EN.520.680.01 Speech and Auditory Processing by Humans and Machines
Final Project, May 7, 2021
Yi-Ting Tsai

Cochlear Implant Sound Coding Strategy – Implementation and Evaluation

Introduction

Cochlear implants (CI) have been invented and developed for decades. Yet, as a mature technology, CI sound coding strategies have not evolved much during the recent years. A CI sound coding strategy works to convert input sound signals captured by the microphones into ordered electric pulses that are sent to stimulate different parts of the basilar membrane of an ear. It bypasses and replaces the normal acoustic hearing process. As a result, a good sound coding strategy directly contributes to the hearing quality of CI users. It is important to keep exploring the limitations of existing CI sound coding strategies and search for possible improvements.

This project aims at exploring the current CI technologies by implementing a CI sound coding strategy that is based on the most popular and basic sound coding strategy used today: the Advanced Combinational Encoder (ACE) strategy from Cochlear Limited. Since ACE strategy only favors the place theory of frequency coding in human hearing, an extra feature—utilization of the temporal theory inspired by other strategies will be added into the system. A synthesizer will also be implemented for evaluation purposes. By evaluating the performance of the implemented strategy, limitations of current CI sound coding strategies will be discussed.

System

The structure of the proposed implementation is as follows:

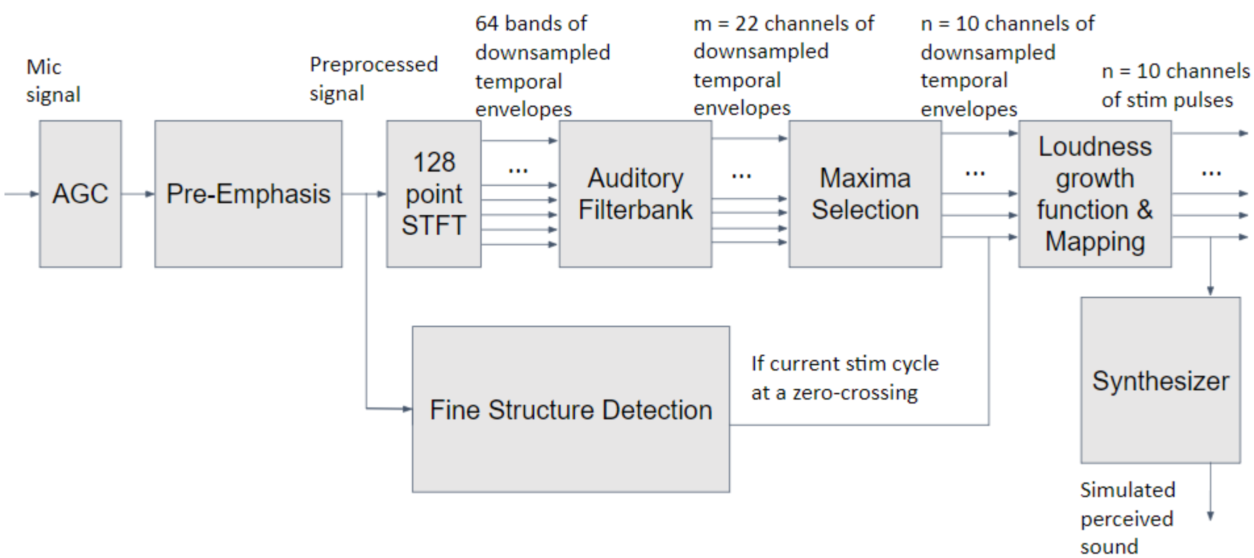

Figure 1 System Block Diagram

Within the system, fine structure detection and synthesizer are two new features that do not belong to the ACE strategy.

For the ACE part, basically the following steps are needed: automatic gain control (AGC), pre-emphasis, frequency analysis based on human auditory perception, temporal envelope extraction on each frequency channel, selection of frequency channels with the largest amplitudes, pulse modulation (done during frequency analysis in this implementation), non-linear mapping of channel amplitudes to stimulation current levels.

ACE is a so-called n-of-m strategy—out of m = 22 activated electrodes, n = 10 electrodes will be selected and stimulated in each stimulation cycle. The default single-channel stimulation rate will be set to 2000 pps.

Implementation

Implementation is done in Matlab. A sampling rate of 16 kHz is assumed in all input signals.

