Improving pitch perception with cochlear implants – Implications for music and speech

Matthias Milczynski, Jan Wouters and Astrid van Wieringen

ExpORL, Department Neurosciences, Katholieke Universiteit Leuven, O & N2, Herestraat 49, bus 721, Leuven, Belgium

The research study described here focuses on the evaluation and further development of a cochlear implant (CI) processing strategy called F0mod (Laneau *et al.*, 2006). The strategy facilitates coding of the fundamental frequency (F0) of an acoustical signal in the electrical stimulation pattern. This document presents concepts of the processing algorithm. Furthermore, an overview about the framework of psychophysical experiments with CI subjects covering music and speech perception related tasks will be given.

IMPORTANCE OF PITCH PERCEPTION IN NORMAL HEARING

Pitch is a fundamental perceptual dimension in normal hearing (NH). Primarily, the property of extracting pitch out of complex acoustical signals is associated with listening to music. In this domain pitch conveys information about the melody contour, which is relevant for melody *learning*, as well as information about intervals between successive notes, which is essential for melody *recognition* (Gfeller *et al.*, 2002). NH persons not only can recognize and learn melodies but e.g. state whether a musical piece is played out of tune (intonation). Beyond the musical scope, pitch perception is crucial for speech perception. Thereby, gender-specific characteristics of voice, prosodic cues but also semantics, as known from tonal languages, are mediated by pitch. These and many other examples (e.g. source separation, speech understanding in noisy environments etc.) emphasize the importance of pitch perception in hearing.

PROBLEMS WITH PITCH PERCEPTION IN ELECTRICAL HEARING

In electrical-hearing (EH) pitch perception cues are sparsely provided. As a consequence, many CI users do not appreciate music. Just noticeable differences (JNDs) in F0 for CI-patients vary between 1 and 24 semitones (0.1 semitones in NH persons), which hampers e.g. perception of melodic patterns. Major problems in EH are missing or respectively inaccurate equivalents for pitch perception mechanisms in NH. Analogies between spectral pitch in NH and place pitch in EH as well as purely temporal pitch in NH and temporal pitch in EH can be deduced. However, a direct complement to periodicity pitch (i.e. pitch percept based on F0 of a sound) in NH is not present in EH.

THE F0MOD STRATEGY

A new DSP strategy called F0mod has been developed with the aim to improve the transmission of temporal pitch cues. The main concept of the strategy is to extract

Auditory signal processing in hearing-impaired listeners. 1st International Symposium on Auditory and Audiological Research (ISAAR 2007). T. Dau, J. M. Buchholz, J. M. Harte, T. U. Christiansen (Eds.).
ISBN: 87-990013-1-4. Print: Centertryk A/S.

the F0 of an incoming sound and to amplitude-modulate each channel of the electrical stimulation pattern at the extracted F0. Thus, the F0, which in many cases can be regarded as a physical counterpart of the perceptual dimension pitch, is coded in the electrical stimulation pattern.

Algorithm description

The basic processing scheme of F0mod (see Fig. 1) is derived from the commercially available Advanced Combinational Encoder (ACE) strategy used in the Nucleus Freedom devices of Cochlear Corporation. First, an incoming signal is processed by a FFT-based filter bank. The resulting magnitudes of the corresponding FFT-bins are then summed into a total number of 22 channels. Second, an F0-estimator extracts the fundamental frequency of the incoming sound. Both parallel processes (i.e. channel magnitude extraction and F0-estimation) are combined by amplitude modulating the channel magnitudes at the extracted F0. Thereby all channels are modulated at full modulation depth and in phase. The subsequent procedures such as maxima selection and compression are common in recent CI strategies and will not be described further.

