NEW METHODS OF FORMANT ANALYSIS-SYNTHESIS FOR MUSICAL APPLICATIONS

Victor Lazzarini and Joseph Timoney

Grúpa Theicneolaíocht Fuaime agus Ceoil Dhigitigh
National University of Ireland,
Maynooth, Ireland
Victor.Lazzarini@nuim.ie, JTimoney@cs.nuim.ie

ABSTRACT

In this article we introduce some novel methods of formant analysis and synthesis. The latter is performed by an implementation of a technique known as Modified FM synthesis, while the former uses an Unscented Kalman Filter algorithm. The proposed techniques provide a good alternative to the existing time and frequency domain methods that is both flexible and elegant. The synthesis technique, based on an extension of FM, allows for an efficient means of generating an easily-describable spectrum. The Kalman Filter analysis provides precise and finely-grained tracking of formant parameters. After a detailed discussion of the methods, we present a number of possible musical applications for them. Sound examples are provided online, illustrating the techniques and their use.

1. INTRODUCTION

The synthesis and manipulation of vocal sounds has been one of the most important areas of sonic design in electroacoustic music. In the signal processing literature, we find a substantial body of research into this subset of audio signals. Among the most important techniques for musical applications, we find analysis-based subtractive synthesis[7] and Linear Prediction Coding (LPC)[1]; direct synthesis algorithms such as VOSIM[3] and Phase-Aligned Formant (PAF)[5] synthesis; and frequency-domain methods, such as SMS[9].

Given that the characteristics of vocal sounds are strongly linked to the occurrence of formants, in this work we will concentrate on the analysis and emulation of such features. We propose an alternative method of direct synthesis, which is complemented by a novel approach to the tracking of formant centre frequencies and bandwidths. The synthesis technique is based on the Modified FM (ModFM) algorithm, which has already been introduced and discussed in relation to Virtual Analogue oscillators[4]. The formant analysis is implemented using an Unscented Kalman filter[13], as an evolution of the Extended Kalman filter technique explored in [12].

2. FORMANT SYNTHESIS

ModFM has its origins on Classic FM[2] synthesis, which we can express, in terms of cosines, as

$$s_{FM}(t) = \Re\{e^{i\omega_c + iz\cos(\omega_m)}\} = \cos(\omega_c + z\cos(\omega_m)) =$$
$$= J_o(z)\cos(\omega_c) +$$
$$\sum_{n=1}^{\infty}\left[(-1)^{\text{int}(\frac{n}{2})}J_n(z)\left(\begin{array}{c}\cos(\omega_c - n\omega_m) \\ + (-1)^n\cos(\omega_c + n\omega_m)\end{array}\right)\right] \quad (1)$$

where $J_n(z)$ stands for the Bessel function of the first kind of order n, ω_c and ω_m are the carrier and modulator frequencies $2\pi f_c t/sr$ and $2\pi f_m t/sr$ respectively (with f_c and f_m in Hz and sr as the sampling rate).

In this case, a change of variable $z = -ik$ will give us the ModFM expression:

$$s_{ModFM}(t) = \Re\{e^{i\omega_c + k\cos(\omega_m)}\} = e^{k\cos(\omega_m)}\cos(\omega_c) =$$
$$= I_o(k)\cos(\omega_c) +$$
$$\sum_{n=1}^{\infty}I_n(k)\left(\cos(\omega_c - n\omega_m) + \cos(\omega_c + n\omega_m)\right) \quad (2)$$

As it can be seen, from Eq.2, we now have a spectrum that is scaled by *modified* Bessel functions of different orders, $I_n(k) = i^n J_n(ik)$. They are purely imaginary-argument versions of the original Bessels[15], with useful musical qualities (for details see[4]). ModFM, with proper normalisation is defined as:

$$s(t) = e^{(k\cos(\omega_m) - k)}\cos(\omega_c) =$$
$$= \frac{1}{e^k}\sum_{n=-\infty}^{\infty}I_n(k)\cos(\omega_c + k\omega_m) \quad (3)$$

In order to synthesise a formant with ModFM, the carrier frequency f_c is made to be an integer multiple of the modulator frequency f_m close to the centre of the formant f_f that we want to reproduce:

$$f_c = nf_m = \text{int}(\frac{f_f}{f_m})f_m \quad (4)$$

Here, we will not want to limit the formant frequency to be an integer multiple of the fundamental frequency (which in

this case is also the modulator one). Instead, we will use a pair of carriers, tuned to adjacent harmonics, which lie each side of the true formant centre. We can then interpolate the outputs of these carriers to produce the correct signal. So if n is int$[f_f/f_0]$ and ω_0 $(= 2\pi f_0 t)$ is the fundamental [radian] frequency, we have:

$$s(t) = e^{(k\cos(\omega_0)-k)} \times$$
$$\left[(1-a)\cos(n\omega_0) + a\cos([n+1]\omega_0)\right] \quad (5)$$

$$a = \frac{f_f}{f_0} - n \quad (6)$$

The expression defined in Eq.5 will then be used to recreate one single formant peak. Figure 1 illustrates this, for a very narrow spectral peak centred at 500Hz.

