APT: Enhanced Speech Comprehension Through Adaptive Pitch Transposition in Cochlear Implants

Kevin Struwe(✉)

Institute of Applied Microelectronics and Computer Engineering,
University of Rostock, 18051 Rostock, Germany
kevin.struwe@uni-rostock.de

Abstract. Cochlear Implants are a marvelous option for hearing-impaired patients to perceive sound again, yet some people still cannot benefit from their implant. The resulting problem is a suboptimal pitch discrimination, which in turn is a vital factor for vowel identification. As an approach to solve this problem, this paper proposes an adaptive pitch transposition, which transposes a given signal to a fixed pitch in which the implant patient has the best comprehension. This pitch is called comfort pitch and has to be determined in a patient test. As a result of these considerations, APT was prototypically implemented in Octave/Matlab and serves as an evaluation platform for the quality of the produced signals. Furthermore, the comfort pitch is a new measure to tune the performance of a cochlear implant patient. The entire concept can be used as a preprocessing stage for state-of-the art speech processing algorithms in cochlear implants.

Keywords: Cochlear Implants · Pitch transposition · Speech processing

1 Introduction

Cochlear implants (CI) are a marvelous way for patients with severe sensorineural hearing loss to gain or regain the ability to perceive sound from their surroundings. In a CI, the sound is captured by a microphone and processed by a speech processor, which transforms the analog sound wave into digital pulses. These pulses are then transmitted to an electrode array in the cochlea, where they invoke an electric field to stimulate the hearing nerve at the electrode's position. However, in spite the success of these implants, some patients still cannot comprehend speech. Low speech comprehension is caused by a variety of circumstances: among others, shallow insertion depth of the electrode array and a low brain plasticity are limiting factors [5]. That leads to a residual hearing range, which is shifted upwards and prohibits the perception of lower frequencies.

To tackle this problem, this paper proposes an adaptive pitch transposition, or APT for short, for transposing the input speech signal to the frequency range in which a particular patient has the best comprehension with his existing

K. Giokas et al. (Eds.): eHealth 2016, LNICST 181, pp. 224–228, 2017.
DOI: 10.1007/978-3-319-49655-9_30

cochlear implant. The perceived speech frequency is called Pitch. The frequency range for best comprehension is thus called comfort pitch and is an empirically determined value that describes the Pitch in which a patient has the best comprehension.

APT uses the optimal range of frequencies for each individual patient, even with a drastically reduced number of active electrodes. Furthermore, by reducing pitch variance by fixing the pitch not to a range but a single value, APT provides the brain with always the same stimulus. This makes it easier for the patient to learn and adapt to the new sound experience.

The remainder of the paper is organized as follows. Section 2 covers the actual problem and how it fits into the context of current research. This leads to the concept of APT, which is explained in depth in Sect. 3. This section describes the individual parts of the concept in detail. The achieved results are presented and discussed in Sect. 4, which concludes the paper.

2 Problem Description

Even with todays sophisticated algorithms, not everyone can understand speech with his CI. The standard insertion depth of a CI electrode array does not allow the patient to perceive frequencies lower than 1000 Hz [3]. With a reduced insertion depth, this limit raises even further. Because of this, and CIs coarse frequency representation, CI patients may have up to 24-times worse pitch perception than a hearing control group [6].

However, research aiming at enhancing pitch perception was recently done by Laneau et al. [2], which developed `f0mod`. `f0mod` modulates the electrode signal with the fundamental frequency, which has been shown to be beneficial for frequencies up to 1000 Hz. Francart et al. [1] have shown that pitch related tasks can benefit from the `f0mod` approach. Vandali et al. developed a training program for pitch and timbre discrimination improvement in [4]. Improvements were shown, but only in environments with little variations. But even with all available improvements, not all patients receive a comprehensible hearing sensation.

Another problem, this paper is addressing, is the impact of low brain plasticity from some of the patients. Brain plasticity means that the brain is able to adapt to new and different stimuli by changing its "wiring". While aging, the plasticity of the brain degrades. CI users of old age, who are severely hearing impaired, have thus problems when using their devices.

These problems led to the concept of adaptive pitch transposition (APT). The shallow insertion depth and therefore the reduced number of electrodes limit the accessible range of frequencies to a minimum. From there, it is obvious that a shift of the most interesting region of sound to the remaining frequencies should be a way to enhance speech perception. This also inherently reduces the influence of a low brain plasticity, since the main cues for comprehension are now always in focus.

3 Adaptive Pitch Transposition

The APT concept works as follows: It assumes that a particular patient has a certain voice pitch where he or she obtains the best recognition rates. This particular pitch is called comfort pitch. Then it transposes all incoming sound samples into that pitch frequency. This approach requires the following processing stages: (1) determine the current pitch, (2) determine the required shift amount, and (3) shift and transpose all incoming signals to the targeted best pitch frequency. The resulting transposed signal may then be relayed into an arbitrary CI simulation model.

The comfort pitch is the first necessity for each patient. This paper defines it as the one Pitch, with that a particular listener has the best speech recognition rates among a given set of pitches. The relationship of comfort pitch and voice pitch range are depicted in Fig. 1.

