SYNTHETIC SIGNALS FOR CLINICAL APPLICATIONS: ALTERED VOICE QUALITY AND NEWBORN CRY

J. Schoentgen[1], C. Manfredi[2] and P.H. DeJonckere[3]
[1]Faculty of Applied Sciences, Université Libre de Bruxelles, Brussels, Belgium jschoent@ulb.ac.be
[2]Department of Information Engineering, Università degli Studi di Firenze, Firenze, Italy
[3]Neurosciences KULeuven and Federal Agency for Occupational Risks, Brussels.

Introduction

Voice quality designates perceived speaker-identifying speech sound properties. The objective of the numerical simulation of voice quality is the synthesis of voiced sustained or concatenated speech sounds with a view to carrying out perceptual experiments, testing clinical analysis software or recording equipment and assessing human voice qualities via synthetic ones. The purpose of the perceptual experiments is to discover acoustic or morphological cues that report or, alternatively, mask voice quality. The simulation includes models that respectively mimic vocal jitter, the evolving shape of the glottis, the glottal airflow rate including acoustic source-tract interaction, additive aspiration or pulsatile noise as well as the vocal tract transfer function and acoustic radiation.

Another issue is high-pitched voices in newborns. The newborn cry is a signal difficult to analyse with standard techniques due to its lack of stationarity and the very high range of vocal frequencies. Therefore, synthetic signals may help in assessing recent and existing software tools, including those not developed for that purpose. The acoustical analysis of the infant cry can assist the clinical specialist in detecting possible neurological disorders.

Synthesized altered voices

The vocal jitter model is inspired by the physiology of multiple motor units that collectively mimic TA muscle tension the fluctuations of which contribute to vocal frequency jitter. The model of the evolving glottis involves distinct oscillators that mimic the motion of the left and right glottal walls, the anterior and posterior glottis as well as the glottal entrance and exit that move out of phase. Distinct oscillators can be operated at different amplitudes and frequencies to simulate diplophonia, aperiodic vibrations, etc.

The motion of the left and right glottal walls comprises distinct oscillators for fold body (muscle) and fold cover to enable controlling the size of vocal jitter fold-internally. A decrease of the amplitude of vibration of the cover with regard to the body indeed boosts the effect of TA muscle tension jitter and possibly explains the observed increase of vocal jitter and perceived hoarseness in conditions that are expected to dampen the motion of the cover of the folds. A decrease of the amplitude of vibration of the body with regard to the cover, on the contrary, lessens the effect of muscle tension jitter and possibly explains, independently of the active tension of the muscle, the observed decrease of frequency jitter in voices the production of which involves a reduction of the amplitude of vibration of the body of the vocal folds. The collision of the left and right glottal walls as well as the projection of the 3D glottal volume on the 2D glottal area are simulated so as to enable smooth flow rate pulse onsets and offsets.

The pulsatile airflow rate is obtained by solving an equation that has been proposed by Rothenberg, but which is reformulated to obtain solutions for all physiologically possible values of the glottal shape and acoustic inertances and resistances of the supra and infra-glottal tracts. The latter skew the flow rate shape and so increase harmonic content. Pulsatile or aspiration noise, the control parameters of which include the spectral contour of the noise, is added to the glottal airflow rate.

The synthetic speech signals are obtained by filtering the glottal flow rate and noise by the vocal tract transfer function, which is modelled by means of a concatenation of resonators the bandwidths and gains of which can be fixed independently, which is an option that is lacking in conventional concatenated formant synthesisers. The closed and open glottis bandwidths differ to take into account the increase of the acoustic losses with glottal opening.

Synthesized voice samples have proven to be very useful for several practical and clinical purposes, either as a sustained /a:/ or as concatenated /aiu/ segments, the latter simulating a short voiced sentence, with intonation. Perceptual experiments have shown that these synthesized voice samples sound realistic and cannot be distinguished by bachelor students in speech language therapy from actual patient's voices.

Overview of clinical applications implemented so far

(1) Defining the limit of validity for jitter measurements with acoustic analysis programs. A widely accepted guideline was that only perturbation measures less than about 5% are reliable (sustained vowels). We have shown that improved acoustic analyses using more reliable algorithms can validly exceed this limit. This is relevant for e.g. spasmodic dysphonia or substitution voices.

