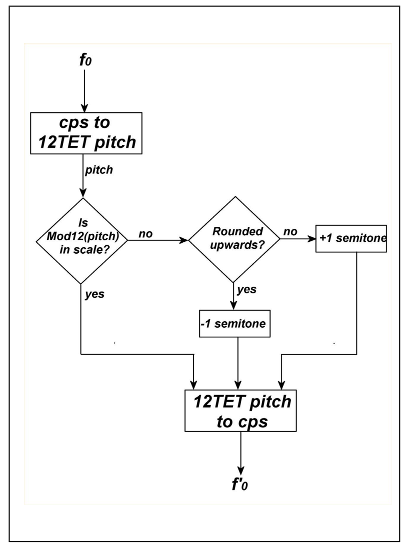

Figure 6. Pitch correction algorithm.

Another useful application, based on the above principles is a harmoniser. By defining a transposition interval, the detected pitch can be scaled to provide a harmony voice that can be mixed to the original signal. This effect can be also be complemented by a similar process to the pitch correction algorithm above to provide harmonies that are based on a given scale. It is important to note that while these processes are typically applied to vocal sounds, they can be used with any input sound (provided that it is monophonic and pitched).

Finally, by substituting the original pitch, the ModFM vocoder can be used as a model-based synthesiser, based on the analysis of instrumental samples. Channel data and envelopes can be obtained from various sources, stored and then reproduced with variable pitch. This data can be significantly decimated in time, so that a degree of data reduction can be obtained. This type of synthesis can be combined with spectral envelope transformations, such as the ones described in the following sections, for various musically interesting effects.

3.2. Spectral envelope effects

Given that the core of the ModFM vocoder operation is based on the analysis of the spectral envelope of an input, some transformations can be applied to this data. For instance, transitions between two or more spectral envelopes can be effected by interpolation of channel amplitudes. In the offline analysis mode, this is simply done by utilising more than one set of channel tracks and performing linear interpolation on them. For realtime processing operation, more than one stream of analysis data can be produced by separate filterbanks. Synthesis is performed in the usual way from the interpolated channel data. Pitch and overall amplitude data can be also interpolated or not, depending on the type of effect desired.

Another interesting effect in this category is spectral envelope warping. This is done by breaking the match between analysis and synthesis centre frequencies, either by a combination of scaling and offsetting of the synthesis channels or by some other non-linear function applied to this parameter. A typical result for vocal re-synthesis is that of gender/age changing and 'donald duck' effects caused by the shifting of formants. However, spectral warping is not limited to these applications, as it can be used in a variety of timbre manipulations of instrumental tones.

Finally, it is worth mentioning that filtering effects can be implemented by applying scaling or masking functions to the channel amplitude data. For instance, a band-reject filter can be created by the application of a mask that zeros the amplitude of one or more channels. Notch filters can be effect by scaling the channel amplitudes by functions with maxima at the frequencies of interest.

3.3. Timescale modifications

As noted before, the offline operation mode also allows for timestretching and compressing of an analysed sound. This is, of course, an operation that is independent of frequency. The quality of timestretching will largely depend on how much decimation has been applied to the channel track data, as the interpolation artifacts will be more noticeable when less analysis points are available. In general, the ModFM vocoder is capable of musically interesting timescale effects.

4. CONCLUSIONS

This paper presented a new formulation of the traditional channel vocoder with the use of ModFM as its synthesis algorithm. We introduced the principles of operation of the vocoder and the technique of ModFM synthesis. The design of the ModFM vocoder was laid out in some detail, where it was shown to be based on a filterbank, pitch and amplitude analysis stage followed by a series of parallel operators tuned to different centre frequencies and sharing a common fundamental. The design was shown to have two basic modes of operation: realtime processing, where the analysis and synthesis as performed together on a streaming basis; and offline, where the analysis is performed before hand and stored channel, pitch and amplitude data are used to synthesise a sound on a separate stage. In general, the ModFM vocoder can be classed as an adaptive synthesis and processing instrument. The paper concluded with a tour of various applications, divided into three categories of pitch, spectral envelope and timescale effects. Typical applications of vocal correction, spectral envelope and timestretching were discussed.

A prototype of the ModFM vocoder has been implemented in the Csound language [2]. Sound examples and application code can be found in

http://music.nuim.ie/synthesis

5. ACKNOWLEDGEMENTS

The authors would like to thank Brian Carty for the original idea on which the pitch correction algorithm is based.

6. REFERENCES

[1] B. Atal and S. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," *J. Acoust. Soc. of Am.*, vol. 50, pp. 637–655, 1971.

[2] R. Boulanger, Ed., *The Csound Book*. Cambridge, Mass: MIT Press, 2000.

[3] R. Boulanger and V. Lazzarini, Eds., *The Audio Programming Book*. Cambridge, Mass: MIT Press, 2010.

[4] C. D. S. Castaneda, "A channel vocoder using wavelet packets," in *124th AES Convention*, Amsterdam, 2008.

[5] J. Chowning, "The synthesis of complex spectra by means of frequency modulation," *Journal of the AES*, vol. 21, no. 7, pp. 526–534, 1973.

[6] H. Dudley, "The vocoder," *J. Acoust. Soc. Am.*, vol. 11, no. 2, p. 169, 1939.

[7] F. Flanagan and R. Golden, "Phase vocoder," *Bell System Technical Journal*, vol. 45, pp. 1493–1509, 1966.

[8] B. P. Gold B. and M. R., "New applications of channel vocoders," *IEEE Trans. on ASSP*, vol. 29, no. 1, pp. 13–23, 1981.

[9] V. Lazzarini and J. Timoney, "New methods of formant analysis synthesis for musical applications," in *Proceedings of the International Computer Music Conference*, Montreal, Canada, 2009.

[10] ——, "The theory and practice of modified frequency modulation synthesis," *Journal of the AES*, vol. 58, no. 6, pp. 459–471, 2010.