WAVE-MORPHING IN THE FRAMEWORK OF A GLOTTAL PULSE MODEL

Julien Hanquinet[1], Francis Grenez[1] , Jean Schoentgen[1,2]

[1]Department "Signals and Waves", Université Libre de Bruxelles, 50, Avenue F.-D. Roosevelt, 1050 Brussels, Belgium, jhanquin@ulb.ac.be
[2]National Fund for Scientific Research, Belgium

ABSTRACT

The presentation concerns a method of wave-morphing applied to a model of the phonatory excitation, the instantaneous frequency and the harmonic richness of which are controlled. This method is based on an interpolation between the Fourier coefficients of two template waveforms. The method enables morphing continuously from one waveshape to another. Possible applications are the simulation of diplophonia, biphonation and different phonation types.

I. INTRODUCTION

The presentation concerns a method of wave-morphing based on Fourier series. It enables continuously changing one waveform into another. This method is applied to a model of the phonatory excitation signal, which is the acoustic signal generated by the vibrating vocal folds and pulsatile glottal airflow.

Conventionally, glottis signals are modeled by means of a concatenation of curves that approximate the glottal pulse shape. The most popular model based on this technique is the Fant-Liljencrants model [1]. A sustained glottis signal is generated by repeating the basic pulse shape periodically.

We proposed here an alternative based on the Fourier signal representation, which offers a more flexible approach to phonatory excitation modeling. It enables controlling continuously the instantaneous frequency and harmonic richness of the synthetic phonatory excitation, as well as glottal pulse morphing. The morphing is carried out by interpolating the Fourier series coefficients between two different template glottal cycles.

II. MODEL OF THE PHONATORY EXCITATION

The model used to synthesize the phonatory excitation is based on Fourier coefficients. The Fourier coefficients are computed for a template cycle of the desired phonatory signal. The template cycle can be modeled or extracted from real speech. Here, we use the Fant-Liljencrants (LF) model [1] to synthesize the desired template. The LF parameters are chosen so that the condition of area balance is fulfilled, i.e. the cycle average is zero.

A discrete periodic signal y of cycle length N can be approximated by its Fourier series truncated at Nh harmonics .

$$y(n) \approx \frac{1}{2}a_0 + \sum_{k=1}^{Nh} a_k \cos(k\frac{2\pi}{N}n) + b_k \sin(k\frac{2\pi}{N}n). \qquad (1)$$

In expression (1), coefficients a_k and b_k encode the shape of the cycles of signal y and parameter N represents the cycle length. By changing the value of N, one can create signals with the same shape as y, but with different cycle lengths. Note that the following condition must be respected.

$$Nh < \frac{N}{2}. \qquad (2)$$

If N is assumed to be real, expression (1) can be written as follows.

$$y(n) = \frac{1}{2}a_0 + \sum_{k=1}^{Nh} a_k \cos(k\theta_n) + b_k \sin(k\theta_n), \qquad (3)$$

where $\theta_n = \theta_{n-1} + 2\pi f\Delta$, f is the instantaneous frequency of signal $y(n)$ and Δ is the sampling step. Condition (2) becomes the following.

$$Nh < \frac{f_{sampling}}{2}\frac{1}{f}. \qquad (4)$$

The generalization of N to real values, because of letting assume f any real positive value, introduces a quantization error of one sample at most in the cycle length. For many applications, this error is negligible when the sampling frequency is chosen sufficiently high.

Therefore, by means of a glottal cycle template, a signal with the same cycle shape, but the instantaneous frequency of which is controlled, can be synthesized by means of (3). Figure 1 shows an example of a phonatory excitation, for which the instantaneous frequency evolves continuously and linearly in time.

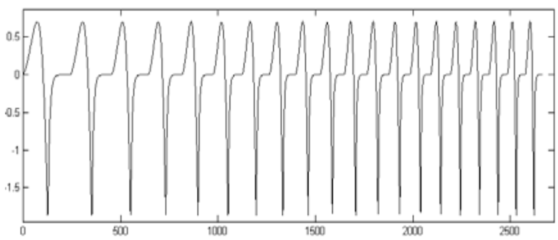

Figure 1 : Synthetic phonatory excitation, the instantaneous frequency of which evolves linearly from 75 to 200 Hz. The vertical axis is in arbitrary

Models and analysis of vocal emissions for biomedical applications. 4th international workshop. October 29-31, 2005 – Firenze, Italy.
Edited by C. Manfredi. ISBN 88-8453-320-1 (online)

units, and the horizontal axis is labeled in number of samples.

The harmonic richness of the synthetic signal can be controlled by modifying the Fourier coefficients as follows. This choice has been loosely inspired by [2].

$$a_k -> a'_k = A^k a_k, \quad b_k -> b'_k = A^k b_k, \quad \text{with } 0 < A < 1 . \qquad (5)$$

One sees in expression (5) that the harmonics decrease, when the parameter A is less than one, the faster the higher their order.

