

# A Neural Excitability Based Coding Strategy for Cochlear Implants

W. K. Lai<sup>1,2</sup>, N. Dillier<sup>1</sup>, M. Killian<sup>3</sup>

<sup>1</sup>ENT Clinic, University Hospital, Zürich, Switzerland; <sup>2</sup>Sydney Cochlear Implant Centre, RIDBC, Sydney, Australia;

<sup>3</sup>Cochlear Technology Centre, Mechelen, Belgium

**Correspondence to:** W. K. Lai, [waikong.lai@scic.org.au](mailto:waikong.lai@scic.org.au)

**Keywords:** Cochlear Implants, Speech Coding, Auditory, Neural Excitability, Channel Interaction

**Received:** June 13, 2018

**Accepted:** July 17, 2018

**Published:** July 20, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## ABSTRACT

A novel cochlear implant coding strategy based on the neural excitability has been developed and implemented using Matlab/Simulink. Unlike present day coding strategies, the Excitability Controlled Coding (ECC) strategy uses a model of the excitability state of the target neural population to determine its stimulus selection, with the aim of more efficient stimulation as well as reduced channel interaction. Central to the ECC algorithm is an excitability state model, which takes into account the supposed refractory behaviour of the stimulated neural populations. The excitability state, used to weight the input signal for selecting the stimuli, is estimated and updated after the presentation of each stimulus, and used iteratively in selecting the next stimulus. Additionally, ECC regulates the frequency of stimulation on a given channel as a function of the corresponding input stimulus intensity. Details of the model, implementation and results of benchtop plus subjective tests are presented and discussed. Compared to the Advanced Combination Encoder (ACE) strategy, ECC produces a better spectral representation of an input signal, and can potentially reduce channel interactions. Pilot test results from 4 CI recipients suggest that ECC may have some advantage over ACE for complex situations such as speech in noise, possibly due to ECC's ability to present more of the input spectral contents compared to ACE, which is restricted to a fixed number of maxima. The ECC strategy represents a neurophysiological approach that could potentially improve the perception of more complex sound patterns with cochlear implants.

## 1. INTRODUCTION

The task of designing an effective coding strategy for cochlear implants (CI) must consider various

limitations inherent to the CI system. Some of these limitations are system specific, while others are more general, for instance, the availability of a limited number of discrete stimulation sites, the reduced dynamic range arising from electrical stimulation (e.g. [1, 2]) or the accompanying electric field spread resulting thereof (e.g. [3, 4]). These limitations generally imply that compromises will arise in the resulting spectral and temporal resolution. Present day CI coding strategies have been quite successful because they have been able to suitably account for the main limitations. There remain, however, other limitations that, if addressed, could result in further improvements upon existing coding strategies.

CI coding strategy developments are often based on signal transmission concepts aimed at optimizing the amount of acoustic information transmitted through suitable conditioning and processing of the incoming acoustic signal. Generally, spectral information is encoded by the stimulation site and amplitude information by the stimulus intensity. An incoming signal to the processor unit is, after some form of conditioning, subjected to spectral analysis and the output then divided and aggregated into a number of channels corresponding to the number of available stimulation sites along the implanted electrode array. The energy content of each of these channels is then used to determine the respective intensity of the stimulus to be presented on the corresponding stimulation site. Typically, the stimuli are in the form of discrete charge-balanced biphasic pulses, although in earlier CI systems, analog stimuli have also been used (e.g. [5, 6]). To avoid interaction between the electrical fields of individual pulses, the stimuli are presented in temporally non-overlapping sequences [7].

The most straightforward approach is taken by the CIS (Continuous Interleaved Sampling) coding strategy (e.g. [8]) whereby the stimuli on each of the total of  $m$  channels are presented sequentially on the corresponding  $m$  electrodes. If the period required to present each set of  $m$  stimuli is defined as a stimulation frame, CIS repeats the above process for subsequent stimulation frames, using new input information for each new stimulation frame. The ACE (Advanced Combination Encoder) coding strategy (e.g. [9]) behaves similarly, but presents only a subset  $n$  of the highest energy channels (maxima) from the total of  $m$  channels, with the stimulation frame here being the period required to present each set of  $n$  stimuli.

In either approach, the stimulation rate does not directly encode the temporal information from the incoming signal. Instead, temporal information is indirectly encoded within the amplitude modulation of the stimuli presented on the individual channels. Other coding strategies have sought to encode the temporal information more directly into the stimuli by explicitly enhancing this amplitude modulation (e.g. MEM [10], F0mod [11], SAM [12], FAME [13], eTone [14] among others) or by encoding such information into the stimuli on specific stimulation channels (e.g. F0F1F2 [15], MPEAK MultiPeak [16], FSP Fine Structure Processing [17]).

The above approaches generally seek to optimize the information deduced from the incoming signal information so that specific acoustic features are either as well represented as possible or enhanced in the resulting stimulation patterns. The incoming signal itself could also be treated to improve the signal to noise ratio, either with pre-processing (e.g. directional microphones, beamformers, intelligent noise cancellation) or using more sophisticated techniques such as sparse non-negative matrix factorization [18].

However, compared to normal hearing, the spectral resolution is already severely reduced due to the limited number of stimulation sites available, necessitating frequency information to be aggregated into discrete channels. This reduction in the spectral resolution is compensated for as best as possible in the spectral channel mapping by ensuring that the useful range of frequencies of interest is tonotopically represented across the range of stimulation sites available. After the incoming signal information has been mapped onto the place of stimulation, the next step in the signal information pathway is the neural interface itself. Here, electrophysiological phenomena such as electric field spread and the reduced dynamic range associated with electrical stimulation will further compromise the fidelity of the signal information being transmitted. The reduced loudness dynamic range is compensated for by using a suitable loudness mapping function [19, 20]. The electric field spread is rarely directly accounted for, although it is known that switching from monopolar to bipolar stimulation modes reduces the electric field spread [3]. One exception is the MP3000™ coding strategy [21] which approximates psychophysical forward masking interactions to predict redundancies in the stimulation patterns. This resulted mainly in a reduction of the

energy consumption as a consequence of reducing the number of stimuli needed to represent a given amount of input signal information, whilst not affecting speech performance [22].

A further limitation of the neural interface is the capacity of the stimulated neural population to convey the encoded information. This is determined partly by the number of surviving spiral ganglion neurons and partly by their neurophysiological behaviour. In particular, the refractory behaviour, in which portions of a stimulated neural population are momentarily incapable of reacting to subsequent stimuli, implies that presenting stimuli to a particular neural population that is momentarily in absolute refractory state will be ineffective and consequently redundant. Instead, it would be more effective to stimulate other sites which are at that moment capable of reacting to stimuli and conveying the information of the incoming sound.

Taking into account neurophysiological factors such as the refractory behaviour could therefore potentially result in a more effective as well as more efficient coding strategy.

### **Excitability and Redundancy**

Generally, with a CI coding strategy, an input signal is analysed and divided into multiple frequency channels. The intensities of the stimuli to be presented are based on the energy content of the corresponding frequency channels. For an input signal consisting of frequency components that are close to one another, such as with harmonics based on the F0 of the sound source, or spectral envelopes like vowel formants, the channels with the most energy will cluster together in adjacent channels. The degree of clustering also depends on the width of the filters used for the frequency analysis, and the clinically used filterbanks tend to have relatively wide filters [23].

Depending on how frequently stimuli are presented on a given channel, the stimulated auditory nerve neurons will not necessarily respond to each and every stimulus due to the variability of the refractory properties. The ability of a given neuron or a given neural population to react to a stimulus is defined here as its excitability. A stimulus presented during the absolute refractory period of an excited neuron will be ineffective and is therefore redundant for this neuron. By extension, when a large proportion of the neural population close to a stimulation site is in a refractory state, a stimulus there will become less effective and ultimately redundant. Such stimuli can therefore be omitted, and it would be more effective to instead use that stimulus interval to present stimuli at alternative sites close to more excitable neural populations.