In ModFM, formant bandwidth is determined by the index of modulation k. Following the example of[5], we can define an expression for k in terms of an intermediary variable γ:

$$k = \frac{2\gamma}{(1-\gamma)^2} \quad (7)$$

The value of γ in its turn can be approximated as a function of the fundamental and bandwidth:

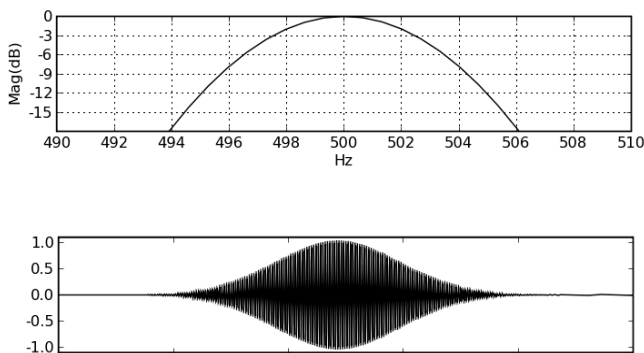$$\gamma \approx 2^{-\frac{f_0}{0.29B}} \quad (8)$$

Figure 1. Spectral envelope of ModFM-based formant and a detail of its waveform

3. ANALYSIS

Complementing our synthesis technique, we propose a means of audio signal analysis that provides the parameters needed to drive our model. The flowchart for our analysis-synthesis scheme is shown on Fig.2. A number of ModFM operators based on eq.3 take in fundamental, centre frequency, bandwidth and amplitude parameters from the analysis process.

3.1. Formant Tracking

The speech signal processing literature shows that there is a wide variety of approaches for the tracking of formants in signals. The most well-known and frequently employed techniques are derived from a frame-based LPC coefficient analysis of the signal [1].

However, the major drawback of any frame-based technique is that continuity of the formant estimates across frames must be imposed [14], normally achieved using some heuristic measures. An alternative to a frame-based analysis is to use a technique that will track the formants on a sample-by-sample basis, whereby continuity should be ensured due to the proximity of the values. This has motivated the application of the well-known Kalman filter, in its extended form, to formant tracking [14], [12]. The Extended Kalman filter (EKF) is a variant that is applied to problems where non-linear relationships exist. In the case of formant tracking, the formants and their respective bandwidths are related to the sound signal by combining them into an all-pole filter description of the signal.
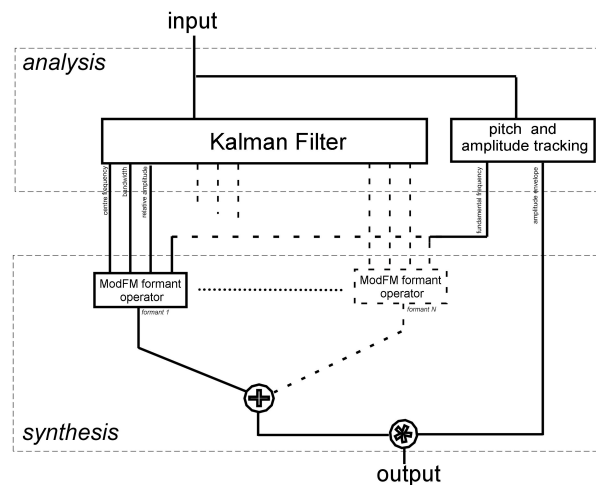

Figure 2. Analysis-synthesis flowchart

The design of an EKF solution relies on the linearisation of the problem using a first-order Taylor series approximation, involving the computation of a Jacobian matrix of partial derivatives. However, if the relationships are highly non-linear the EKF can perform poorly. An alternative approach is to use the Unscented Kalman filter (UKF) [13]. The UKF works under the premise that it is easier to approximate a Gaussian distribution than to approximate an arbitrary nonlinear function [11], and thus produce a more accurate estimation of the variables of interest. This approach also obviates the need to compute Jacobians which is an added benefit [13].