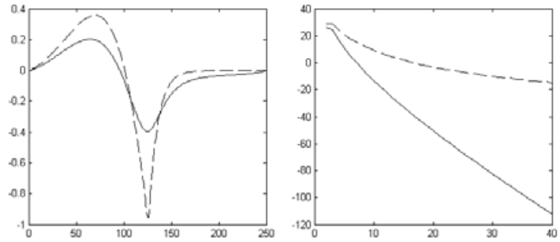

Figure 2 : The graph to the left shows two different cycles of the phonatory excitation. The dashed line is obtained with parameter *A* set to 1 and the solid line is obtained with parameter *A* set to 0.5. The vertical axis is in a.u. and the horizontal axis is labeled in samples. The graph to the right shows, dashed, the values in db of $|a_k + jb_k|$ and, solid, the values in db of $|a'_k + jb'_k|$ with *A* set to 0.5. The horizontal axis is labeled in the values of Fourier index *k*.

The control of the harmonic richness of the phonatory excitation may also be used to simulate onsets and offsets as illustrated in Fig.3.

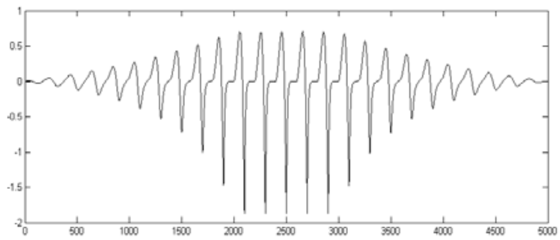

Figure 3 : Synthetic phonatory excitation where parameter *A* evolves linearly from 0 to 1 and from 1 to 0. The vertical axis is in a.u. and the horizontal axis is labeled in number of samples.

III. WAVE-MORPHING

Given two sets of Fourier coefficients X_1 and X_2 ,in complex notation, computed for two different template cycles, intermediary shapes can be synthesized by interpolating the Fourier coefficients as follows (Figure 4).

$$X_k = X_{1k}^{1-Int} X_{2k}^{Int} , \qquad (6)$$

where *Int* is an interpolation coefficient comprised between 0 and 1.
As a consequence, the Fourier phase and the logarithm of the Fourier magnitude are linearly interpolated. Therefore, coefficients a_k and b_k change as follows :

$$a_k = \frac{2}{N}|X_{1K}|^{(1-Int)}|X_{2K}|^{Int}\cos(Arg(X_{1K})(1-Int)+Arg(X_{2K})Int) \qquad (7)$$
$$b_k = \frac{2}{N}|X_{1K}|^{(1-Int)}|X_{2K}|^{Int}\sin(Arg(X_{1K})(1-Int)+Arg(X_{2K})Int)$$

To avoid possible phase distortions in morphed signals, care should be exercised to respect the following condition.

$$|\arg(X_{1k}) - \arg(X_{2k})| < |\arg(X_{1k+1}) - \arg(X_{2k+1})| \qquad (8)$$

To satisfy this condition, one computes the arguments of the two sets of complex Fourier coefficients X_1, X_2, and subtracts 2π from the argument of X_2 if condition (8) is not satisfied. The reason is that the phase of the morphed shape must be intermediary between the phases of the template cycles, which is possible provided that the arguments of coefficients X_1 and X_2 evolve quasi-monotonously.

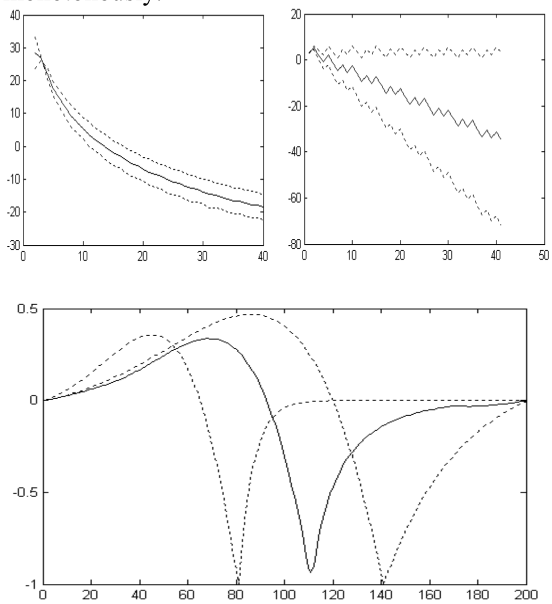

Figure 4 : Above, the graphs show the magnitude in db (to the left) and phase in radians (to the right) of the complex Fourier coefficients. Below, the dotted lines correspond to the template glottal cycles, and the solid line corresponds to the interpolated glottal cycle, with interpolation coefficient *Int* set to 0.5. Above, the horizontal axis is labeled in the values of Fourier index *k*. Below, the horizontal axis is labeled in number of samples.