The comfort pitch can be determined by a patient test. This test could be comprised of voices of different pitch, uttering randomized phrases from a fixed set. Pitch-score pairs are then ranked for best speech recognition rates on the patients side. The voice pitch with the best score will be the new comfort pitch for optimal recognition when using the system. With this test, the comfort pitch adapts to the patient.

Since the goal is to transpose a given signal to a specific comfort pitch, the pitch of the current input signal has to be determined to calculate a transposition coefficient, which in turn defines how much shift has to be applied.

The Autocorrelation function (ACF) is an algorithm for this very purpose. Pitch is defined by the signals fundamental frequency and its harmonics, which are equally spaced in the frequency spectrum. The ACF slides a portion of data with a varying offset lag over itself to determine a similarity measure. This measure can then be interpreted as fundamental frequency under certain conditions.

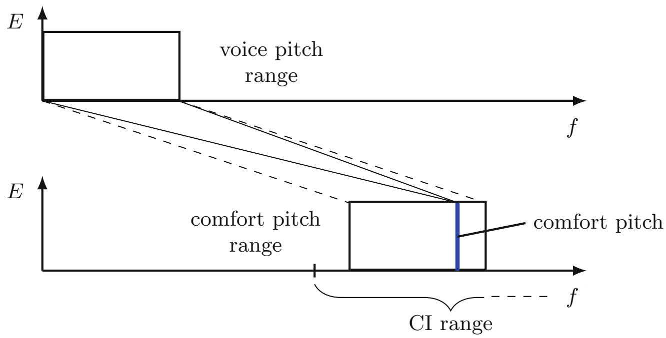

Fig. 1. Different pitch ranges in the frequency domain. The voice range of possible voice pitches is spacial disjunct from the range, a CI can address. Therefore the pitch has to be transposed into the CI range.

For input signals which actually have a pitch, the ACF gives a good estimate. These signals are called *voiced*. However, for signals that do not have a pitch (*unvoiced* signals), the ACF results appear pseudo-random. Therefore, it is necessary to verify these results. This process is called voice-activity-detection (VAD) for speech audio signals, and is used in speech recognition software or other speech related tasks.

Comfort pitch and determined pitch now allow for the calculation of the transposition coefficient.

Pitch transposition is usually seen as a combination of time stretching and re-sampling of an audio signal. First the signal is time-stretched while maintaining the original signal pitch. Then it is re-sampled to the original signal length. A signal stretched by the factor of two, re-sampled to the original length, will thus have a pitch at the double frequency. As an example one can imagine a glass of water, where the fill level of the water is the pitch and the diameter of the glass is the signal length. If we can make the diameter bigger while keeping the fill level, and afterwards make the diameter small again, we will have a raised fill level (thus: pitch).

4 Conclusion

This paper has presented a concept for preprocessing signals for cochlear implants. The complexity of an input sound signal in terms of pitch gets reduced to benefit the brains ability to adapt to stimuli.

This approach is only applicable, if an improvement in speech recognition can be observed for specific pitches. The extent to which the proposed system enhances the speech recognition rates in patients is subject of future research.

One major drawback of the current approach is the loss of speaker differentiation through the pitch fixing. Nevertheless, in prospect of regaining the ability to understand speech, this seems negligible.

The APT concept has been prototypically implemented in Octave/Matlab. The resulting system has yet some drawbacks regarding quality and computing time, but it acts as a foundation for a testbed to evaluate the effect on speech recognition compared to a simulation of a pure speech processing algorithm. Furthermore, the comfort pitch constitutes an additional parameter to tune the efficacy of speech processing algorithms.

Future research will be directed to the evaluation and testing with hearing people as well as CI users. Performance and quality increase are desired for real time usage.

Lastly APT is a powerful preprocessing stage, which is adaptable to any other speech processing strategy. It thus yields the potential to theoretically enhance the speech recognition rate for every CI user.

Acknowledgments. The author thanks the welisa graduate school for its support. Part of this research was funded by the German Research Foundation (DFG) grant number GRK 1505.

References

1. Francart, T., Osses, A., Wouters, J.: Speech perception with f0mod, a cochlear implant pitch coding strategy. Int. J. Audiol. **54**(6), 424–432 (2015)
2. Laneau, J., Wouters, J., Moonen, M.: Improved music perception with explicit pitch coding in cochlear implants. Audiol. Neurotol. **11**(1), 38–52 (2006)
3. Shannon, R.V., Fu, Q.J., Galvin, J., Friesen, L.: Speech perception with cochlear implants. In: Zeng, F.-G., Popper, A.N., Fay, R.R. (eds.) Cochlear Implants: Auditory Prostheses and Electric Hearing, pp. 334–376. Springer, New York (2004)
4. Vandali, A., Sly, D., Cowan, R., van Hoesel, R.: Training of cochlear implant users to improve pitch perception in the presence of competing place cues. Ear Hear. **36**(2), e1–e13 (2015)
5. Wilson, B.S., Dorman, M.F.: Cochlear implants: a remarkable past and a brilliant future. Hear. Res. **242**(12), 3–21 (2008). http://www.sciencedirect.com/science/article/pii/S0378595508001251, frontiers of auditory prosthesis research: Implications for clinical practice
6. Zeng, F.G., Tang, Q., Lu, T.: Abnormal pitch perception produced by cochlear implant stimulation. PloS One **9**(2), e88662 (2014)