AGC

CI systems need AGC to compress the large dynamic range (~120 dB) of the acoustic environment into a smaller input dynamic range for processing and electrical stimulation [1]. Since AGC is not the focus of this project, a simple normalization is used instead. Assume the input dynamic range we want is 75 dB (30 dBSPL~105 dBSPL). Assign *xmax* (the signal amplitude that corresponds to “105 dBSPL”) as 1, then *xmin* (the signal amplitude that corresponds to “30 dBSPL”) will become 0.000178:

$$20log_{10}\left(\frac{1}{0.000178}\right) = 75\ dB$$

Parameters *xmin* and *xmax* will be used later in the electrical current level mapping step. Using the same concept, in order to simulate the volume of normal conversations (~55 dBSPL), the maximum amplitude of all input signals will be normalized to 0.0025 in Matlab.

Pre-Emphasis

Since speech signals typically have higher energy in low-frequency components, without the varying thresholds of perception provided by human ears, we need to flatten the energy presented to the frequency analysis filterbank. Pre-emphasis is applied here by attenuating the frequency components below 2 kHz using an IIR high-pass filter. This will also balance the energy ratio between vowels and consonants, while reducing the dominating low frequency noise [1].

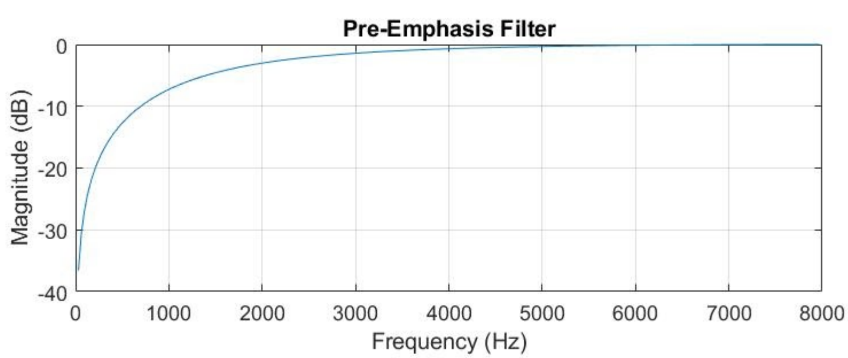

Figure 2 Pre-Emphasis Filter Frequency Response

Frequency Analysis & Temporal Envelope Extraction

The input signal is analyzed with a 128-point fast Fourier transform (FFT), and the temporal envelopes of all the resulting bins are estimated, based upon the definition of the envelope in the Hilbert transform, by calculating the magnitude of the complex spectrums [2]. Envelopes of each bin are then combined into 22 separate channels through RMS, matching the frequency distribution of ACE's 22 electrodes.

Channel no.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Lower cut-off frequency (Hz)	188	313	438	563	688	813	938	1063	1188	1313	1563	1813	2063	1313	2688	3063	3563	4063	4688	5313	6063	6938
Center Frequency (Hz)	**250**	**375**	**500**	**625**	**750**	**875**	**1000**	**1125**	**1250**	**1438**	**1688**	**1938**	**2188**	**2500**	**2875**	**3313**	**3813**	**4375**	**5000**	**5688**	**6500**	**7438**
Higher cut-off frequency (Hz)	313	438	563	688	813	938	1063	1188	1313	1563	1813	2063	1313	2688	3063	3563	4063	4688	5313	6063	6938	7938
Band-width (Hz)	125	125	125	125	125	125	125	125	125	250	250	250	250	375	375	500	500	625	625	750	875	1000

Figure 3 ACE 22 electrodes frequency distribution [3]

