Analysis of cochlear implant vocoder simulation including the current spread effect in the presence of background noise

Kauê Werner[1], Rafael Chiea[2], Júlio A. Cordioli[3], Stephan Paul[4]

[1] *Laboratório de Vibrações e Acústica, UFSC, Florianópolis, Brasil, Email: kaue.werner@lva.ufsc.br*

[2] *Laboratório de Processamento Digital de Sinais, UFSC, Florianópolis, Brasil, Email: rafaelchiea@gmail.com*

[3] *Laboratório de Vibrações e Acústica, UFSC, Florianópolis, Brasil, Email: julio.cordioli@ufsc.br*

[4] *Laboratório de Vibrações e Acústica, UFSC, Florianópolis, Brasil, Email: stephan.paul@ufsc.br*

Introduction

A cochlear implant (CI) is an implantable device that restores auditory recognition of patients with severe to profound hearing loss. It extracts band-pass-filtered acoustic envelope information and modulates current pulse trains to stimulate the auditory nerve. Nowadays, the ability to recognize speech by CI recipients in quiet environments has achieved parity with normal hearing subjects (NH), but the presence of background noise greatly reduces the speech recognition performance in IC users.

A common simulation method used to acoustically represent speech recognition by CI users, that has been used extensively to study aspects of speech understanding, is the vocoder. It represents envelope information by modulating an acoustic carrier, discarding the acoustic temporal fine structure from the original waveform. These vocoder simulations have been used in listening tests with NH and have been shown by many to provide results consistent with the outcome of cochlear implants. However, this method fails to consider important parameters such as pulse shape, stimulation rate, channel selection and current spread, that may be strongly related to speech recognition.

The present work is therefore focused on the effect of current spread on speech recognition. The envelope information given by the vocoder was analysed using vocoder methods of simulation taking into account current spread in the cochlea and objectve metrics for speech recognition, in quiet and noisy speech conditions.

Current spread and vocoder

The current spread in CIs is the spatial spread of electrical field generated by current pulse stimulation along neuron populations. It is known that the number of neurons excited by an electrical stimulus increases considerably as the stimulus level increases. The effect also depends on the electrode configuration of the device. In a monopolar configuration (MP), current pulses are delivered to individual intra-cochlear electrodes with reference to a far-field ground. For bipolar configuration (BP), the stimulation occurs when a potential difference is created between neighbouring electrodes leading to a current flow between them.

Through experimental measurements, some authors have proposed models for the spatial current spread along the auditory nerve using exponential functions [1, 2, 3], given that the electric current at a position x is given by

$$I(x) = I_0 e^{-\frac{|x_0 - x|}{\lambda}}, \qquad (1)$$

where I_0 is the current pulse level, x_0 is the electrode position and λ is the length constant, that depends on the CI electrode configuration. The model demonstrates that as the stimulus current amplitude increases, both the width and the peak of the spatial current spread profile increase. Several studies reported measurements of current spread in bipolar and monopolar configurations. The length constant for bipolar configuration is 2 - 4mm and for monopolar 8 - 11mm [4].

A novel approach presented by Boghdady et al.[5], named neural-based vocoder, considers the CI pulse trains as the input that electrically stimulate neuron populations along the auditory nerve. Figure 1 presents a schematics of the proposed method. For each channel of the 22-electrode array, they used a local neuronal population of 47 to 48 neurons.

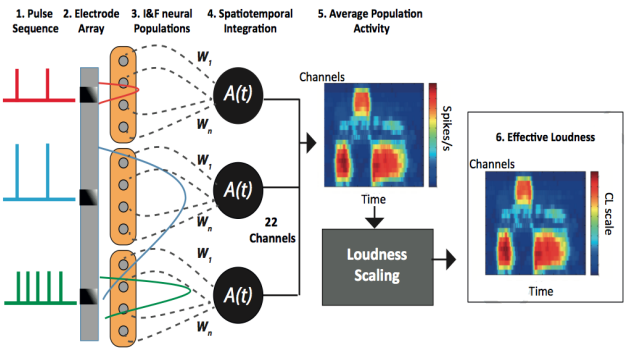

Figure 1: Neural-based vocoder simulation blocks [5]: (1) current pulse sequence as input; (2) implantable electrode array; (3) integrate and fire model applied in neural populations; (4) spatiotemporal integration; (5) average population activity for each channel; (6) effective loudness as temporal envelopes for each channel.