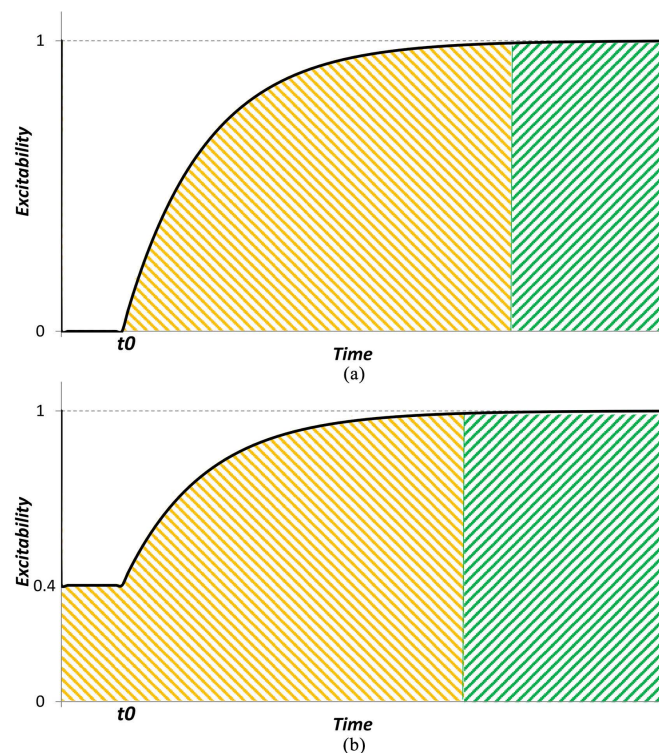
Electric field spread effects from a particular stimulus on neighbouring sites must also be accounted for, especially when the stimuli are clustered together both spatially and temporally. Depending on the stimulus intensity, neural populations associated with adjacent stimulation sites will also react to this stimulus, causing part of these neighbouring neural populations to be activated and thus driven into a refractory state as well.

This paper presents a new cochlear implant coding strategy, called Excitability Controlled Coding (ECC), in which the excitability of the spiral ganglion is modelled based on neurophysiological refractory properties of the neurons. The model also takes into account the electric field spread to calculate the excitability of the spiral ganglion population close to the stimulating intracochlear electrode array during active stimulation. The main distinguishing feature of this strategy is that the decision to present a stimulus on a given channel is based on a combination of the momentary state of that channel's neural excitability and the amplitude of the corresponding incoming sound signal. The aim of ECC is to improve the effectiveness of the stimuli that are actually selected for presentation. The ECC methodology is described and illustrated using the outputs of a Matlab implementation. Preliminary test results from a pilot study are also presented, and their implications discussed.

## **2. EXCITABILITY WEIGHTED STIMULUS SELECTION**

At the core of the ECC strategy is a model that computes the excitability state of the auditory nerve. The model divides the spiral ganglion into a number of neural populations corresponding to the number of stimulation electrodes of the cochlear implant electrode array. The excitability state of each population

is a time-dependent function that varies depending upon the stimulation signal. In its resting state, the population has 100 percent excitability, denoted as an excitability state of 1. When a neural population is stimulated by a stimulus of intensity  $A$  (which is also scaled between 0 and 1, as computed from the energy content of the corresponding channel), its excitability is reduced accordingly by the same amount  $A$ . Note that, depending on the initial excitability state  $X$ , there will be also be a portion  $(X - A)$  of the neural population that remains excitable. This remaining excitability is also defined as the remnant excitability. The portion  $A$  of the neural population which reacted to the stimulus is then driven into an absolute refractory state which remains constant for a fixed duration before the excitability begins to recover towards the resting state of 1. This is illustrated in **Figure 1** for two instances with  $A = 1$  and  $A = 0.6$  respectively. The excitability state at any given time is then computed from this time-dependent recovery function [24] and subsequently applied as a weighting to the corresponding input signal for this channel. The weighted input signals are then used to determine which channels will be expected to be most effective at any given time for presenting on the electrode array.



**Figure 1.** Logarithmic recovery functions of the form  $y = 1 - \exp(-\alpha(t - t_0))$  where  $\alpha$  is the inverse of the time constant and  $t_0$  is the absolute refractory interval, illustrating the recovery time course for a neural population that has been driven (a) fully and (b) partly into refractoriness by stimuli of intensity  $A = 1$  and  $A = 0.6$  respectively. In both cases, the excitability is reduced by an amount corresponding to the stimulus intensity  $A$  immediately after stimulation, followed by a flat segment (till  $t_0$ ) where the stimulated portion of neurons are in absolute refractoriness and therefore their excitability does not change. At the end of the absolute refractory interval  $t_0$ , the excitability begins to recover. The shaded area under the curve therefore represents the proportion of neurons available for stimulation. The orange shaded portion corresponds to the relative refractory interval, while the green shaded portion denotes full recovery of the neural population. In (b), only 0.6 of the neurons in the population are in a refractory state immediately after the stimulation and the excitability is reduced to  $(1 - 0.6) = 0.4$ . Note that  $\alpha$  is the same in both examples above.



In a system with  $m$  stimulation channels, there are  $m$  corresponding neural populations associated with and assumed to be close to the corresponding stimulation site. Each neural population has its own respective excitability state. Similarly, the input signal is divided into  $m$  frequency-band components corresponding to the respective stimulation channels. Stimuli are then selected one at a time from the input signal components, with each stimulus presented at regular time intervals corresponding to an overall stimulation rate of choice. Since these time intervals are known, the momentary excitability state of the system can be easily computed using the time dependent recovery function for any time interval. Prior to selecting any stimulus for presentation, the input signal components are weighted with their respective momentary excitability states. The highest weighted signal component is then selected for presentation on the electrode array.

Immediately after each stimulus, the excitability states of up to  $m$  affected channels are then modified. The extent to which the neural population of the stimulated channel as well as those of its neighbouring channels are affected will depend on the estimated electric field spread function associated with the stimulus intensity above. At the next stimulus interval, the excitability state is computed once more and again used for weighting the input signal components for this next interval, and the process then repeated.

### Regulating the Channel Stimulation Rate

Selecting the stimuli based on the neural population's excitability in the manner described above puts the various channels in competition against one another to be selected for stimulation at each time interval. The excitability of a previously stimulated channel will eventually recover over subsequent intervals, and depending on the combination of momentary excitability and input signal intensity used for the weighting, this same channel could be reselected for stimulation. The frequency of reselection of any given channel, in other words the channel stimulation rate, is generally variable, depending on its momentary weighted excitability and that of the other competing channels.

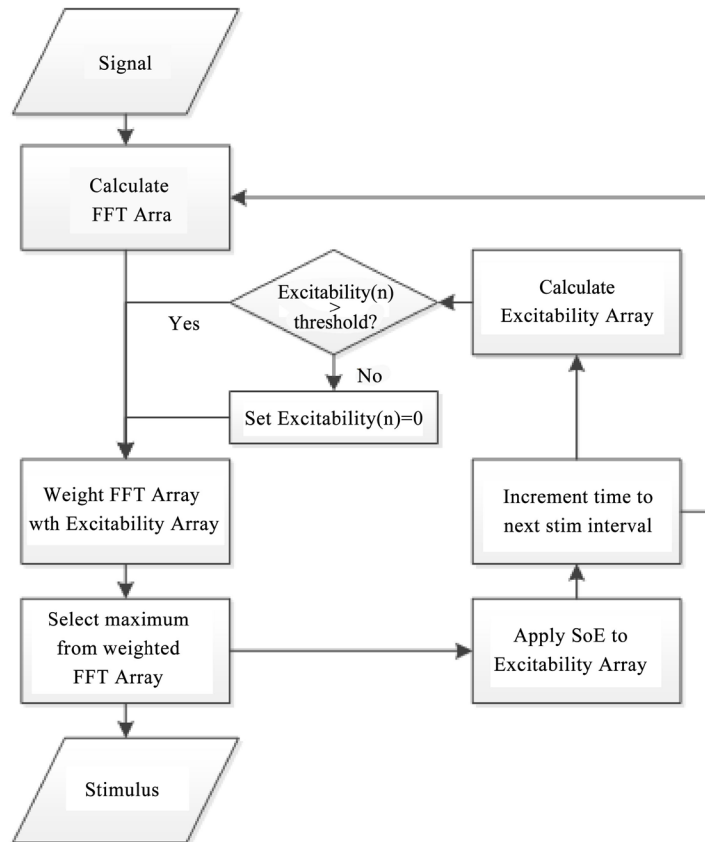
Selecting the channels based on the weighted excitability alone has one drawback, especially with sparse input signals that only activate very few channels. Because any channel with non-zero input signal intensity is eligible to compete for reselection whenever its excitability is also non-zero, an input signal on only a single channel, for instance, would be reselected every time its excitability recovers slightly above zero, regardless of its input signal intensity, due to the lack of competing channels. This effect diminishes as the number of competing channels is increased. To prevent this effect, a selection threshold dependent on a channel's input signal intensity is necessary. The excitability has then to exceed this threshold value before the corresponding channel is considered for selection. This would also allow channels with higher input signal intensities, which contain more information, to be represented more often, and vice-versa when the input signal is sparse. The iterative process of weighting, selection and updating of the excitability state, together with how the threshold affects the decision making process, is summarized in the flow chart in [Figure 2](#).