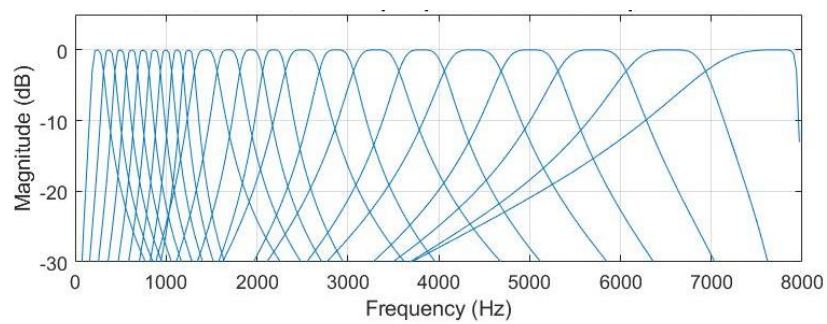

Figure 4 ACE 22 electrodes frequency distribution - 2

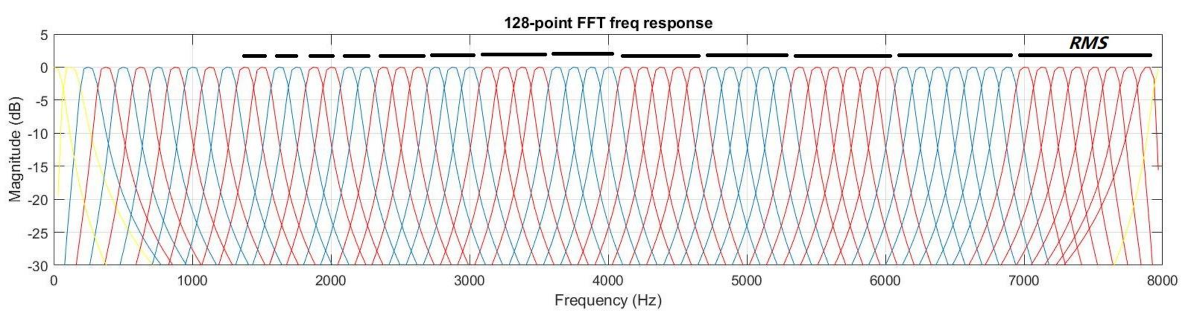

Figure 5 Frequency bands derived from 128-point FFT, grouped into 22 channels (blue and red)

By setting the FFT hop size to 8 (0.5 ms), each analysis frame is overlapped with the previous frame to make the envelope sample rate the same as the 2000 pps single-channel stimulation rate.

Maxima Selection

The channel selection method for the ACE strategy is called maxima selection. The maxima selection block scans the amplitudes of the channel envelopes and selects the channels with the highest amplitudes [1]. In this implementation, n = 10 out of m = 22 channels will be selected each stimulation cycle. Maxima selection is good for ignoring unimportant noise, as well as enabling a higher single-channel stimulation rate. Electrodes are stimulated sequentially to avoid interferences in ACE. Since the total stimulation rate is fixed (32,000 pps for Cochlear Nucleus 7 [4]), the amount of time spent in each stimulation cycle is limited. If all 22 electrodes are to be stimulated each cycle, the single-channel stimulation rate can only stay at 1400 pps.

Mapping

To match human perception of changes in loudness, mapping of signal amplitudes to electrical current levels is done using a loudness growth function [5].

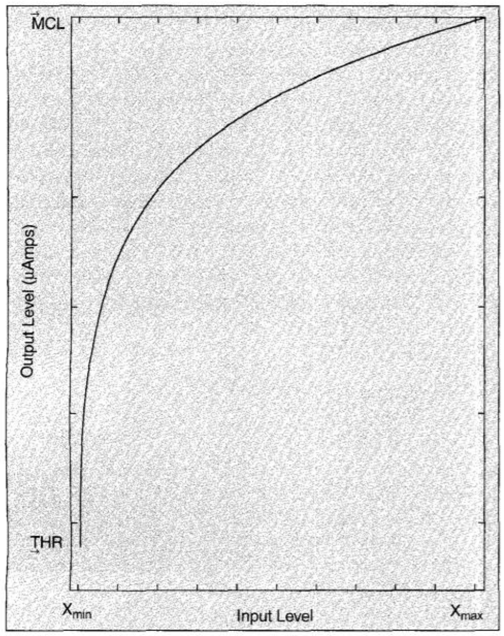

Figure 6 Loudness growth function, mapping from [xmin xmax] to [THR MCL]

Current levels are expressed in integers between 0~255 (each corresponds to a value in μA). Usually, the threshold level *THR* or *T* is set to 220, and the most comfortable level *MCL* or *C* is set to 100 [6]. With an input amplitude *x*, the mapping function will be:

$$\mathrm{Y} = \begin{cases} 0, & if\ x < xmin \\ C, & if\ x > xmax \\ A\log(x) + B, & otherwise \end{cases}$$

$$where \quad A = \frac{C - T}{\log(xmax) - \log(xmin)},$$

$$B = T - A\, log(xmin)$$

Fine Structure Detection

Inspired by F0mod strategy [2], the original attempt to incorporate temporal fine structures into ACE was to estimate the fundamental frequency F0 of the input signal through cepstrum analysis and change the single-channel stimulation rate of electrode 1 to the detected F0.