IV. RESULTS

A. MORPHING

Figure 5 illustrates the phonatory excitation signal while morphing from one cycle template to another, e.g. illustrating the transition from one phonation type to another. The interpolation coefficient evolves, between samples 600 and 3600, linearly from zero to one.

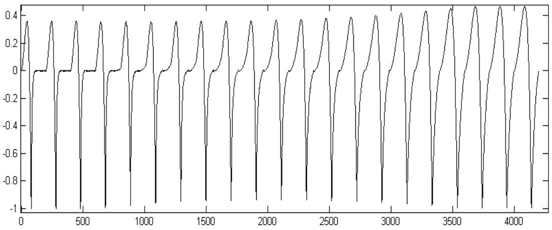

Figure 5 : Morphed synthetic phonatory excitation. The vertical axis is in a.u. and the horizontal axis is labeled in number of samples.

B. DIPLOPHONIA

Diplophonia refers to periodic phonatory excitation signals whose mathematical periods comprise several unequal glottal cycles. A repetitive sequence of different glottal cycle shapes can be simulated by modulating the interpolation coefficient, i.e. by continuously interpolating the Fourier coefficient between two sets of template coefficients X_1, X_2, computed from two different reference glottal cycles. Similarly, a modulation of the instantaneous frequency may simulate a repetitive sequence of glottal cycles of unequal lengths. The temporal evolution of the interpolation coefficient as well as phase may then be written as follows.

$$Int_n = \frac{1}{2}(1 + \sin(\theta_n / Q)) \qquad (9)$$

$$\theta_{n+1} = \theta_n + 2\pi\Delta(f_0 + f_1 \sin(\theta_n / Q)) \qquad (10)$$

The instantaneous frequency oscillates between $f_0 - f_1$ and $f_0 + f_1$. Parameter Q fixes the number of different glottal cycles within the mathematical period of the phonatory excitation. In practice, parameter Q is a small integer.

Figure 6 shows an example of diplophonia obtained by modulating the interpolation coefficient as well as the phase according to expressions (9) and (10), with Q set to two.

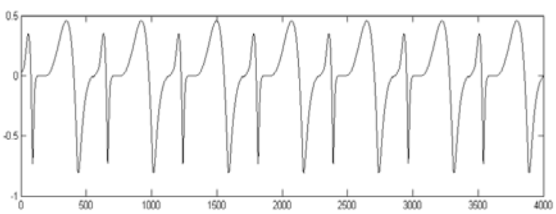

Figure 6 : Synthetic phonatory excitation demonstrating diplophonia. The vertical axis is in a.u. and the horizontal axis is labeled in number of samples.

C. BIPHONATION

Biphonation is also characterized by a sequence of glottal cycles of different shapes and lengths. But in this case, two glottal cycles are never identical. Biphonation reflects the presence in the spectrum of the signal of at least two harmonic series, the fundamental frequency of which form an irrational ratio. Biphonation is therefore characterized by discrete spectra with irrational ratios between the frequencies of some of the partials. Biphonation is also simulated by means of expression (9) and (10), with parameter Q equal to an irrational number.

Figure 2 shows an example of biphonation obtained with Q set to the constant e (2.71).

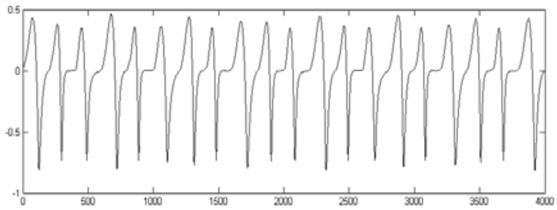

Figure 7 : Synthetic phonatory excitation demonstrating biphonation. The vertical axis is in a.u. and the horizontal axis is labeled in number of samples.

Note that diplophonia and biphonation can also be simulated by modulating phase (10) and parameter A instead of interpolation (9). This is because parameter A controls the harmonic richness and therefore the shape of the cycle. The control is less flexible however.

V. CONCLUSION

This presentation concerns a model of the phonatory excitation based on Fourier series. This model enables the control of the instantaneous glottal cycle length, instantaneous harmonic richness and glottal cycle shape via distinct parameters. This model also enables interpolating between two template cycle shapes. The shape of the cycles of the phonatory excitation may morph continuously from one shape to another. These possibilities are useful to simulate onsets and offsets, intonation, phonation type transients as well as diplophonia and biphonation.

REFERENCES

[1] Fant G., Liljentcrants J., Lin Q., " A four-parameter model of glottal flow ", STL-QSPR, 4: 1-13, 1985.
[2] Schoentgen, J., "Shaping function models of the phonatory excitation signal", J.Acoust. Soc. Am. 114(5): 2906-2912, 2003.