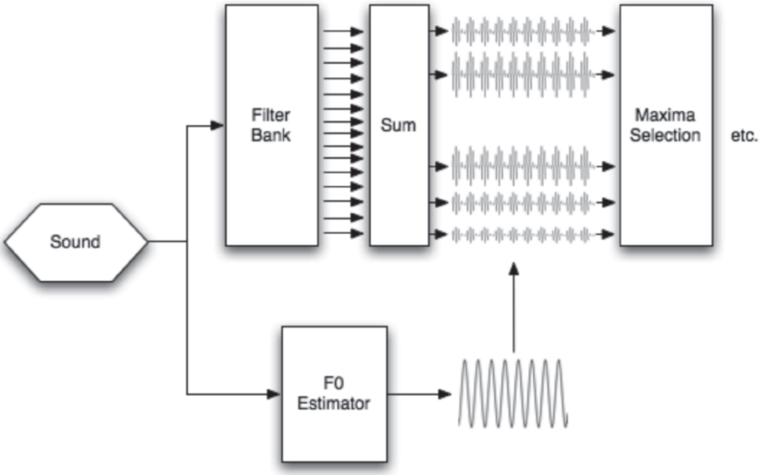

Fig. 1: F0mod processing scheme. Each channel is amplitude-modulated at the extracted F0.

In the pilot-implementation of the F0mod strategy the F0 was estimated by a simple autocorrelation-approach. However, no robustness features, essential for F0-estimation under real-life conditions were included. For the recent study, again the autocorrelation-approach serves as a starting point for F0-estimation. However, several subcomponents have been added in order to refine the F0-extraction procedure.

F0-estimation using the short-term autocorrelation function

A possible representative for the F0 of a currently processed buffer of length W is given by the maximum of the short-term autocorrelation function $r_t(\tau)$ (see de Cheveignè and

Kawahara, 2002) the digitized waveform:

$$r_t(\tau) = \frac{1}{W-\tau} \sum_{i=t+1}^{t+W-\tau} x_i x_{i+\tau} \qquad \text{(Eq. 1)}$$

In Eq. 1 the estimated F0 corresponds to fs/τ_{max}, where fs is the sampling frequency and τ_{max} is the lag for which the autocorrelation function $r_t(\tau_{max})$ is maximal. In order to facilitate a reasonable F0-estimation in real-life situations, such as e.g. conversations in noisy environments the F0-estimation algorithm was enhanced by a number of additional robustness features.

Subharmonic-error correction

Due to variations in amplitude of a particular waveform the lag τmax is not necessarily representative for the actual F0. In other words, the autocorrelation function might have several local maxima of which the appropriate one has to be chosen. For this purpose a correction routine is implemented that detects valuable harmonics of the extracted F0. This is illustrated in Fig. 2. Besides the maximum at a lag L1 (lag 244) a strong second harmonic at L2 (lag 122) lies above a specified harmonic-threshold (upper horizontal line). L2 will be chosen as the F0-representative. Thus, the term subharmonic-error refers to (falsely) deciding on a subharmonic of the true fundamental frequency.

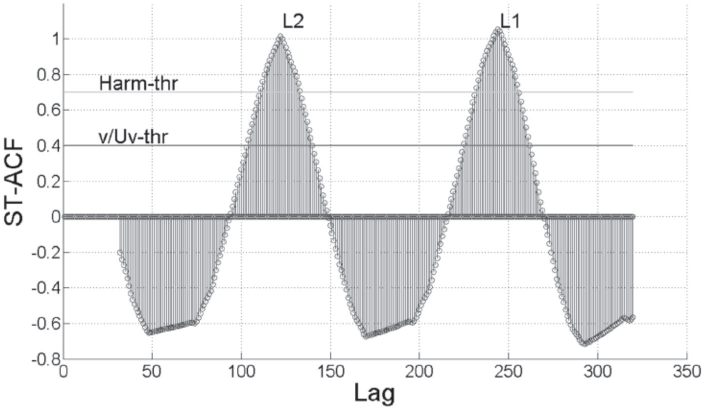

Fig. 2: Illustration of subharmonic-error correction and voiced-unvoiced classification.

The estimated F0 in this case will be 131 Hz (16 kHz sampling rate).

Voiced-unvoiced discrimination

For the transmission of speech a different approach is desired for voiced and unvoiced segments. Stimulation patterns representing e.g. unvoiced fricatives are not intended to be amplitude modulated in contrast to those representing vocals. In order to realize this purpose a voiced-unvoiced threshold (v/Uv-thr, see bottom horizontal line in Fig. 2) is introduced into the F0-estimation algorithm. If the autocorrelation value of the F0-corresponding lag lies above this threshold the signal segment is classified as voiced. Otherwise, the segment is classified as unvoiced and consequently no explicit amplitude modulation will take place.