The threshold  $thr$  is set such that higher intensity signals have a lower threshold and vice-versa, allowing channels with higher signal intensities to be proportionately more likely to be reselected than lower signal intensity channels. This is implemented according to:

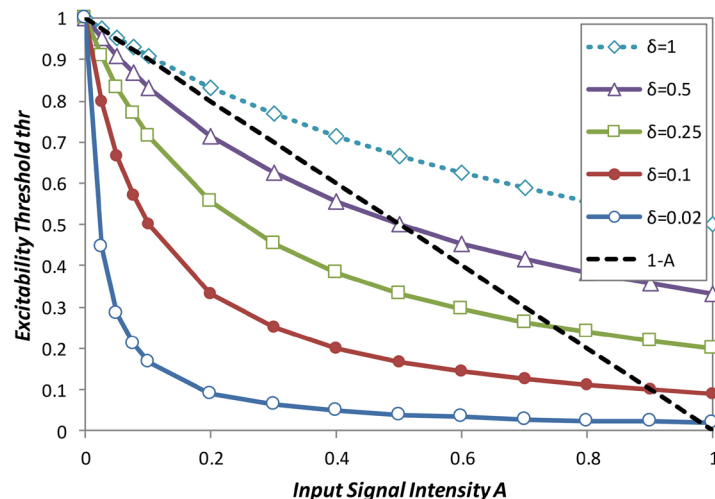
$$thr = \delta / (A + \delta) \quad (1)$$

where  $A$  is the stimulus intensity expressed as a ratio relative to the input dynamic range, and  $\delta$  is a constant which can be used to modify the function. For instance, in a system with an input dynamic range between 25 and 65 dB SPL, an input signal level of 35 dB would correspond to  $A = (35 - 25) / (65 - 25) = 10 / 40 = 0.25$ . The way the excitability threshold  $thr$  varies as a function of  $A$  is illustrated in [Figure 3](#) for different values of  $\delta$ .

When a channel is stimulated, its excitability will be reduced proportionately according to the stimulation intensity. With a weak stimulus, this poses a problem as the remnant excitability for that channel may still be greater than its corresponding  $thr$  threshold arising from that stimulus, thereby indicating that



**Figure 2.** Flow chart illustrating how the excitability is iteratively recomputed and used to weight the input signals for the stimulus selection process. The excitability of a given channel has to exceed a threshold before the channel is eligible to be reconsidered for selection in the next interval. Otherwise, the channel is excluded from the next selection interval by setting its excitability to zero.



**Figure 3.** Excitability threshold thr versus input signal intensity A functions for various values of  $\delta$ . thr values are lower for higher intensity signals, facilitating higher reselection probabilities and consequently higher stimulation rates, and vice versa. The input complement ( $1 - A$ ) is also shown here (dashed black line) for comparison.

this channel is still eligible for reselection in the following interval. If this happens, it would result, at least momentarily, in a very high stimulation rate on that channel, which is undesirable. To illustrate this, consider a single persistent low level input signal of say  $A = 0.2$  on a given channel, with  $\delta = 0.25$ . The threshold  $thr$  for subsequent intervals (since the input remains constant at  $A = 0.2$ ) is computed from (1) above as 0.556. After the initial selection of that channel, its excitability is reduced accordingly by  $A$ , *i.e.* from 1.0 to 0.8. In the following interval, the excitability (0.8) is still larger than  $thr$  (0.556) and will therefore result in another stimulus despite the fact that the channel actually contains a low level input signal which ought to result in less frequent stimulation.

Thus, the threshold  $thr$  alone is not sufficient to account for instances with low signal input levels. To specifically prevent the above from happening, a further threshold condition can be defined. In the example above, the total excitability must also exceed the value of 0.8 or more generally,  $(1 - A)$ , before the channel is eligible for reselection. The term  $(1 - A)$  can also be called the “input-complement”. A stimulus is then only generated when the corresponding excitability exceeds both  $thr$  and the input-complement. Together,  $thr$  and the input-complement define a selection threshold that provides the necessary differentiation, in terms of the stimulation rate, between channels that are stimulated at different intensities. The input-complement threshold is also illustrated in [Figure 3](#).

### 3. IMPLEMENTATION

#### 3.1. Description of Excitability Model

The algorithm for implementing the ECC strategy is based on the description in Patent WO2009/143553A1 [25]. The central feature is a neural excitability variable associated with each stimulation channel, and these excitability values will be tracked over time at every stimulation interval. The excitability variable is thus persistent over time, allowing subsequent stimulations on any channel to also be accounted for. The model assumes that stimuli are presented at regular time intervals corresponding to an overall stimulation rate of choice. At any given stimulation interval, the channel with the highest weighted combination of excitability and corresponding input signal intensity will be selected for presenting the stimulus. This selection process based on the excitability-weighted input signal is repeated for every stimulation interval.

When a channel is selected for stimulation, its neural excitability is initially reduced in the following interval but this excitability will gradually recover to 100 percent over subsequent time intervals. This recovery function is modelled after the refractory properties of a stimulated neural population, incorporating an “absolute refractory” period, where the tissue is not excitable and a “relative refractory” period over which the excitability recovers to full excitability. Note that a neurophysiological based recovery function was chosen here to reflect the neuronal nature of the excitability considerations behind the ECC strategy, but in practice, any other similar time-varying function would also be usable. [Figure 1\(a\)](#) shows the logarithmic recovery function used for ECC which mimics the recovery functions found in CI patients [24]. The absolute refractory interval  $t_0$  and inverse of the time constant  $\alpha$  parameters for this recovery function can be varied in order to find optimal combinations through perceptual experiments.

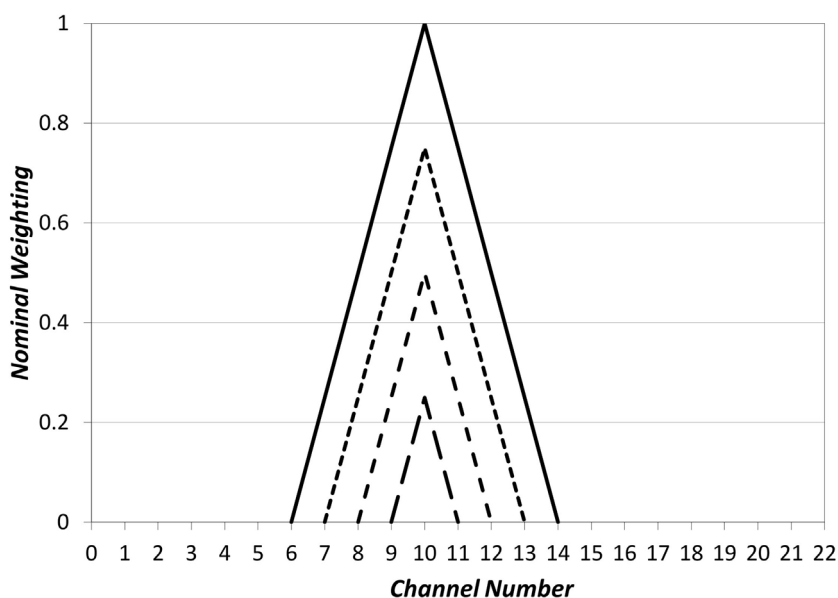
Immediately after a stimulus is presented on a given channel, the corresponding excitability is reduced proportionally by the stimulus intensity presented in that particular time interval. For instance, a stimulus corresponding to an input signal intensity of  $x$  (where  $0 \leq x \leq 1$ ) would cause the respective excitability to be reduced by  $x$ . Depending on the initial excitability state of the neural population associated with that particular channel, the excitability state after the stimulus is reduced by  $x$  and this could still result in a remnant excitability value greater than zero. [Figure 1\(b\)](#) shows how, after a stimulus of intensity  $x = 0.6$ , the available excitability is initially reduced to 0.4, and how this remnant excitability recovers over time. This remnant excitability is also taken into account in the excitability computations for subsequent time intervals. Should the same channel be re-selected for stimulation, the channel’s excitability will once again be reduced accordingly.

Recall that the portion of the excitability that has been reduced by any previous stimulation on this

channel will also be recovering and its contribution to the overall excitability must also be accounted for. At any given time, the channel's total excitability is therefore taken as the sum of the remnant excitability at that moment and the recovered excitability at that moment from previous stimulation. The persistent nature of the excitability variable allows for the effects of stimulation on the excitability to be tracked over time, and the different excitability components from each stimulus then summed together.

Closely associated with the excitability variable is the selection threshold consisting of  $thr$  and the input-complement  $(1 - A)$ . At the beginning of each time interval, the selection thresholds for each channel are computed based on the channel's corresponding input signal intensity  $A$ . The excitability values of channels that are below their respective selection thresholds are first set to zero, and the remaining non-zeroed excitability values then used to weight the corresponding input signal intensity. The channel with the largest excitability weighted input signal intensity is then selected as the next stimulus, and the process is repeated.