Current spread is modelled as a Gaussian distribution of the stimulus amplitude along neighbouring neuronal populations nearby each electrode. The *leaky integrate and fire* model was implemented to simulate spiking activity through a finite difference approach of the model's equation. After modelling the spiking activity for each

neuron, a total of 22 spatiotemporal integrators were calculated to obtain the average population activity, where each neuron had its synaptic weight related to the electrode position. A sigmoidal function was used for loudness scaling, considering a relationship with the average spike rate. The threshold of this function was empirically set to one quarter of the maximum firing rate among all neuronal populations. The extracted envelopes were used to modulate acoustic carriers in order to simulate CI recognition as the standard vocoder. Listening tests with NH subjects were carried out using consonant-vowel and cons.-vowel-cons. speech token recognition. They achieved considerable results for processed signals in quiet environment. The presence of background noise was not evaluated.

Speech intelligibility metrics

In some cases, when listening tests with NH subjects are not carried out, one must objectively evaluate speech intelligibility. Modulation spectrum metrics can be used as a feasable way to evaluate the envelope information of speech signals. The Speech Reverberant Modulation Ratio (SRMR-NH) [6] is a non-intrusive metric developed originally for reverberant and dereverberated speech and evaluated against subjective NH listener data. The metric is divided in the following signal processing procedures, where the input passes through: acoustic gammatone filterbank; temporal envelope and modulation spectrum; modulation filterbank and ratio computation of energy between groups of modulation bands.

An adaptation of SRMR tailored for CI users was proposed by Santos et. al [7]. For this metric, a CI filterbank was applied instead of the gammatone filters, and a different frequency range was considered for the modulation filterbank (2 - 64Hz). Listening tests were carried out with eleven CI users (fitted with the Nucleus Freedom device) [8] and after acquiring the results the authors proposed models to fit speech intelligibility and SRMR-CI, under different Signal-to-Noise Ratios (SNR) and Reverberant conditions with high correlation.

Analysis procedure

The analysis procedure of the current work can be seen in Figure 2. The same audio bank of 100 English sentences used by [8] to evaluate the SRMR metrics was considered here. The bank was divided in 5 groups of 20 sentences, where each group was set to a different SNR condition: -5dB, 0dB, 5dB, 10dB, ∞ (clean). Standard and neural-based vocoder simulations were implemented and the evaluation of current spread was only considered in the latter one. For both simulations, two types of carrier were used: white noise and a harmonic complex. The SRMR-NH was calculated for all processed and unprocessed signals and SRMR-CI was calculated for all unprocessed signals.

For the standard vocoder (SV) simulation, the input signal passes through a filterbank of 22 channels, similar to the one used in the Nucleus Freedom device from

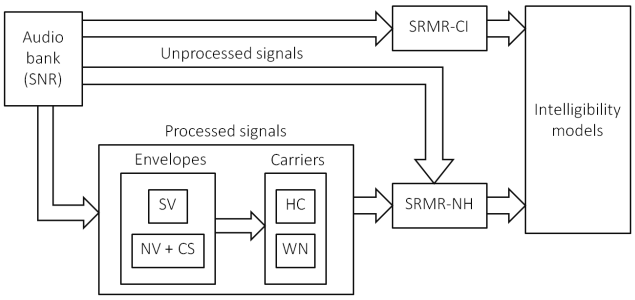

Figure 2: Analysis procedure for objective evaluation of speech intelligibility using vocoder simulations. SV: standard vocoder; NV: neural-based vocoder; HC: harmonic complex; WN: white noise.