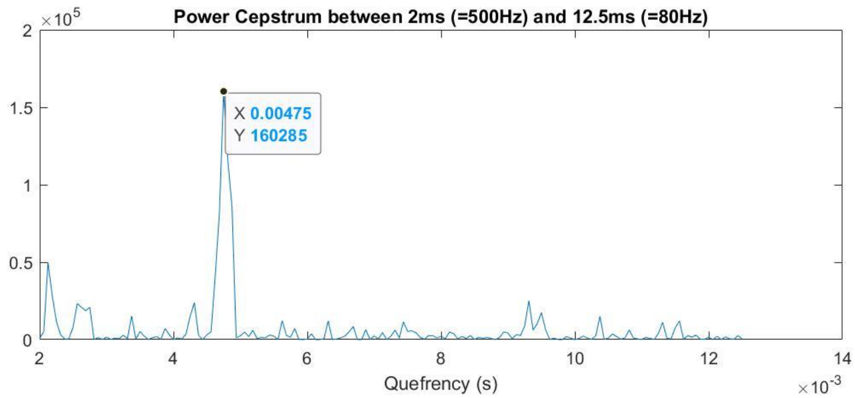

Figure 7 F0 Estimation through cepstrum analysis (working case)

However, the F0 detection results were unstable across different time frames without more complex algorithms. In addition, deriving cepstrums requires a large point FFT, which introduces time delays in the detection of changes in F0.

As a result, the original plan was not used. Inspired by the Channel-Specific Sampling Sequences (CSSS) technique used in the Fine Structure Processing (FSP) strategy [7] from MED-EL, stimulation pulses in electrode 1 will only be triggered by zero crossings in the downsampled (2000 Hz) band-pass filtered (frequency range the same as ch1 in ACE) input signal.

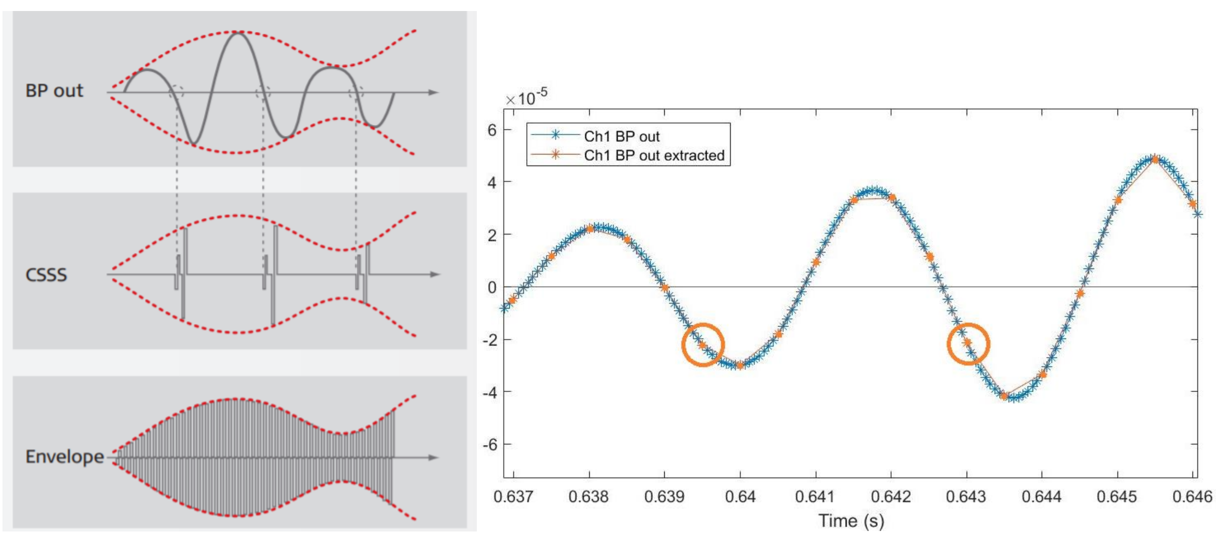

Figure 8 CSSS and zero crossing detection on downsampled BP output signal (triggered points are circled)

The specific process is as follows: (1) Input signal gets band-pass filtered with the same cutoff frequencies as channel 1 in ACE. (2) Downsample the band-pass filtered output signal to 2000 Hz. (3) Check if the downsampled band-pass filtered signal is at or next to a zero crossing in the current time frame. (4) Zero the current time frame's electrode 1 output stimulation data if not.

Synthesizer

Inspired by vocoders, a synthesizer is implemented to simulate what CI users perceive. Two methods, noise vocoder and tone vocoder, are tried.