The selection of any stimulus to present on a given channel is made in competition with other channels. It is therefore important to also account for channel interaction effects. Whenever a stimulus is presented on a given channel, depending on the stimulus intensity, auditory neurons associated with adjacent neighbouring channels will also be stimulated due to the resultant electric field spread. In order to account for this, a model of the spread of excitation (SoE) function is used which spatially describes the excitation caused by a pulse on a channel. The SoE function is defined as a set of weights centred on the stimulated channel, with the central weight being the largest and set corresponding to the input signal intensity  $A$ . The SoE function is assumed to be symmetric, and its extent described as the number of channels  $n$  it spans on either side of the stimulated channel when the central weight is set to its maximum value of 1. Weights for channels at  $n$  and beyond are set to 0. For simplicity, the intermediate weights are linearly interpolated. For input signal intensities less than the maximum of 1, the central weight as well as the extent is reduced accordingly as shown in Figure 4.



**Figure 4.** Example of linearly interpolated spread of excitation (SoE) functions centred on channel 10 with various central weights and extents. Intermediate values are linearly interpolated. For an input signal of intensity 1, the central weight is 1 with an extent of  $n = 4$ , shown by the full line. The weighting on channels beyond  $n = 4$  are set to 0. As the input signal intensity is reduced, both the central weight and extent are reduced proportionately, as shown above by the dashed lines for input signal intensities of 0.75, 0.5 and 0.25 respectively.



### 3.2. Matlab Implementation

The Matlab model, consisting of a series of processing blocks, is derived from an implementation of the ACE coding strategy provided by Cochlear® Pty Ltd in its Nucleus™ Implant Communicator (NIC) [26] software library and the Nucleus Matlab Toolbox (NMT) [27], and modified accordingly to accommodate ECC (see Figure 5). Processing begins with a WAV file as the input signal. Frequency shaping of the input signal is then applied to simulate the pre-emphasis of the microphone output of a real Nucleus SP12 speech processor, before this is amplified and passed on to the Automated Gain Control (AGC) block whose task is to limit input signal intensities to some predefined level, such as 65 dB SPL. Input signals with intensities above this level are then simply presented at this level.

The signal is then processed by a 128-point FFT block followed by an aggregation block which combines the FFT output bins into maximally 22 channels for the Nucleus CI. Note that a logarithmic frequency to channel mapping is used to account for the cochlea's tonotopicity [28]. It is this FFT output array of 22 combined channel values that will be weighted by the corresponding excitability array of 22 values in the subsequent ECC block.

The ECC block essentially performs the selection of the channel to be stimulated which is repeated at regular time intervals corresponding to the specified overall stimulation rate. The block keeps track of the excitability state of each channel over time. During each stimulation interval, the excitabilities are computed and the channel with the highest excitability exceeding the respective selection threshold is then selected for stimulation. The excitability state variables values are persistent, being saved at the end of each interval and then made available again in the following interval. As the time interval till the next pulse is known, the excitability state of each channel at the next time interval can be computed for each time interval based on the excitability model. Note that by storing the excitability of each channel in a persistent variable, the additional computation time needed to retrieve, update and store the excitability state is minimal.

The selected stimulus information from the ECC block is then mapped and used to specify the corresponding stimulus pulse parameters, namely the active and reference electrodes, pulse amplitude, phase width, phase gap and duration. This mapping accounts for individual differences between actual CI listeners, such as the number of active electrodes or the individual sensitivity of these electrodes to the biphasic stimulation pulses. The output of the Matlab model is thus a sequence of CI stimulus pulses that can be examined for analysis.

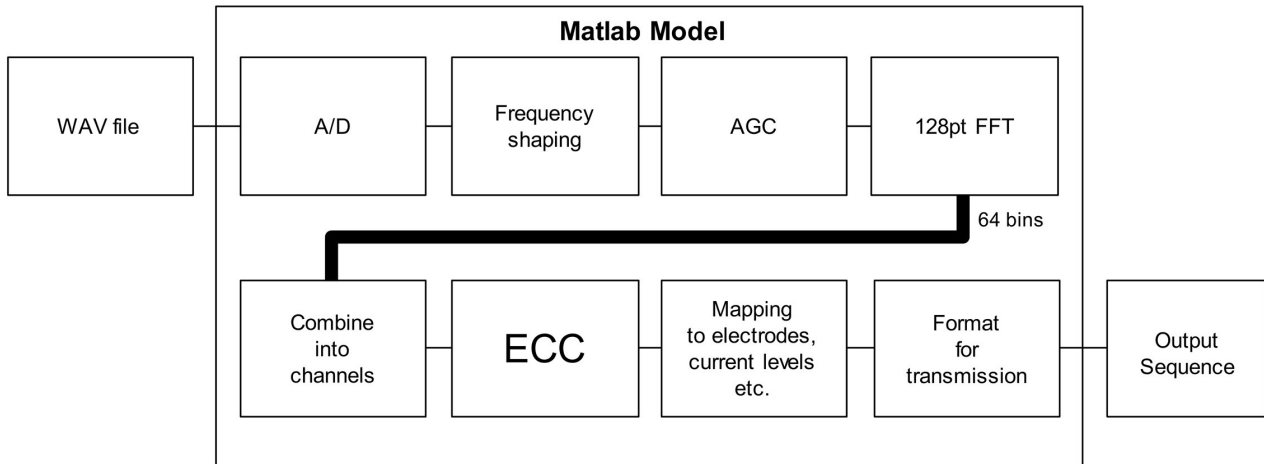
## 4. MATLAB MODEL OUTPUTS

The Matlab model is verified using various artificial input signals as well as realistic speech tokens, whereby the output from the model is examined and analysed. The analysis includes examining how the different variables involved in the decision making process, namely the excitability state, *thr* and the input-complement, change from interval to interval. In particular, their deterministic behaviour should be observable using the artificial input signals.

### 4.1. Testing with Artificial Input Signals

#### 4.1.1. Single Channel Input

An artificial single channel stimulus of finite duration was input directly into the ECC block in order to bypass the preceding blocks. The corresponding changes in the key variables at each stimulation interval were then examined in detail. Figure 6 shows how the excitability changes over time with respect to *thr* and the input-complement, with a constant amplitude ( $A = 1$ ) single-channel input signal, and with  $\delta = 0.25$ . The x-axis depicts individual stimulation intervals when the Matlab model decides whether a stimulus should or should not be presented, depending on whether the corresponding excitability value exceeds the selection threshold. The y-axis shows the excitability of the neurons corresponding to the stimulation channel being activated here.



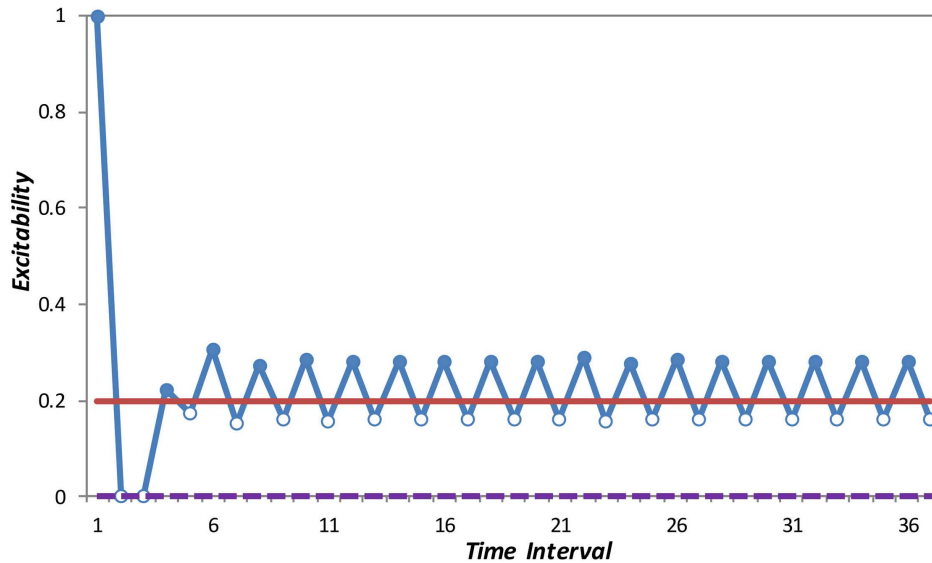
**Figure 5.** Schematic representation of the processing blocks in the Matlab Model. The crucial processing block is the ECC block where the excitability-based stimulus selection takes place.

With  $A = 1$ , this yields a  $thr$  (after Equation (1)) that remains constant at 0.2 throughout, while the input-complement is  $(1 - A) = 0$ . Since the selection threshold is effectively the larger one of the two values, the input-complement can be disregarded in this example and the effective selection threshold is therefore 0.2.