Cochlear$^{\text{TM}}$, and the envelope was extracted from each one using a Hilbert transform. Two types of carriers were used: white noise (WN) and harmonic complex (HC),being the latter the same HC used in [9] (fundamental frequency of 100Hz, 240 harmonics with 100Hz of increment and maximally dispersed phase between them as a normal distribution between 0 and 2π). The carriers were bandpass filtered using 22 gammatone synthesis filters and modulated by the envelope information of each channel. The synthesis filters for the envelope modulation were extracted from the gammatone filterbank of the SRMR-NH. The sum of all channels was defined as the processed output signal (or simulated).

For the neural-based vocoder (NV) implementation, all parameters used in [5] were considered and applied. The audio input signal passed through a CI simulation of the ACE strategy, with 22 channels, 10 maxima and 900pps. The output was a set of current pulses used as the input for the neural model. Both Gaussian and exponential current spread were considered. The same two types of carriers were used and bandpass filtered as in the standard vocoder. They were modulated by the loudness envelope obtained with the neural-based model for each neuronal population.

Results

The normalized results of SRMR-NH with respect to the unprocessed signal, for SNR $= \infty$ (clean condition), can be seen in Figures 3 and 4. Figure 3 shows the results for the different vocoder-processed signals: the standard vocoder, the neural-based vocoder with gaussian distribution, and the neural-based vocoder with exponential distribution; using both carriers. The length constant of the exponential distribution was set to 2.8mm (correlates to a bipolar configuration). It can be seen that the harmonic complex carrier provides a higher modulation ratio when compared to white noise, for all vocoder methods. White noise produces a higher variation between methods than the harmonic complex. Regarding modulation energy, the harmonic complex carrier leads to a better correlation of the neural-vocoder with the standard vocoder, without background noise. Figure 4 shows the results for different length constants in the neural-vocoder with exponential spread. For both car-

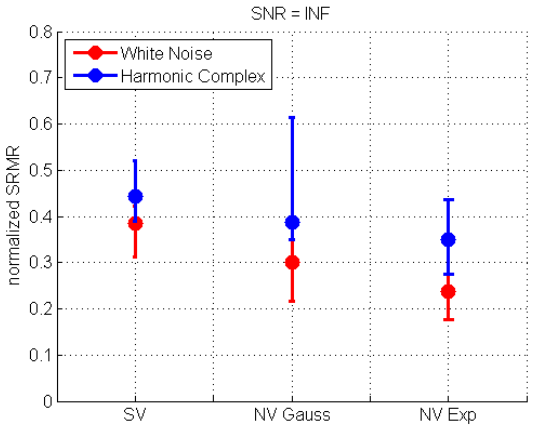

Figure 3: Normalized SRMR (with respect to unprocessed signal) results for the different vocoder-processed signals in clean condition: the standard vocoder, the neural-based vocoder with gaussian distribution, and the neural-based vocoder with exponential distribution; using harmonic complex (blue) and white noise (red) as carriers.

ries, the modulation ratio did not present considerable differences when varying the length constants. From bipolar to monopolar length constants, it presented a decay of around 5% of the unprocessed modulation ratio.

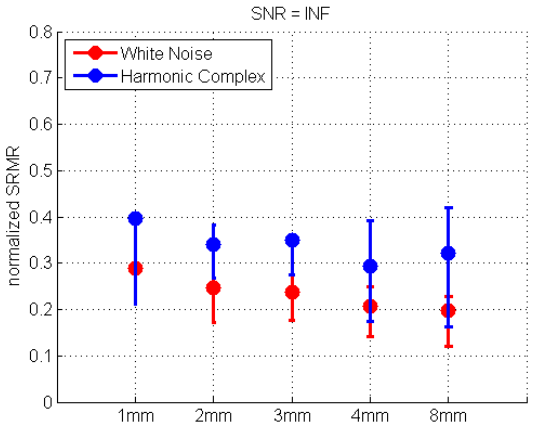

Figure 4: Normalized SRMR (with respect to unprocessed signal) results for different length constants in the neural-vocoder with exponential spread, also using harmonic complex (blue) and white noise (red) as carrier.