Fig. 3 illustrates example electrodograms obtained with ACE (electrodogram at the top) and F0mod (electrodogram at the bottom). Both electrodograms show the transition between the voiced 'a' vowel and the unvoiced 's' consonant in more detail. While the segment representing the vowel exhibits clearly the in-phase modulations across channels the segment representing the unvoiced 's' consonant remains unmodulated.

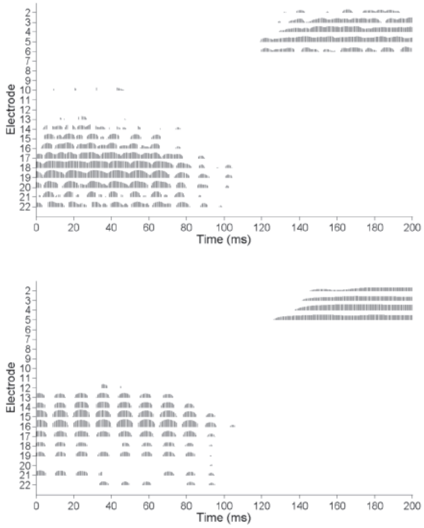

Fig. 3: Electrodograms obtained with ACE (top) and F0mod (bottom) showing the transition from voiced to unvoiced in the syllable "ASA".

FRAMEWORK FOR PSYCHOPHYSICAL EVALUATION OF F0MOD STRATEGY AND RESULTS FROM PRELIMINARY EXPERIMENTS

The evaluation of F0mod contains test-protocols that address music and speech perception. In all tests F0mod will be evaluated in comparison to the commercially available ACE strategy. The objectives that we target are to improve pitch perception performance in music, maintain speech recognition abilities in quiet and eventually enhance speech understanding in noise. Thus a universal processing strategy is desired. In the next sections an overview of the procedures that will be conducted in the near future will be given.

Psychophysical procedures related to music perception

One of the music related psychophysical procedures will include pitch ranking of a tone of a musical instrument. Thereby, the subject will be presented with 2 stimuli of which the higher on has to be indicated. Five different MIDI-synthesized instruments will be used in order to evaluate a variety of spectrally different tones. In addition, melody recognition of familiar Flemish melodies, as already done in the previous study (see Laneau *et al.*, 2006 for details) will be addressed. In this case the subject is asked to recognize a monophonic melody (with and without rhythmic cues) from a closed set off possible answers. Since this type of melody recognition test (with isochronous melodies) is a difficult task for the CI subjects we plan to include tests such as the melodic contour identification test (see Galvin *et al.*, 2007). This and other similar tests, which focus on recognition and discrimination of musical patterns, do not require familiarity with a particular song. However, at the same time these tests reflect perceptual abilities that are important for melody learning and recognition.

Psychophysical procedures related to speech perception

In a first phase speech perception in quiet will be tested. First, consonant and vowel recognition experiments will be conducted with samples taken from the Leuven Analytical Test (LAT). In addition, sentence recognition in quiet will be tested with sentences from the LIST database (see van Wieringen and Wouters, 2005). From these tests we expect similar scores with F0mod and with ACE. It is possible that F0mod is also better for some phonemes (LAT) example transmission of Voice Onset Time, one of the cues for distinction b-p or t-d. With more redundant speech materials, e.g. sentences, we would not expect to see a difference in quiet either).

In a second phase speech recognition in noise will be accessed with the same procedures as described above but with speech samples corrupted with noise at different SNRs. At SNRs about 5-10 dB we expect advantages with F0mod, assuming that the algorithm will provide a reliable F0 estimation under these conditions.

