SUMMARY DRAFT

Synthesis of disordered voices

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

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

Abstract. This presentation concerns the simulation of disordered voices. The synthesis is based on shaping functions, which are nonlinear memoryless input-output characteristics that transform a trigonometric driving function into a synthetic phonatory excitation signal. One advantage of the shaping fuction model is that the instantaneous frequency and the spectral balance of the phonatory excitation signal are controlled by two separate parameters. It is shown how to synthesize different types of dysperiodicities by modulating both the amplitude and instantaneous frequency of the driving function. The voice disorders that are simulated are short- and long-term perturbations of the vocal frequency, biphonation, diplophonia and raucity. Turbulence noise is modeled by additive white noise.

1 Introduction

The presentation concerns a model of the phonatory excitation signal, which enables simulating several types of voice disorders. The phonatory excitation signal is the acoustic signal that is generated at the glottis by the vibrating vocal fold and pulsatile airflow. Conventional models, such as the Liljencrants-Fant model [1], enable controlling the total glottal cycle length only. This means that any change of the cycle length must be synchronized with the onsets or offsets of the excitation cycles. This synchrony has no physiological basis. It is therefore proposed to synthesize the phonatory excitation by means of a shaping function, which is an operator that transforms a trigonometric driving function into any desired waveform. One property of this model is that the instantaneous frequency and spectral balance of the signal are controled by two separate parameters which are the instantaneous frequency and the amplitude of the driving function respectively.

2 Phonatory Excitation and Vocal Tract Model

The model used to synthesize disordered voices is based on a shaping function model [2]. The shaping function is a nonlinear memoryless input-output characteristic that transforms a cycle of a harmonic into any desired cycle shape. The shaping function involves two polynomials f and g whose coefficients are obtained via a linear transformation of the Fourier series coefficients of the shape of a cycle of the desired phonatory signal. This template cycle can be modeled or extracted from real speech [2].

The formal expression of the shaping function is the following:

$$s(n) = f[\cos\theta(n)] + \sin\theta(n) g[\cos\theta(n)] \ . \quad (1)$$

Conventionally, the phonatory excitation is considered to be the derivative with respect to time of the glottal airflow rate. In the framework of the modeling of the phonatory excitation signal, the coefficients of the polynomials *f* and *g* are therefore computed by means of a template flow rate, after which the derivative of expression (1) is taken. To remove the dependency of the phonatory signal amplitude on the phonatory signal frequency, the derivative is taken with respect to phase.

Also, to control the spectral balance of the synthetic excitation signal, the polynomials are driven by a trigonometric function whose amplitude A may differ from unity. The spectral balance decreases with decreasing A.

Consequently, the phonatory excitation is written as follows.

$$e(n) = \frac{d}{d\theta}\{f[A\cos\theta(n)] + A\sin\theta(n) g[A\cos\theta(n)]\} \text{ with } 0 \le A \le 1 \ . \quad (2)$$

By means of expression (2), several kinds of dysperiodicities may be simulated by modulating the amplitude and frequency of the trigonometric driving function. A condition is that the product of the upper bound of the effective bandwith of the driving function times the order of the shaping function is less than half the sampling frequency, to avoid aliasing.

To simulate the propagation of the phonatory excitation through the vocal tract, a concatenation of cylindrical tubelets is used. Each tubelet has the same length. For each tubelet, viscous, thermal and wall vibrations losses are modeled by means of filters. To simulate the transition at the lips from one-dimensional to three-dimensional wave propagation, a conical tubelet, the opening of which is controlled, is added at the lip-end of the vocal tract model.

3 Synthesis of Disordered Voices

3.1 Vocal Jitter and Microtremor

Jitter and microtremor designate small random cycle-to-cyle perturbations and a low-frequency random modulation of the glottal cycle lengths respectively. Jitter and microtremor are therefore inserted into the phonatory excitation model by perturbing the instantaneous frequency of the driving function by two random components [3]. Consequently, the discrete-time evolution of the phase of the sinusoidal driving function is written as follows.

$$\theta_{n+1} = \theta_n + 2\pi(f_0\Delta + j_n + m_n) \ . \quad (3)$$

Symbol f_0 is the unperturbed instantaneous vocal frequency; Δ is the time step; j_n is uniformly distributed white noise that simulates intra-cycle frequency perturbation that give rise to jitter; m_n is uniformly distributed white noise filtered by a linear second order filter, which sets the microtremor frequency and bandwith.