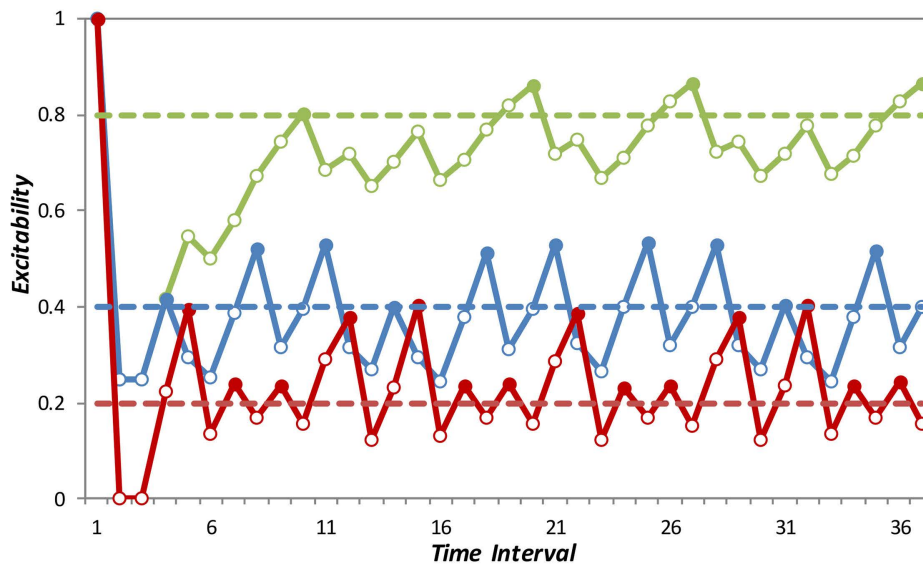
In the first interval, the initial excitability of 1 is obviously above the selection threshold 0.2, and yields a stimulus, indicated by a filled circle in **Figure 6**. After the first stimulus has been selected, the excitability is reduced to zero and remains so for the next two intervals, corresponding to the absolute refractory interval. In these two intervals, no stimuli are presented at all, indicated by empty circles in **Figure 6**. In the following interval, the excitability begins to recover. At this point, it is important to differentiate between two neural subpopulations: namely the portion of neurons that were not activated by the last stimulus (yielding a remnant excitability), and the portion of neurons that had been activated previously but have in the meantime recovered, and continue to recover. Together, the remnant plus the recovered excitability comprise the current excitability state at any given moment in time. In the fourth interval, the excitability state has recovered to 0.22, exceeding the selection threshold of 0.2, and a stimulus is therefore presented in this interval. The excitability is again reduced proportionally to the stimulus intensity for the following interval, while the previously stimulated subpopulation continues to recover. In subsequent intervals, the excitability state continues to be monitored and checked against the selection threshold, eventually settling down and resulting in a regular stimulation pattern, in this example on every second interval.

#### 4.1.2. Complex Input

A more complex, but more realistic scenario would be with multiple competing channels with different input signal intensities. **Figure 7** shows the excitability traces on each of three immediately adjacent channels, each with an input signal intensity of 0.2, 1.0 and 0.6 respectively. The corresponding  $thr$  values are 0.556, 0.2 and 0.294, while the input-complement values are 0.8, 0.0 and 0.4. Consequently, the corresponding selection threshold, namely the larger of the  $thr$  and input complement, of each of the three channels are 0.8, 0.2 and 0.4 respectively. For simplicity, only the selection thresholds are plotted as broken lines in **Figure 7**. In any single interval, only channels whose excitabilities are greater than their corresponding selection thresholds are considered for selection. From these candidates, the channel with the highest excitability-weighted input signal intensity is then selected for stimulation, and is indicated with a filled circle. Channel interactions arising from the spread of excitation are also accounted for, for instance, in the first interval, the stimulus on the middle channel also results in the excitabilities of both neighbouring channels to be reduced accordingly.



**Figure 6.** The excitability (blue trace) yields a stimulus only when it exceeds both the thr (dark red trace) and the input-complement (purple trace). Note that the larger thr is indicated by a full line while the smaller input-complement is indicated by a broken line. Intervals resulting in a stimulus being presented are indicated by filled circles, while empty circles indicate no stimulation.



**Figure 7.** When multiple channels compete with each other, only channels with an excitability greater than their corresponding selection threshold are firstly considered. When more than one such channels exist, the channel with the highest excitability-weighted input signal intensity is selected (filled circles) for stimulation, while the others remain unselected (empty circles), even though their excitability may still be above their selection threshold. The selection thresholds of each component are shown here as respective dashed lines. In this example, the channel with the highest input signal intensity of 1.0 (dark red trace) has the highest reselection rate, while the second highest intensity (0.6) channel (blue trace) has the next highest reselection rate and the third channel (input signal intensity 0.2, green trace) has the lowest reselection rate.

Altogether, it can be seen that the channel with the highest input signal intensity of 1.0, with 15 stimuli altogether, has the largest number of stimuli over the entire input signal duration. The channel with the next highest input signal intensity of 0.6 has 10 stimuli, followed by the last channel with input signal intensity of 0.2 having only 4 stimuli over the same input signal duration.

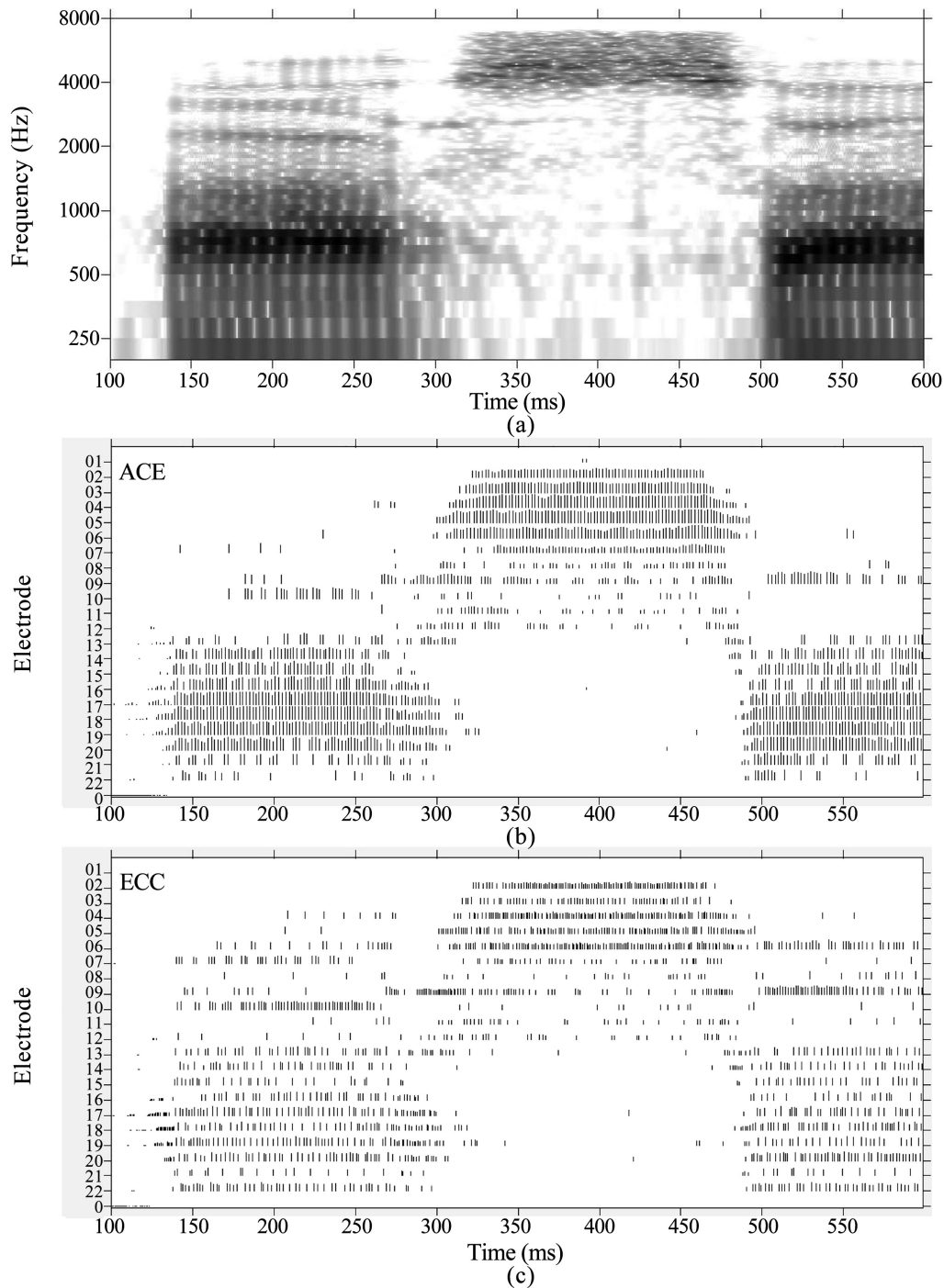
## 4.2. Testing with Realistic Speech Tokens

For even more complex stimuli such as speech tokens, the interval-by-interval behaviour can also be examined visually but plotting the corresponding excitability and selection threshold of more than three channels simultaneously in a single figure is not practical. Alternatively, the output sequence from the Matlab model can be plotted in the form of an electrodogram [29, 30], which display how the stimulus pulses presented on individual channels or electrodes vary as a function of time. The electrodogram resembles a spectrogram but with the frequency axis replaced by discrete electrodes, ordered from low (apical electrodes) to high (basal electrodes) frequencies. Note that for Nucleus implants, the electrode numbering is in reverse order to the frequency: e22 has the lowest frequency and e01 the highest. The x-axis, which depicts time, indicates the time of occurrence of individual stimulation pulses in the output sequence. Furthermore, instead of the intensity being coded by colour shades or a gray scale, the intensity of individual pulses in the output sequence is displayed as the height of a corresponding bar.