Preliminary results

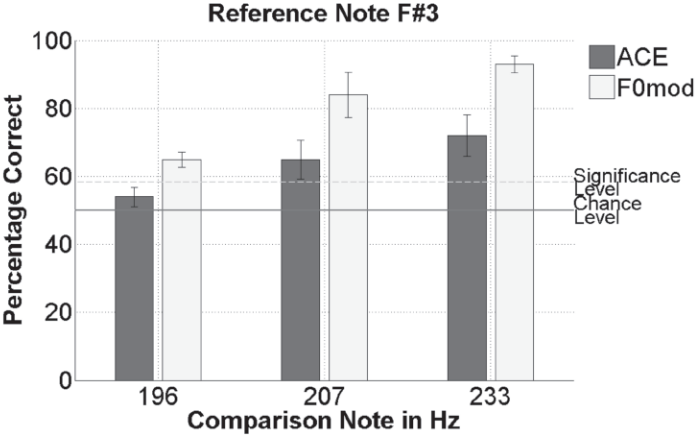

Fig. 4: Results taken from a pitch ranking experiment. The x-axis shows the frequency of each comparison note. On the y-axis the % correct scores are shown.

In a preliminary study pitch ranking, as described above, has been addressed in five CI users. Upon each trial the subjects were presented with a reference note at 131 Hz

(C3), 185 Hz (F#3) or 370 Hz (F#4) respectively and a comparison note that was one, two or four semitones higher than the reference. The order of reference and comparison note was randomly chosen. Each of the different conditions (3 frequency registers represented by reference note) was tested in a separate run for ACE and F0mod. The experiments were carried out by using a real-time implementation of both strategies stored on a Freedom Nucleus speech processor provided by Cochlear Ltd.

Fig. 4 shows the results for the F#3 condition. On the x-axis the frequency of each comparison note is shown. Each bar depicts the correct score in percent, averaged over 5 subjects, for ACE (green bars) and F0mod (yellow bars). The error bars represent the standard error of the mean. In this case, the scores obtained with F0mod were significantly better than scores obtained with ACE ($p < 0.05$). Note that for each comparison note the average F0mod score is significantly above chance level, which is clearly not the case for ACE for the smallest interval of one semitone. For the other conditions the scores were note significantly different between both strategies.

SUMMARY

The research presented here addresses the improvement of pitch perception with CIs with a new processing strategy called F0mod. The algorithm adds a F0 estimation component into an ACE-like processing scheme and amplitude-modulates the resulting stimulation pattern at the extracted F0. The F0 estimator includes several robustness parameters in order to provide a controlled modulation only for particular types of signals (e.g. no modulation for unvoiced consonants in speech). The strategy will be evaluated relative to the commercially available ACE strategy. Both, music and speech recognition related psychophysical procedures will be conducted. An improved performance with F0mod for music related tasks is expected. On the other hand no degradation in speech recognition performance in quiet, but increased speech recognition scores in noise are expected.

In a preliminary experiment a significant improvement in pitch ranking could be shown with F0mod in comparison with ACE for a limited frequency interval. The results support the importance of temporal cues for pitch coding in electrical stimulation.

This project is carried out in the frame of an IWT (Institute for the Promotion of Innovation by Science and Technology in Flanders) project and is partly sponsored by Cochlear Ltd.

REFERENCES

de Cheveignè A., and Kawahara H., (**2002**). “Yin, a fundamental frequency estimator for speech and music,” J. Acoust. Soc. Am., **111**, 1917–1930.

Galvin, J. J. III, Fu, Q., Nogaki, G., (**2007**). “Melodic Contour Identification by Cochlear Implant Listeners,” Ear & Hearing. **28**(3), 302-319.

Gfeller, K., Turner, C., Mehr, M., Woodworth, G., Fearn, R., Knutson, J., Witt, S., and Stordahl, J., (**2002**). “Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults,” Cochlear Implant **3**, 29–53.

Laneau, J., Boets, B., Moonen, M, van Wieringen, A., and Wouters, J., (**2005**). “A flexible auditory research platform using acoustic or electric stimuli for audults and young children,” Journal of Neuroscience Methods, **142**, 131–136.

Laneau, J., Moonen, M., Wouters, J., (**2006**). “Improved music perception with explicit pitch coding in cochlear implants,” Audiology and Neuro-Otology, **11**, 38–52.

Van Wieringen, A.,Wouters, J., (**2005**) “LIST en LINT, Nederlandstalige spraakaudiometrielijsten met zinnen en getallen,” realisatie Lab. Exp. ORL-NKO, K.U. Leuven, CD SIG0501/2.

van Wieringen, A., Wouters, J., “List en lint: Dutch speech audiometry lists with sentences and numbers,” in review for a special issue of the International Journal of Audiology.