The normalized results of SRMR with respect to the clean condition of each method can be seen in Figures 5 and 6, for each carrier. The idea was to observe the normalized modulation ratio variation when decreasing SNR levels. For both carriers, and both current spread models, the neural vocoder did not represent well the variation of the envelope information, as the metrics predict for the unprocessed signals. The standard vocoder presented a lower energy decay rate when compared to both unprocessed results. In order to obtain a prediction of the speech intelligibility of the analysed vocoder methods, we used the models proposed for the metrics.

It can be seen in Figures 7 and 8 that for clean signals, all vocoder methods have achieved a better intelligibility using harmonic complex as carrier, when compared to white noise. When applying the noise background, neural-vocoder presented unintelligible sentences if evaluated using SRMR.

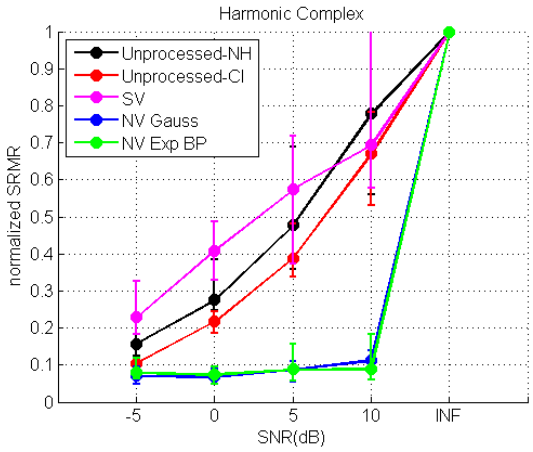

Figure 5: Normalized results of SRMR-NH and SRMR-CI with respect to the clean condition of each method using band-filtered harmonic complex as a carrier: (black) SRMR-NH for the unprocessed signal; (red) SRMR-CI for the unprocessed signal; (pink) SRMR-NH for the standard vocoder, (blue) for the neural vocoder using gaussian distribution; and (green) for the neural vocoder using exponential distribution, with length constant of 2.8mm.

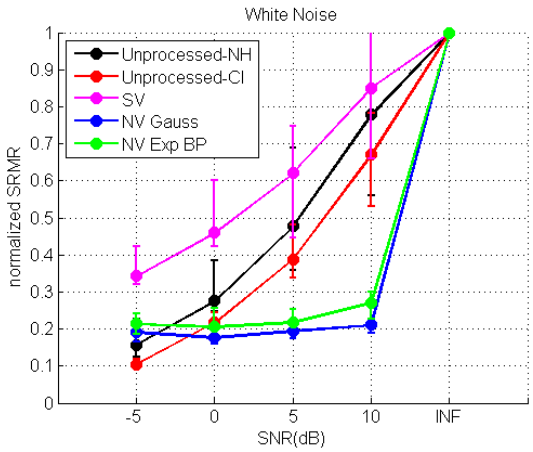

Figure 6: Normalized results of SRMR-NH and SRMR-CI with respect to the clean condition of each method using band-filtered white noise as a carrier: (black) SRMR-NH for the unprocessed signal; (red) SRMR-CI for the unprocessed signal; (pink) SRMR-NH for the standard vocoder, (blue) for the neural vocoder using gaussian distribution; and (green) for the neural vocoder using exponential distribution, with length constant of 2.8mm.

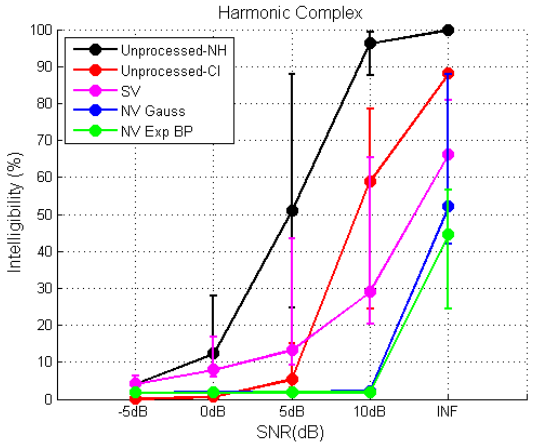

Figure 7: Intelligibility predictions using band-passed harmonic complex as vocoder carrier: (black) NH intelligibility of unprocessed signals; (red) CI user intelligibility of unprocessed signals; (pink) NH intelligibility of the standard vocoder-processed signals; (blue) NH intelligibility of the neural vocoder-processed signals using gaussian distribution; and (green) NH intelligibility of the neural vocoder-processed signals using exponential distribution, with length constant of 2.8mm.