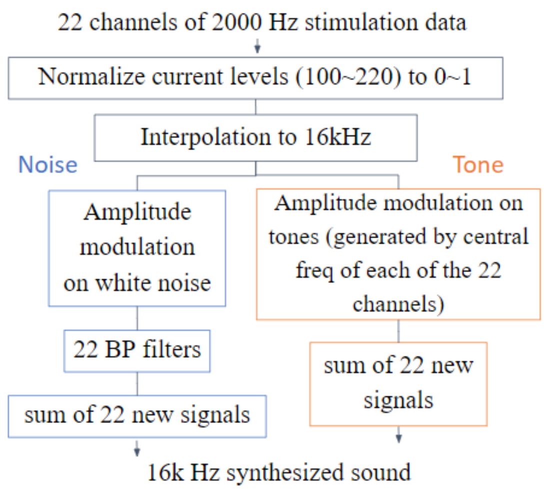

Figure 9 Synthesizer flowchart

The tone vocoder method achieves better sound quality, so it is used in the final system. During the synthesis process, it is also found that the temporal fine structure added into the stimulation data cannot be incorporated into the synthesized signal properly. If fine structure detection is turned on, since the carrier frequency is close to the modified single-channel stimulation rate, new low-frequency components will be generated during the amplitude modulation part of the synthesizer (see Fig. 11).

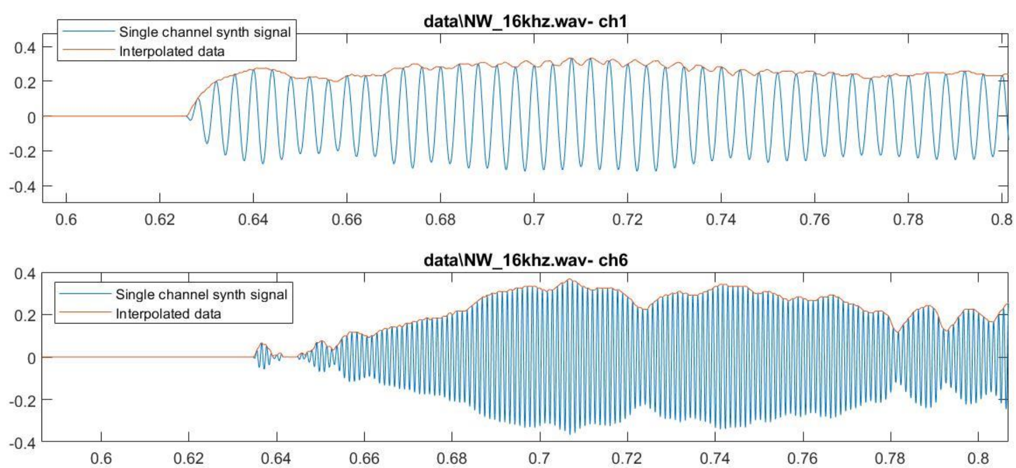

Figure 10 Amplitude modulations on tones for ch1 and ch6, without fine structure detection

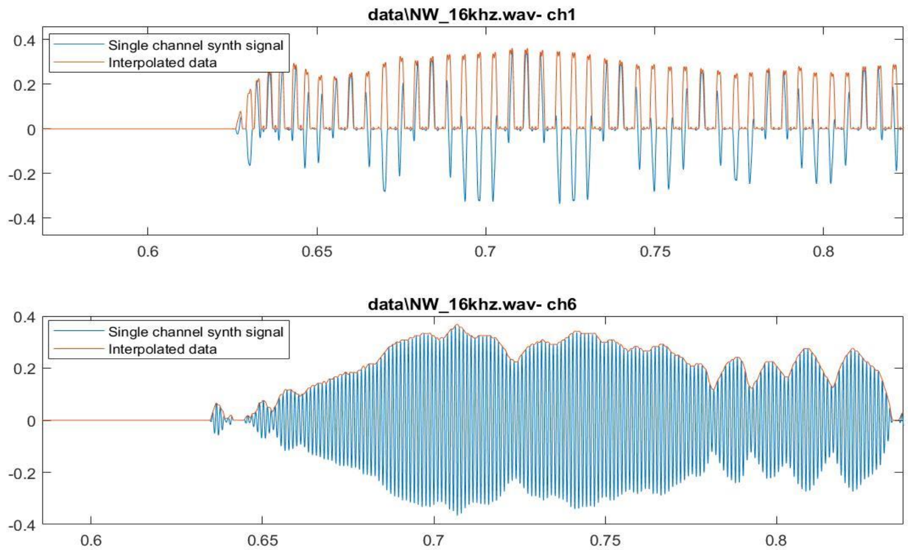

Figure 11 Amplitude modulations on tones for ch1 and ch6, with fine structure detection

Results and Evaluation

Testing the system with an input speech signal "Northwest", the following results are achieved.