(2) Testing various software tools. Even when (synthetic) irregular voice signals (sustained vowels) are corrupted by additive noise, several software tools provide reliable jitter measurements up to 15%.
(3) Comparing the performances of trained visual raters and those of computer programs with regard to vocal cycle identification, when synthetic sustained vowels are corrupted by increasing cycle irregularity and additive noise. Visual cycle pattern recognition by trained clinicians gives the best results, but for high noise levels, the raters tend to overestimate the inputted jitter. Also large differences between software tools are observed.
(4) Checking the ability of software tools and of perceptual raters to discriminate ten levels of additive noise in synthetic concatenated [aiu] segments with intonation and small levels of jitter, comparable to breathy voices of female patients with vocal fold nodules. Such synthetic voice samples may be relevant anchors for medicolegal experts who need to quantify the severity of the voice deviance. Vocal fold nodules are indeed the most common occupational voice disease, likely to be financially compensated in several European countries.
(5) Testing whether smartphones are adequate for the reliable recording and sampling of pathologic voices, to be transferred to a voice lab. Repeated measurements (e.g. pre- and post-treatment) are the basis of single-subject and baseline research designs. The tested smartphones demonstrate a good reliability with regard to jitter and noise-to-harmonics ratio for sustained vowels only. The same does not apply (so far) to voice samples with formant shifts as in connected speech. Further, the ambient noise level must not exceed 50 dB_A.

Synthetic newborn cries

Along with the perceptual analysis, the automated analysis of the cry is usually performed by means of software tools that however may not have been developed with a view to analysing un-stationary high-pitched vocal signals. To assess existing software, a synthesizer was specifically developed to reproduce the melodic shape of the neonatal cry. Synthetic signals are obtained implementing a new accurate simulation model that takes into account the variability of the newborn cry as well as the basic shapes of the newborn cry melody. It consists of a pulse train generator and a filter bank with 3 parallel resonance filters (2^{nd}-order all-pole filters). White noise of increasing amplitude is added to the signal.

Three analysis methods have been tested and compared, one freely available and the other two specifically built using two different approaches: autoregressive adaptive models and wavelets. Results point out strengths and weaknesses of each method, thus suggesting the most appropriate use according to the goals of the analysis.

References

Manfredi C , Lebacq J , Cantarella G, Schoentgen J, Orlandi S, Bandini A, DeJonckere, PH. Smartphones offer new opportunities in clinical voice research. J Voice 2016. 31(1), 111.e1-111.e7.doi:10.1016/j.jvoice.2015.12.020

Lebacq J, Schoentgen J, Cantarella G, Bruss FT, Manfredi C, DeJonckere PH. Maximal ambient noise levels and type of voice material required for valid use of smartphones in clinical voice research. JVoice 2017, in press.

Rruqja N., Dejonckere PH, Cantarella G., Schoentgen J., Orlandi S., Barbagallo D., Manfredi C. Testing software tools with synthesized deviant voices for medicolegal assessment of occupational dysphonia. Biomed Signal Process Control 2014; 13: 71–78.

DeJonckere PH, Schoentgen J, Giordano A, Fraj S, Bocchi L, Manfredi C. Validity of jitter measures in nonquasi-periodic voices. Part I. Perceptual and computer performances in cycle pattern recognition ; Logoped. Phoniatr. Vocol. 2011 ; 36 (2) : 70-77.

Manfredi C, Giordano A, Schoentgen J, Fraj S, Bocchi L, DeJonckere PH , Validity of jitter measures in non-quasi-periodic voices. Part II. The effect of noise. Logoped. Phoniatr. Vocol. 2011 ; 36 : 78–89.

DeJonckere PH, Giordano A, Schoentgen J, Fraj S, Bocchi L, Manfredi C . To what degree of voice perturbation are jitter measurements valid ? A novel approach with synthesized vowels and visuo-perceptual pattern recognition. Biomed Signal Process Control, 2012 ; 7 : 37– 42.

Manfredi C , Giordano A , Schoentgen J , Fraj S , Bocchi L , DeJonckere PH . Perturbation measurements in highly irregular voice signals : performances/ validity of analysis software tools. Biomed Signal Process Control. 2012 ; 7 : 409–416.

DeJonckere PH. Voice evaluation and respiratory voice assessment. In Anniko M & al. Eds. Otorhinolaryngology, Head & Neck Surgery. Springer, Heidelberg Berlin, 2010. 563 – 574.

Orlandi, S., Bandini, A., Fiaschi, F. F., Manfredi, C.. Testing software tools for newborn cry analysis using synthetic signals. Biomedical Signal Processing and Control 2016 doi:10.1016/j.bspc.2016.12.012 (in press)