**Figure 8** shows the spectrogram of the first 500 msec of the speech token “asa” followed by the corresponding electrodograms for the ACE and ECC coding strategies. Note that the frequency-axis (y-axis) of the spectrogram is logarithmically scaled, as are the y-axes of the electrodograms, which represent channels, but are plotted in terms of the corresponding physical electrodes, with e22 being the lowest frequency channel and e01 the highest. Typically, the input frequency range of 180 - 7938 Hz is divided logarithmically into  $m$  channels corresponding to  $m$  physical electrodes [31], although the first few low frequency channels are linearly distributed due to discretization effects from the FFT analysis.

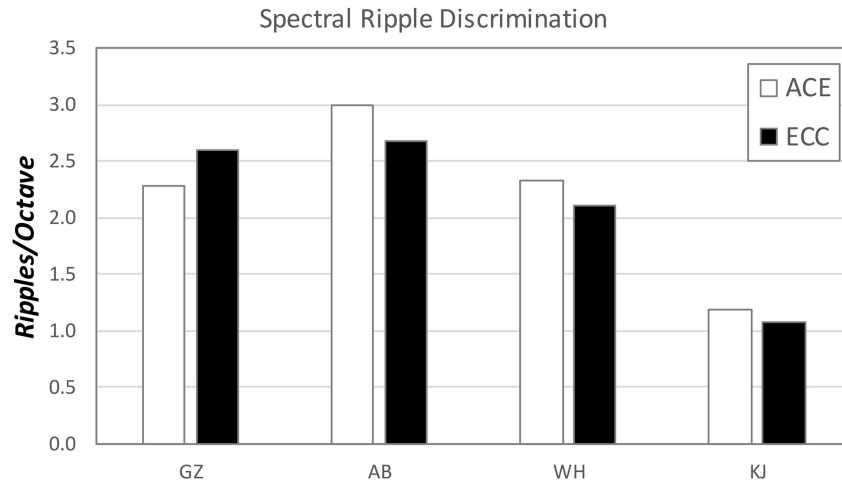
Timing differences can be seen between the two output sequences. ACE always selects a subset  $n$  of the highest energy channels (maxima) from the total of  $m$  channels at a time, presenting the  $n$  selected stimuli on their corresponding channels sequentially and equally spaced in time over the duration of a so-called stimulation frame. The stimulation frame is in turned defined as  $1/R$ , where  $R$  is the corresponding channel stimulation rate in pulses per second (pps). For example, ACE with a channel stimulation rate  $R = 500$  pps and  $n = 8$  will nominally present 8 stimuli on different channels at  $1/(8 \times 500) = 250$  us intervals within the stimulation frame of duration  $1/500 = 2000$  us. This time interval between individual stimuli is also known as the overall stimulation rate, which is derived as  $n$  times the channel stimulation rate. In the example here, the overall stimulation rate  $n \times R = 8 \times 500 = 4000$  pps. As a result of this stimulus selection approach, the stimulation rate on a given channel is nominally equal to the specified channel stimulation rate  $R$ , producing the regular and similar timing structure observed in the stimulated channels shown in **Figure 9**.

ECC, by contrast, does not employ a stimulation frame with multiple stimuli per frame. Instead, it repeats its selection stimulus by stimulus, in other words, at the overall stimulation rate. Compared to an ACE channel stimulation rate of 500 pps with  $n = 8$ , ECC would select its stimuli at an equivalent rate of  $8 \times 500 = 4000$  pps. Unlike the ACE output, the ECC output has more variation in the stimulation timing pattern observed on each channel, which arises from the competing nature of the ECC stimulus selection procedure. Of particular interest is the visibly increased density of pulses in the fricative “s” portion compared to the vowel portions of the signal. In the fricative portion with relatively fewer frequency components and hence fewer channels to pick from, ACE does not always find  $n$  channels to stimulate within each stimulation frame, leaving some stimulus intervals empty. ECC, in comparison, is more likely to find a channel with an excitability exceeding its selection threshold in each interval. As a result, ECC is more frequently stimulated, and the larger number of ECC stimuli are shared out between the small number of channels, effectively increasing their frequency of stimulation. In the vowel portions with their larger number of frequency components, the resulting output is now shared out amongst a larger number of



**Figure 8.** Spectrogram and electrodograms of the first 500 msec of the speech token “asa” with ACE versus ECC. Note that the spectrogram’s y-axis (200 - 8000 Hz) is logarithmically scaled. The ACE output (total 3026 stimuli) has the same regular temporal structure, seen here as a uniform grating pattern, on all channels, corresponding to the channel stimulation rate, due to its  $n$  of  $m$  stimulus selection. In contrast, the ECC output’s (total 3188 stimuli) temporal structure shows greater variability (the visual density of the grating patterns varies) within a channel (e.g. e02 rates range from 190 - 1333 pps), as well as between channels (e.g. e10 ranges from 15 - 1000 pps) over time, due to its interval-by-interval stimulus selection.





**Figure 9.** Results for spectral ripple discrimination test with 4 CI listeners. The y-axis is in units of ripples per octave. Higher scores are better. ACE (white) is marginally better than ECC (black) in 3 of 4 cases.

channels, resulting in each channel being stimulated less often. It is unclear if such an effect will be perceptually desirable or not, and this may have to be modified in a future iteration of ECC.

Another important aspect of ECC that is illustrated in the example above is that the stimulation levels used for the output pulses are different from those of ACE. The ACE pulse stimulation level is derived from the corresponding input signal intensity  $A$  on each channel. With ECC, one has the possibility to use a stimulation level that is related to the excitability state. In the ECC example shown in Figure 8, the stimulation level is derived from the excitability weighted input signal intensity, with the reasoning that a given channel's capacity to react to a stimulus depends on its excitability state. For the duration of any persistent input signal, the overall excitability is generally reduced from the resting state of 100% excitability. As a result, stimulation levels obtained via excitability-weighting will also be lower than those from ACE.

### 4.3. Modifying ECC Parameters

The ECC strategy involves several parameters that affect the excitability computations.

Firstly, the selection threshold, which is determined jointly by  $thr$  and the input complement  $(1 - A)$ , basically regulates the likelihood of selection as a function of the input signal intensity  $A$ . Higher intensity input signals are more likely to be selected more often and vice versa. The  $thr$  function is determined by the variable  $\delta$  according to Equation (1) described earlier. However, as can be seen in Figure 3, the selection threshold is dominated by the  $(1 - A)$  threshold, and changing  $\delta$  (and in turn  $thr$ ) has little effect especially when  $\delta < 0.5$ . This was confirmed by examining electrodiagram outputs with different values of  $\delta$ .

Secondly, the recovery function itself determines how quickly the excitability of a stimulated neural population corresponding to a particular channel recovers to allow the channel to be eligible again for selection and stimulation. Faster recovery means that a given channel is more often considered for selection, leading to a higher stimulation rate on that channel. This will in turn favour channels with higher input signal intensities. This was also confirmed by varying the recovery time constant and examining corresponding electrodiagram outputs.

Thirdly, the overall stimulation rate, which determines the stimulation time intervals, also directly affects how often the excitability state is updated. Slower update rates allow previously stimulated neural populations to recover to higher levels, while faster update (overall stimulation) rates gives these neural populations less chance to have recovered as much. Changing the overall stimulation rate will, in addition

to affecting the number of stimuli generated per unit time, also change the mixture of channels competing for selection at any given time will also be affected by the overall stimulation rate selected, thereby yielding a different distribution of activity across the electrode array.