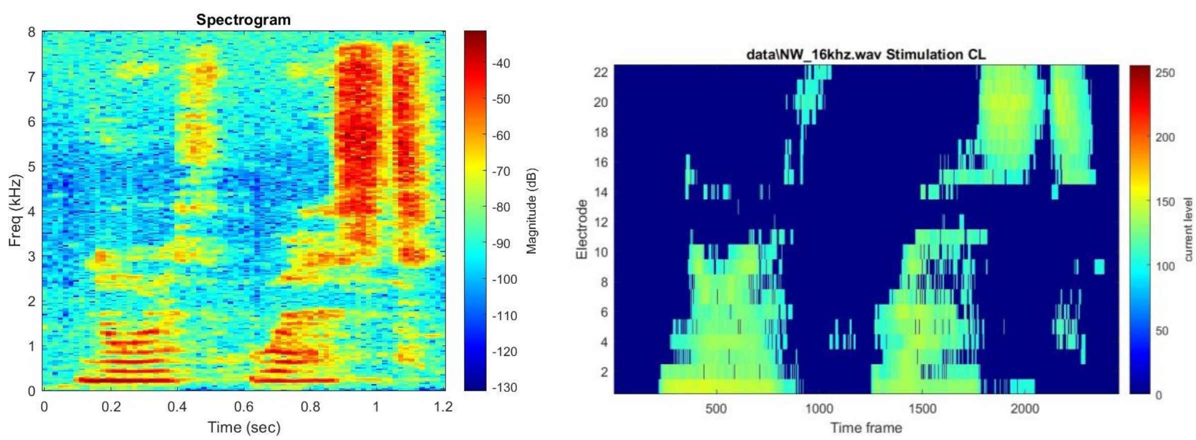

Figure 12 Input signal spectrogram vs 22-channel stimulation data

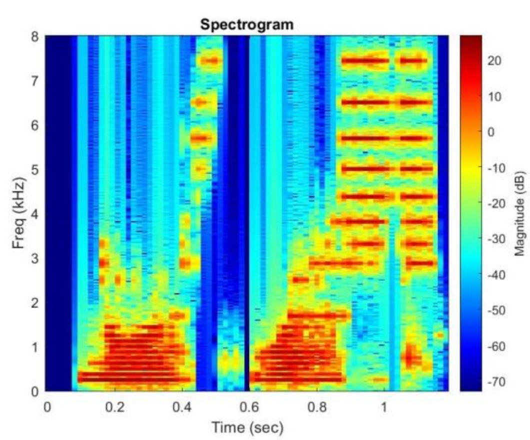

Figure 13 Synthesized signal spectrogram

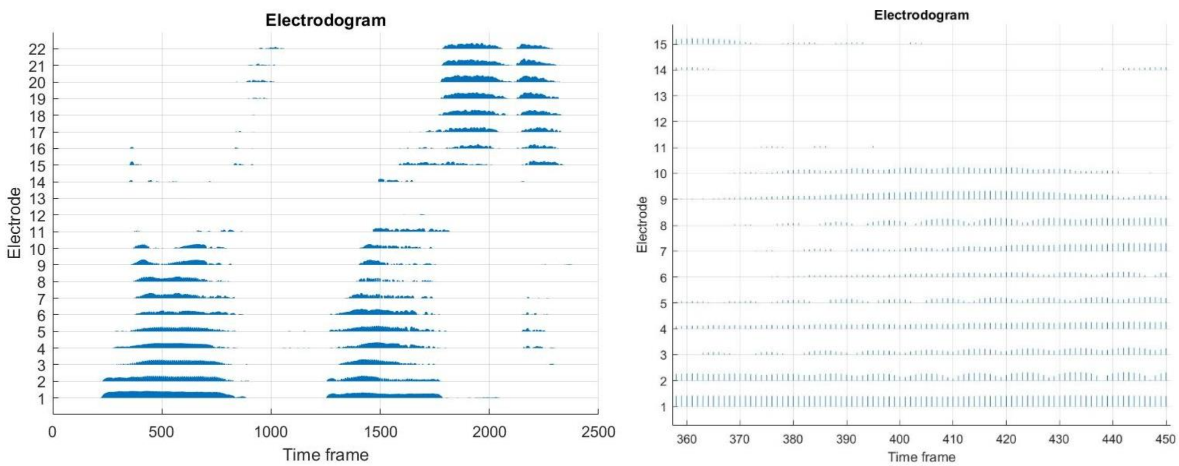

Figure 14 22-channel stimulation data electrodogram and the enlarged version

With the addition of fine structure detection:

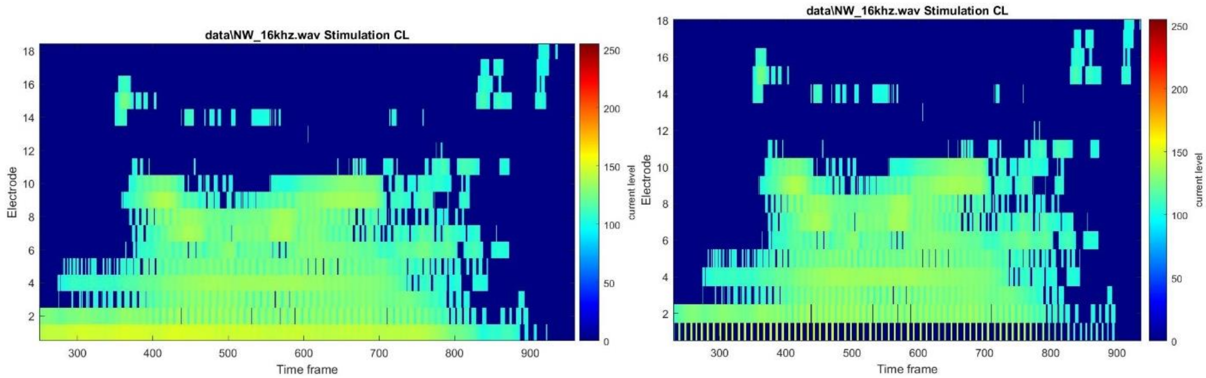

Figure 15 Enlarged 22-channel stimulation data without fine structure vs with fine structure

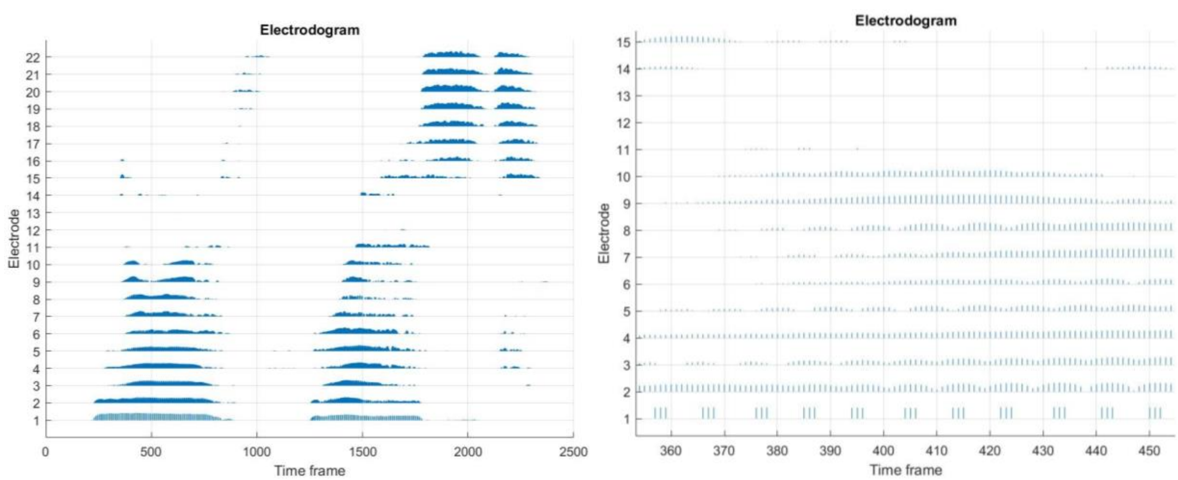

Figure 16 22-channel stimulation data electrodogram and the enlarged version – with temporal fine structure

Listening to different synthesized signals, it is found that the speech intelligibility of the strategy output is stable and good (as shown in the above plots too), and some differences between male and female speeches can be perceived. However, tones and pitches are not obvious, and it is hard to identify speakers more specifically. The effect of adding temporal fine structures into the stimulation data cannot be evaluated either because of the problem mentioned above in the Synthesizer section. On the other hand, the maxima selection technique in ACE is effective in ignoring some noise, as shown in the figures below. Yet, CI strategies can only do so much in noise reduction.