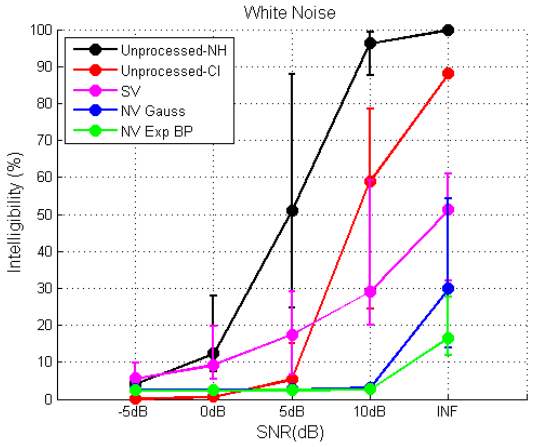

Figure 8: Intelligibility predictions using band-passed white noise as vocoder carrier: (black) NH intelligibility of unprocessed signals; (red) CI user intelligibility of unprocessed signals; (pink) NH intelligibility of the standard vocoder-processed signals; (blue) NH intelligibility of the neural vocoder-processed signals using gaussian distribution; and (green) NH intelligibility of the neural vocoder-processed signals using exponential distribution, with length constant of 2.8mm.

Conclusions

There was a difference between harmonic complex and white noise carries for the SRMR in clean signals, where the first had the best intelligibility prediction results for all vocoder methods. None of the current spread methods lead to a feasible representation of envelope information in the presence of background noise when compared to SRMR of unprocessed signals. The neural-based vocoder must be improved in order to provide a better representation of the current spread effect. The loudness scaling function is not precisely represented, one should consider a variation of the threshold and sigmoidal shape between neurons regarding their locations along the auditory nerve.

Acknowledgements

We would like to thank CNPq for financial support; João Santos and Oldooz Hazrati for the shared data and assistance.

References

[1] Black, R. C., Clark, G. M., & Patrick, J. F. (1981). Current distribution measurements within the human cochlea. Biomedical Engineering, IEEE Transactions on, (10), 721-725.

[2] O'Leary, S. J., Black, R. C., & Clark, G. M. (1985). Current distributions in the cat cochlea: a modelling and electrophysiological study. Hearing research,18(3), 273-281.

[3] Cohen, L. T., Richardson, L. M., Saunders, E., & Cowan, R. S. (2003). Spatial spread of neural excitation in cochlear implant recipients: comparison of improved ECAP method and psychophysical forward masking. Hearing research, 179(1), 72-87.

[4] Bingabr, M., Espinoza-Varas, B., & Loizou, P. C. (2008). Simulating the effect of spread of excitation in cochlear implants. Hearing research, 241(1), 73-79.

[5] Boghdady, N. E., Kegel, A., Lai, W. K., & Dillier, N. (2016). A Neural-Based Vocoder Implementation for Evaluating Cochlear Implant Coding Strategies.Hearing research.

[6] Falk, T. H., Zheng, C., & Chan, W. Y. (2010). A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech. Audio, Speech, and Language Processing, IEEE Transactions on, 18(7), 1766-1774.

[7] Santos, J. F., & Falk, T. H. (2014). Updating the SRMR-CI metric for improved intelligibility prediction for cochlear implant users. Audio, Speech, and Language Processing, IEEE/ACM Transactions on, 22(12), 2197-2206.

[8] Chen, F., Hazrati, O., & Loizou, P. C. (2013). Predicting the intelligibility of reverberant speech for cochlear implant listeners with a non-intrusive intelligibility measure. Biomedical signal processing and control, 8(3), 311-314.

[9] Churchill, T. H., Kan, A., Goupell, M. J., Ihlefeld, A., & Litovsky, R. Y. (2014). Speech perception in noise with a harmonic complex excited vocoder.Journal of the Association for Research in Otolaryngology, 15(2), 265-278.