Lastly, the SoE function determines how a particular stimulus affects adjacent or neighbouring channels, with the effect of reducing their excitability and also the likelihood of their subsequent selection. The general effect is to allow channels with lower input signal intensities to also be presented, that otherwise, for instance, with simple maxima selection strategies like ACE, would be ignored. Broadening the SoE function should therefore achieve a greater representation of the entire input signal across the electrode array, generally spreading out the activity to more channels across the array. Note that such an effect is also observed with the MP3000™ coding strategy [21], whose forward masking function resembles the SoE function. With ECC, this spreading also tends to reduce the stimulation rate on individual channels. Conversely, narrowing the SoE tends to concentrate the stimuli on fewer channels, resulting in a net increase in the stimulation rate on the activated channels. A further effect of spreading the effects of each stimulus across the array like this is that the original input signal intensity tends to be evened out across the array with broader SoE functions, while the original input signal intensities are better represented in the output pulse amplitudes with narrower SoE functions.

Obviously, some degree of interplay between the various ECC parameters described above can be expected and the perceptual effects of changing these parameters either individually or in conjunction with one another will need to be assessed in subjective tests with cochlear implant users.

## 5. REAL-TIME IMPLEMENTATION AND TESTING

The output from the Matlab model can also be presented to a Nucleus implant for assessment by a CI-listener via streaming. However, the stimuli to be presented would need to be processed in advance so they can be streamed when needed. This pre-processing can be time consuming, requiring additional planning depending on the number and types of stimuli to be assessed. Consequently, both ACE and ECC Matlab models were implemented as Simulink xPC Target real-time models, in conjunction with a Speedgoat™ real-time hardware system. This allows more flexibility in the range of sounds that can be presented, including running input (speech or otherwise), in order to allow the listener to be familiarized with the sound impressions. The input signal from either a microphone or direct connection to a sound card's output is processed in the same manner as in a CI speech processor, and with the appropriate custom hardware, the output is then encoded for transmission to a CI. The Speedgoat™ real-time target system therefore essentially functions as the CI-listener's speech processor. The real-time system was then used to present signals encoded either using ACE or ECC in a pilot trial involving 4 experienced (average 12 years of CI use) adult CI-listeners (average age 54). Approval of the Ethics Committee of the University of Zurich was obtained (KEK-ZH 2014-0202). All participants gave written informed consent after a comprehensive explanation of the procedures.

For these pilot tests, the ACE model used a speech processor map with 8 maxima presented at a channel rate of 500 pps. These ACE maps for each CI-listener were all prepared separately using routine clinical Nucleus Custom Sound fitting software.

The ECC model used an equivalent overall stimulation rate of 4000 pps, and otherwise used the same stimulation parameters in the ACE map. The relevant ECC parameters were set to  $\delta = 0.25$ , with absolute and relative refractory intervals of 300  $\mu$ s and 1000  $\mu$ s respectively for the recovery function parameters, and an SOE function extent of 4 electrodes wide. For the test here, the stimulation level used for ECC was derived from the excitability-weighted input signal amplitude. The overall loudness with ECC was reported by all 4 CI-listeners as being slightly softer than the ACE counterpart, but the loudness was still judged as being adequate for performing the tests. The stimulation level was not increased to compensate for the loudness difference in order not to also affect the reduction in channel interaction expected with using reduced stimulation levels with ECC.

The following two tests were carried out:

## 5.1. Spectral Ripple Discrimination Test

The spectral ripple discrimination test [32] was intended to examine if and how the spectral resolution differs between the two coding strategies. As in [33], ripple amplitudes that are sinusoidal on a logarithmic scale were used. The test signals were calibrated to match a free-field loudness of 65 dB SPL with ACE on a clinical speech processor. The results summarized in Figure 10 suggest that the spectral resolution was marginally better with ACE than ECC for 3 out of 4 CI-listeners.

## 5.2. OLSA Adaptive Sentences in Noise Test

The OLSA adaptive sentences in noise test [34] were intended to examine how ECC fares against ACE with complex listening situations. The test signals were calibrated such that 0 dB SNR corresponded to a free-field loudness of 65 dB SPL for both the test and noise signals with ACE on a clinical speech processor. The results summarized in Figure 10 suggest that ECC yielded better speech reception thresholds (SRT) than ACE for 3 out of 4 subjects. The SRT improvements ranged from 0.4 to 1.3 dB.

## 6. DISCUSSION

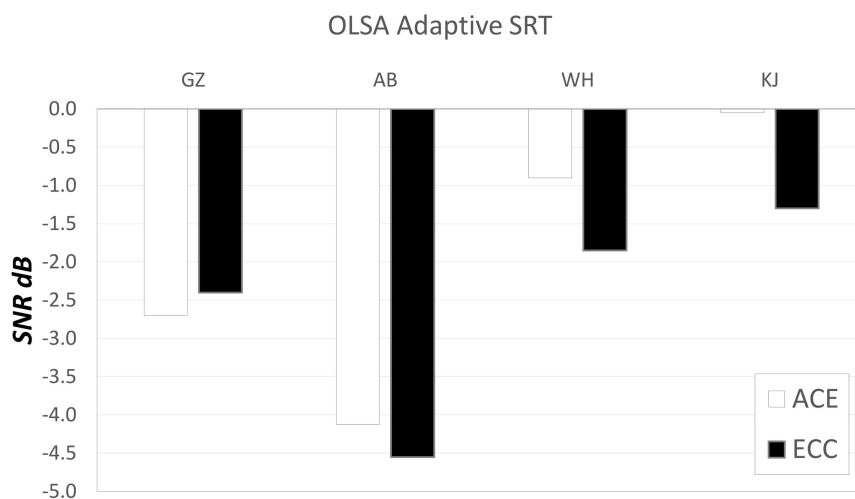
### 6.1. Matlab Model Outputs

The interval-by-interval analysis of the Matlab model outputs with simplified artificial inputs demonstrate that the ECC coding strategy's stimuli selection based on the weighted excitability threshold can be deterministically verified, and behaves as expected. This was also the case with an artificial input signal on three channels. The ACE and ECC outputs with a speech token "asa" input illustrate the different distribution of stimuli across the array as well as in time. Although the same interval-by-interval analysis was not conducted with these outputs, it could be seen from the corresponding electrodiagrams in Figure 8 that the outputs demonstrated characteristics which are consistent with the two coding strategies.

The observed differences can be expected to translate into various perceptual effects.

### 6.2. Increased Spectral Representation

One of the expected effects of using the excitability to regulate the stimulus selection is a greater efficiency in presenting the input signal to the neural interface since redundant stimuli are not produced



**Figure 10.** OLSA adaptive sentences in noise test results. The SRT in dB units is shown on the y-axis, and lower SRTs are better. In 3 of 4 cases, the CI listeners performed better with ECC (black) than ACE (white), with improvements ranging from 0.4 to 1.3 dB.

when the target neural population is not excitable. Improved efficiency is important for systems with capacity limitations as it can increase the amount of information transmitted for a given cost. Cochlear implants are subject to such limitations in that presenting the entire input frequency spectrum would result in slowing down the refresh rate. The ACE coding strategy attempts to avoid this by limiting the spectral information to only the largest  $n$  maxima. Note that in this way, ACE may be regarded as behaving effectively like a spectral sharpener, picking only the largest spectral components and setting the unselected components to zero. For input signals with many frequency components close to each other such as vowels, the ACE output also tends to be clustered together, with many redundant stimuli of the adjacent channels within the cluster. ECC, on the other hand, avoids such clustering by considering the excitability of the activated channels and allowing the activity to spread to other more excitable sites. Compared to an  $n$  of  $m$  coding strategy such as ACE, ECC is more likely to spread out its activity across more channels and thereby, present a greater amount of the input spectral information. Note that the spread of activity arises primarily due to ECC accounting for SoE effects. The MP3000™ coding strategy [21], which mimics masking effects in a similar manner, also spreads the activity similarly. ECC differs in that the reselection of a particular channel depends on its excitability and can occur at any time interval, whereas MP3000™ reselects its stimuli strictly at the stimulation frame rate. ACE with a larger  $n$  would also attain greater input signal representation, but at a higher energy cost. When  $n = m$ , this is equivalent to CIS and has the highest energy cost. ECC, like MP3000™, does not need to present as many stimuli as CIS while having comparable representation of the input signal, resulting in corresponding power savings over CIS.

### 6.3. Reduced Channel Interaction

One of the concerns with CI stimulation is the channel interaction that arises from the accompanying electric field spread [35]. Channel interaction is widely regarded to be a major limitation in the search for more refined coding strategies. With most present day coding strategies, the stimulation intensity is derived directly from the input signal intensity of the corresponding stimulation channel. It is known that changing the stimulation rate also affects the perceived loudness. ECC provides a mechanism to introduce a rate loudness cue by regulating the stimulation rate on a given channel depending on the corresponding input signal intensity. As a result, the loudness cue which is normally regulated by the stimulation intensity could potentially be augmented by rate loudness cues, and the stimulation intensity can be reduced accordingly.