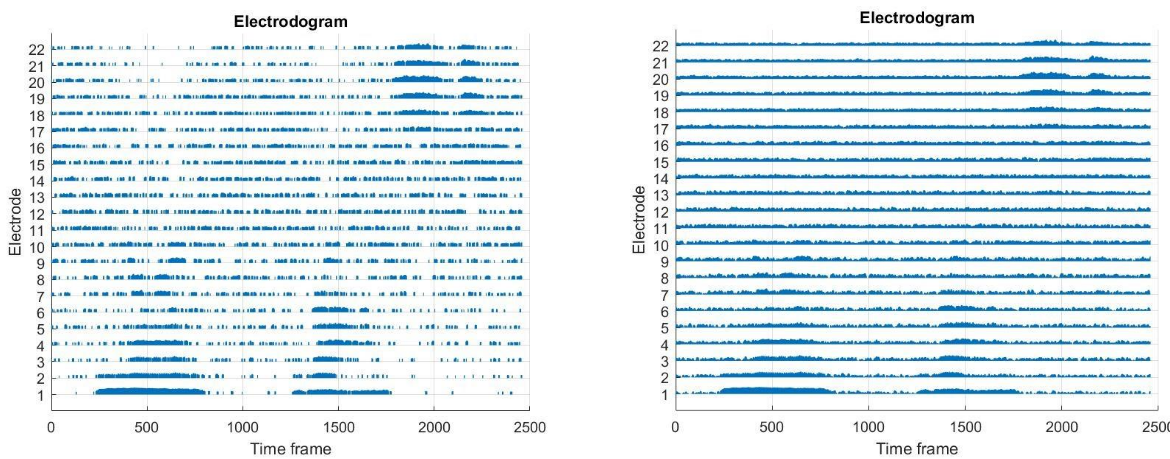

Figure 17 Electrodograms of noisy input (5 dBSNR pink noise) with maxima selection vs without maxima selection

Conclusions

Speech intelligibility is not a problem in the current CI strategies. Poor expression of pitch and tone information is the key limitation. CI users' perception of tonal languages and music can be greatly improved if the problem can be addressed. Given a lot of physical and biomedical restrictions (e.g. number of electrodes that can be inserted into the cochlea or stimulated at the same time), this can only be done by incorporating as much crucial information as possible into the stimulation data, which means both the place and the temporal theories should be utilized.

Adding temporal fine structures into CI stimulation data is feasible, and may be the key to improve tone and pitch perception. However, since it is hard to evaluate methods related to temporal fine structures, standard evaluation procedures should be researched and developed first.

Lastly, besides improvements in CI strategies, effective pre-processing algorithms such as noise reduction should still be further developed and relied on.

Reference

[1] P. P. Khing, *Gain Optimization for Cochlear Implant,* Sydney: The University of New South Wales, 2013.

[2] Johan Laneau, Jan Wouters, Marc Moonen, "Improved Music Perception with Explicit Pitch Coding in Cochlear Implants," *Audiology and Neurotology,* vol. 11, pp. 38-52, 2006.

[3] Hussnain Ali, Jack H. Noble, René H. Gifford, "IMAGE-GUIDED CUSTOMIZATION OF FREQUENCY-PLACE MAPPING IN COCHLEAR IMPLANTS," in *ICASSP*, 2015.

[4] "Cochlear Implant Comparison Chart," Apr 2021. [Online]. Available: https://cochlearimplanthelp.files.wordpress.com/2021/04/cochlearimplantcomparisonchart_v11.3a.pdf. [Accessed 07 May 2021].

[5] P. C. Loizou, "Mimicking the Human Ear," *IEEE SIGNAL PROCESSING MAGAZINE,* pp. 101-130, 1998.

[6] A. Zarowskia, A. Moliszb, L. De Coninck, "Influence of the pre- or postlingual status of cochlear implant recipients on," *International Journal of Pediatric Otorhinolaryngology,* Jan 2020.

[7] "FineHearing™ Technology," [Online]. Available: https://s3.medel.com/downloadmanager/downloads/maestro/en-US/FocusOn_FineHearing.pdf. [Accessed 07 May 2021].