3.2 Diplophonia

Diplophonia refers to periodic phonatory excitation signals whose periods comprise several unequal glottal cycles. A repetitive sequence of different glottal cycle shapes can be simulated by modulating the amplitude of the driving function because the amplitude of the driving function influences the spectral balance of the phonatory excitation. Similarly, a modulation of the instantaneous frequency of the driving function may simulate a repetitive sequence of glottal cycles of unequal lengths. The temporal evolution of the amplitude and phase of the driving function are then written as follows.

$$A_n = A_0 + A_1 \sin(\theta_n / Q) \tag{4}$$

$$\theta_{n+1} = \theta_n + 2\pi\Delta(f_0 + f_1 \sin(\theta_n / Q)) \ . \tag{5}$$

Parameter Q sets the number of different glottal cycles within the mathematical period of the vocal excitation. In practice, parameter Q is a small integer.

3.3 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 is therefore characterized by discrete spectra with irrational ratios between the frequencies of the partials. Biphonation is simulated by means of an expression similar to (5). The difference is that parameter Q is equal to an irrational number.

3.4 Random Cycle Lengths

Contrary to jitter, which is due to external perturbations of a dynamic glottal regime that is periodic, random vibrations are the consequence of a random dynamic regime. The relevant model parameter is therefore the total cycle length. In the framework of model (2), the choice of a new instantaneous frequency must be synchronised with the onsets of the excitation cycles. The reason is that the statistical distribution of the cycle lengths is requested to follow a gamma distribution. The gamma distribution is the simplest distribution that enables determining independently the variance of the positive cycle lengths and their average.

3.5 Turbulence Noise

Turbulence noise is taken into account by means of additive noise, which simulates the effect of turbulent airflow through the glottis. These turbulences are expected to occur in the vicinity of glottal closure. Uniformly distributed white noise the amplitude of which is proportional to the phonatory excitation signal is therefore added to the phonatory excitation signal when it is negative. No noise is added when the signal is positive or zero.

4 Methods

The template flow rate used to compute the coefficients of the shaping polynomials is a synthetic cycle simulated by the Liljencrants-Fant model. Its Fourier coefficients are computed numerically. The total number of coefficients is 80: 40 cosine coefficients and 40 sine coefficients. The polynomial coefficients are obtained from the Fourier coefficients by means of a linear transformation. The sinusoidal driving function is sampled at 100KHz.

The simulated vocal tract shape is fixed on the base of published data. The number of cylindrical tubelets used in the vocal tract model is comprised between 20 and 30.

5 Results

The presentation will comprise auditory examples of normal and disordered voices. Preliminary tests show that the model that is presented here enables synthesizing vocal timbres that are perceived as plausible exemplars of disordered voices. We here illustrate graphically the ability of the synthesizer to simulate different vocal timbres by means of examples of synthetic diplophonic, biphonic and rough voices.

Figure 1 shows an example of diplophonia obtained by modulating the driving functions following expression (4) and (5) with Q set to two.

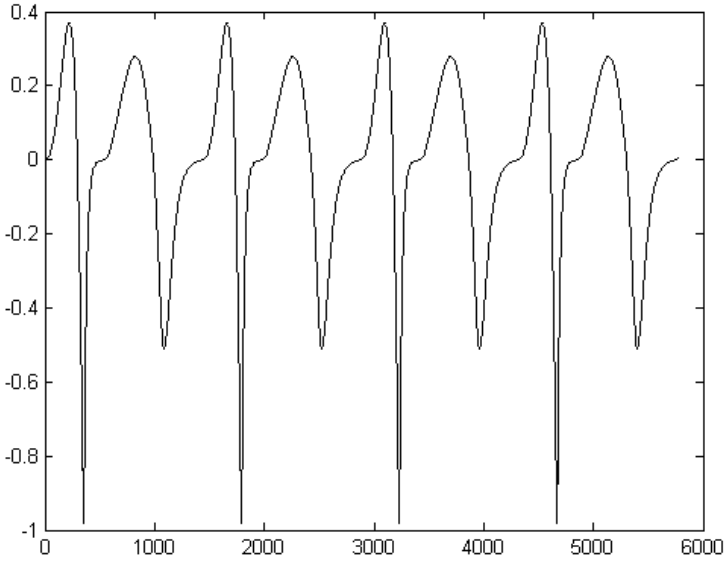

Fig. 1. Synthetic excitation signal simulating diplophonia. The horizontal axis is labeled in number of samples and the vertical axis is in arbitrary units.