The UKF uses a non-linear state space description of the system as given by

$$\mathbf{x}_{k+1} = \mathbf{f}(k, \mathbf{x}_k) + \mathbf{v}_k \qquad (9)$$

$$\mathbf{y}_k = \mathbf{h}(k, \mathbf{x}_k) + \mathbf{n}_k \qquad (10)$$

where $\mathbf{y_k}$ is the observed output; $\mathbf{x_k}$ is the state; $\mathbf{f}(.)$ and $\mathbf{h}(.)$ describe non-linear, and possibly time-varying, state transition and measurement matrices, respectively; $\mathbf{v_k}$ and $\mathbf{n_k}$ are independent zero-mean white Gaussian noise processes with covariance matrices $\mathbf{P_v}$ and $\mathbf{P_n}$ respectively.

In the formant tracking problem the state vector is assumed a set of m formants frequencies and their respective bandwidths [19]:

$$\mathbf{x} = [F_1, F_2, \ldots, F_m, B_1, B_2, \ldots, B_m]^T \qquad (11)$$

It is assumed that the dynamic model for state update is linear, that is the formant values and bandwidths at time $k+1$ are equal to the current values plus some deviation $\mathbf{v_k}$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{v}_k \qquad (12)$$

Thus, the state transition function $\mathbf{f}(.)$ is given by the identity matrix $\mathbf{I}$. It is assumed that the sound signal is produced by an all-pole model of order n, with coefficients $a_1, \ldots a_n$ giving an output value

$$y(k) = -a_1 y(k-1) - \cdots - a_n y(k-n) + n_k \qquad (13)$$

Letting

$$\mathbf{S} = [y(k-1) \, y(k-2) \ldots y(k-n)]^T \qquad (14)$$

$$\text{and} \quad \mathbf{a} = [a_1 \, a_2 \ldots a_n]^T \qquad (15)$$

means that Eq.14 can be rewritten as

$$y(k) = -\mathbf{S}^T \mathbf{a} \qquad (16)$$

The transfer function of the all-pole model can be expressed as the cascade of $m=n/2$ resonators, each one representing a single formant resonance,

$$\frac{1}{1 + a_1 z^{-1} + \cdots + a_n z^{-n}} = \left(\frac{1}{1 + c_1 z^{-1} + d_1 z^{-2}}\right) \cdots \left(\frac{1}{1 + c_m z^{-1} + d_m z^{-2}}\right) \qquad (17)$$

The formants and bandwidths are contained in the resonator coefficients and are given by

$$c_j = -2 e^{-\pi B_j T} \cos(2\pi F_j T) \qquad (18)$$

$$d_j = e^{-2\pi B_j T} \qquad (19)$$

Furthermore, the amplitude A_j of each formant can then be found by first defining

$$|H_j(\omega)| = \frac{1}{\sqrt{1 + c_j^2 + d_j^2 + 2\cos(\omega)(c_j + c_j d_j) + 2\cos(2\omega)(d_j)}} \qquad (20)$$

giving,

$$A_j = \prod_{l=1}^{m} |H_l(2\pi F_j T)| \qquad (21)$$

Figure 3 shows the tracking of 4 formants though a sequence of vowels 'aeiou'. The spectrogram of the sound is displayed in grayscale where high magnitude values are indicated in white. The formants tracks are overlaid using black lines. Before tracking the higher frequency formants were enhanced by passing the signal using a simple one zero high pass filter.
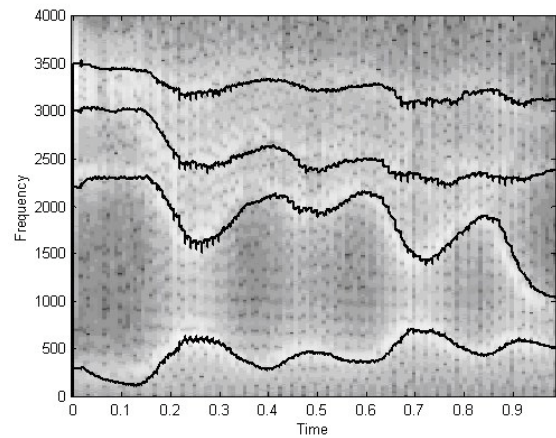

Figure 3. Tracking of formants in a sequence of vowels.

3.2. Pitch and Envelope

In addition to the tracking of formant parameters, we also include optional pitch and amplitude envelope estimation, which complements the analysis model. The added information on the fundamental frequency allows a more realistic resynthesis. The amplitude envelope is also useful in reinforcing the shape of the original signal.

4. MUSICAL APPLICATIONS

The proposed analysis and synthesis methods have a number of useful applications. Sound examples illustrating the techniques discussed here can be found at http://music.nuim.ie/synthesis

Time and frequency-scale modifications: Given that our synthesis is driven by independent parameters, we can easily change the time and frequency scales of an analysed signal separately. Simply scaling the pitch-tracked

fundamental is enough to realise, for instance, a harmoniser effect. Time-scale modifications can be performed by altering the playback rate of the analysed data.

Formant transformation: Instead of modifying the pitch and timescale while keeping formants intact, we can do the opposite. A number of transformations can be applied to formants: non-linear scaling of formant frequencies; bandwidth narrowing/widening; amplitude/weight modifications; etc. These, for instance, can be used for gender change, as the tonal colour of male and female (as well as children's) voices depend significantly on the formant structure. It is also possible to make transitions between various levels of natural and synthetic sounding resyntheses.

Cross synthesis and morphing: Finally, combining analyses from different source sounds can be used in various ways for cross-synthesis of spectra. For instance, we might apply a pitch and amplitude track from one origin, say an instrumental melody, to the formants of another provenance, say a spoken set of vowels. Also, it is possible to use data interpolation to create hybrid sounds from two sources, morphing from one spectral type to another.

In addition to the above uses, since the analysis and synthesis stages are independent, it is possible to employ them separately. For instance, we may choose to emulate a choral sound by employing the ModFM formant operators with generalised parameters from well known vowel estimates. Or, conversely, we might employ the analysis stage to generate material for other purposes than direct synthesis.

5. CONCLUSION

In this article we have presented some novel methods of formant analysis and synthesis. The sound generation technique, based on Modified FM, was discussed in detail and its key features highlighted. We presented a method of formant tracking based on Unscented Kalman Filters, which in addition to amplitude and pitch estimation, can be used to feed our synthesis model. The method has been shown to have various musical applications, which are ultimately the goal of this research.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1]Atal, B and Hanauer, S. "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave." *Journal of the Acoustical Society of America*, 50, pp.637-655, 1971.

[2]Chowning, J. "The Synthesis of Complex Audio Spectra by Means of Frequency Modulation". *Journal of the Audio Engineering Society*, 21, pp. 526-34, 1973.

[3]Kaegi, W. and Tempelaars, S, "VOSIM – A New Sound Synthesis System", *Journal of the Audio Engineering Society*, 26(6), pp. 418-25, 1978.

[4] Lazzarini, V, Timoney, J. and Lysaght, T. "A Modified FM approach to bandlimited signal generation". *Proceedings of the 11th International Conference on Digital Audio Effects*, Espoo, Finland, 2008.

[5] Puckette, M. "Formant-Based Audio Synthesis Using Nonlinear Distortion." *Journal of the Audio Engineering Society*, 43(1), pp. 40-47, 1995.

[6] Rodet, X. "Time Domain Formant-Wave-Function Synthesis". *Comp. Music Journal*, 8 (3), pp. 9-14, 1984.

[7] Rabiner, L . "Digital Formant Synthesizer for Speech Synthesis Studies". *Journal of the Acoustical Society of America*, 43, pp.822-828, 1968.

[8] Rigoll, G., "A new algorithm for estimation of formant trajectories directly from the speech signal based on an extended Kalman filter", ICASSP 86, Tokyo, Japan, 1986.

[9] Serra, X. "Musical Sound Modelling with Sinusoids plus Noise". in: G.D. Poli et al (eds.), *Musical Signal Processing*, Swets & Zeitlinger, Amsterdam. 1997.

[10] Snell, R.C., and Milinazzo, F., "Formant location from LPC analysis data", IEEE Trans. on Speech and Audio Proc., vol. 1, no. 2, April 1993, pp. 129-139.

[11] Stastny, N.A., Bettinger, R.A., and Chaveze, F. R., "Comparison of the Extended and Unscented Kalman Filters for angles based relative navigation", Proc. of AIAA/AAS Astrodynamics specialist conference and exhibit, Honolulu, Hawaii, USA, Aug. 2008.

[12] Timoney, J, Lysaght, T, Lazzarini, V, and Gao, R (2005). "A Reinvestigation of the Extended Kalman Filter applied to Formant Tracking". *Proc. of European Signal Processing Conference,* Bratislava, Slovakia.

[13] Wan, E.A., and Van Der Merwe, R., "The Unscented Kalman Filter for nonlinear estimation", Proc. of IEEE Symposium on Adaptive Systems for Signal Processing, Communication and Control 2000, Lake Louise, Alta, Canada, Oct. 2000.

[14] Watanabe, A, "Formant estimation method using inverse-filter control", IEEE Trans. on Speech and Audio Proc., vol. 9, no. 4, May 2001, pp. 317-326.

[15] Watson, G.N. *A Treatise on the Theory of Bessel Functions*. Cambridge Univ. Press, , 2nd Ed., 1944.