Figure 2 shows an example of biphonation obtained by modulating the driving functions following expression (4) and (5) with Q set to the constant e (2.71…).

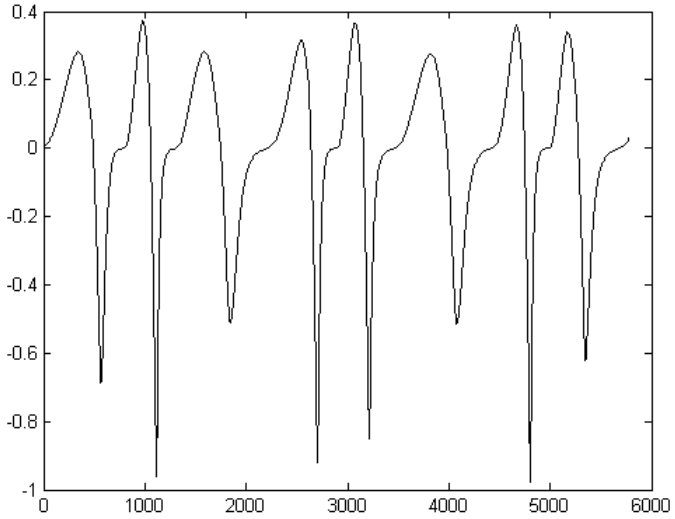

Fig. 2. Synthetic excitation signal simulating biphonation. The horizontal axis is labeled in number of samples and the vertical axis is in arbitrary units.

Figure 3 shows an example of rough voice (random cycle lengths). The instantaneous frequencies have been randomly selected from a gamma distribution in synchrony with the onsets of the cycles. The mean and standard deviation of the gamma distribution have been equal to 100 Hz and 25 Hz respectively.

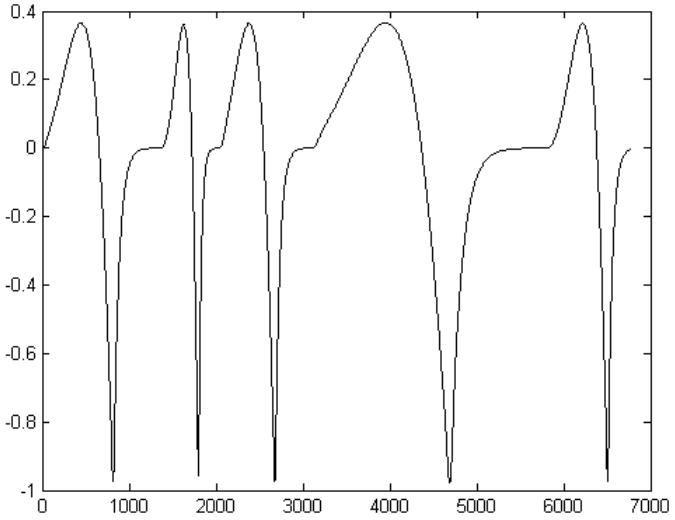

Fig. 3. Synthetic excitation signal simulating random vibration. The horizontal axis is labeled in number of samples and the vertical axis is in arbitrary units.

6 Conclusion

The presentation concerns a simulation of disordered voices. The model is based on a nonlinear memoryless input-output characteristic that transforms a trigonometric driving function into a synthetic phonatory excitation signal. Several types of

dysperiodicities can be simulated by modulating the amplitude and/or frequency of the trigonometric driving function. The model enables synthesizing a wide range of vocal phenomena, such as jitter, microtremor, diplophonia, biphonation, raucity and breathiness. Preliminary tests show that the simulated voices are perceived as plausible exemplars of voice disorders.

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.
3. Schoentgen, J., "Stochastic models of jitter", J. Acoust. Soc. Am. 109 (4): 1631-1650, 2001.