The exact amount of reduction that is required is presently unknown and would need to be determined experimentally. In the ECC implementation described in this paper, a simple initial estimate is used, whereby the stimulation level is derived from the excitability-weighted input signal intensity, following the assumption that a given channel's capacity to react to a stimulus is dependent on its excitability state. If its excitability has been reduced due to prior stimulation, the intensity of the next stimulus on this channel can be reduced accordingly, thereby avoiding unnecessarily excessive stimulation and resulting in reduced electric field spread and channel interaction. The reduction in the stimulation level can be expected to result in a softer loudness percept with ECC compared ACE, especially if the expected rate loudness cues do not contribute adequately to the perceived loudness.

Compared to ACE, ECC will spread out its stimuli over a larger portion of the electrode array. While the increased spread of activity may potentially reduce the saliency of specific signal components, when combined with reduced channel interaction, this could conceivably still lead to the individual channels and their corresponding signal components being better perceived compared to the more clustered and interacting activity that results from ACE.

Note that there are other ways to further influence the electric field spread such as by using bipolar, tripolar or even phased array (e.g. [36, 37]) as opposed to monopolar stimulation modes. However, these alternative stimulation modes are often also associated with higher stimulation levels, and the trade-off in the electric field spread resulting from the electrode configuration compared to the stimulation level needs to be studied more thoroughly first before a conclusive decision can be made on this matter. Also, the

spread of activity could also potentially be reduced by using narrower filters in the analysis stage by reducing the amount of filter overlap. This is because broader filters are more likely to duplicate frequency components in multiple adjacent channels. However, the electric field spread arising from the stimuli themselves does not diminish merely by having narrower filters. Although these features are not specific to ECC, they could be used in combination with ECC to possibly achieve more prominent results.

#### 6.4. Some Expected Outcomes

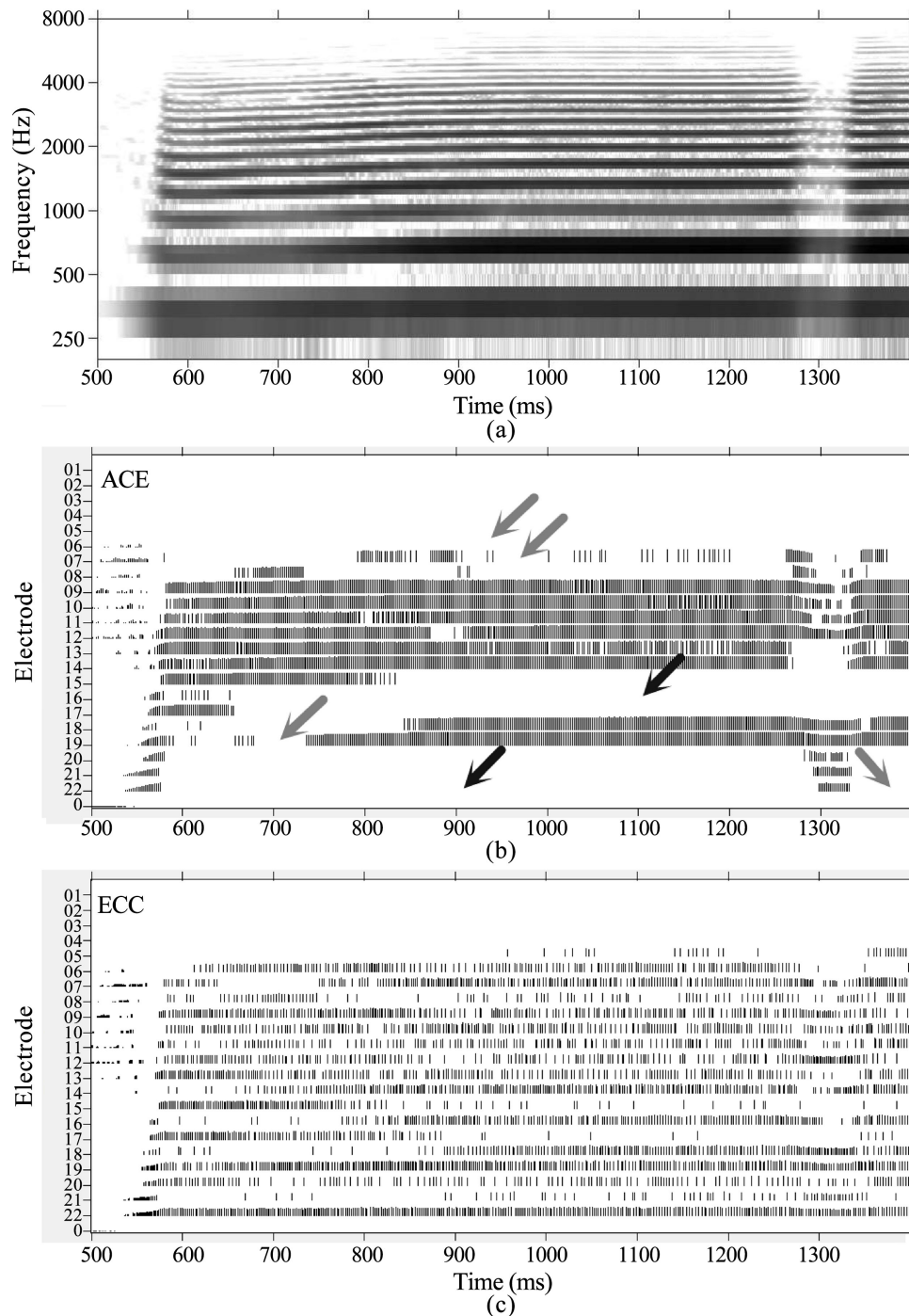
The ACE coding strategy, which extracts the dominant frequency components of an input signal, is obviously robust for signals with simple spectral structures such as vowels or even consonants. The amount of spectral information presented can also be increased by raising the number of maxima. However, merely increasing the amount of information presented will not necessarily make them more perceptible, especially when channel interaction effects arising from the accompanying electric field spread will limit the perceptibility of the additional information. Also, the tendency for ACE to concentrate on components with larger amplitudes will also miss weaker but possibly still important components. ECC, on the other hand, could fare better in making this information perceptually more salient due to reduced channel interaction achieved firstly by spreading and not clustering the resultant stimuli, and secondly through reducing the stimulation levels used. Also, ECC is more likely to select less dominant frequency components compared to ACE. ECC could therefore be a better choice for presenting signals with more complex spectral structures such as music, where a greater saliency in the perceived information conveying differences in melody and timbre is desirable. Tasks such as musical instrument identification, where the timbre information is highly encoded within the harmonic structure could possibly benefit from ECC. Even simple melodic tone discrimination may be better if more information about the harmonic contents is present in otherwise very similar stimulation patterns. Potentially, reduced channel interaction would also be helpful to better resolve harmonic components.

**Figure 11** shows how the electrodiagrams for a short extract from a saxophone piece differ between ACE with 8 maxima, and ECC. One striking difference between the ACE and ECC output is how the input signal is much better represented throughout with ECC, whereas the ACE outputs are generally missing the slightly weaker first and third harmonics due to the maxima selection approach. These missing harmonics will in turn change the timbre of the perceived sound. Additionally, the higher stimulation intensities produced by ACE which could smudge out the signal components due to the associated channel interaction. ECC, on the other hand, with lower stimulation levels, may help to keep signal components more salient.

#### 6.5. Real-Time Implementation: Perceptual Test Results

The spectral ripple discrimination test results indicate that the expected improvement in spectral resolution with ECC failed to materialize. One possible explanation for this is that ACE, in selecting a limited number of maxima from the input signal spectrum, effectively acts as a spectral sharpener. By picking only the strongest channels, it is more likely to be able to more effectively represent the peaks in the spectral ripples. In particular, this also produces gaps in the input spectrum where spectral components are left out. ECC, on the other hand, with its tendency to spread out its activity across the electrode array, is more likely to smooth out the gaps in the input spectrum. That ECC is able to even match ACE at all could perhaps be due to additional perceptual cues such as rate loudness due to the input signal amplitudes being encoded within the corresponding channel stimulation rates. It is possible that this effect was simply too weak compared to the loudness contribution from the stimulation levels. It should also be noted that at this stage in the development, the ECC parameters are unlikely to be optimized. Alternatively, the reported softer overall loudness due to using lower stimulation levels derived from the excitability-weighted input signal amplitudes may also have weakened this effect. This will have to be investigated further, using for instance, the original unweighted input signal amplitudes. Note that the spectral resolutions greater than 2 ripples/octaves obtained for 3 of the CI listeners here are rather high compared to the average resolution





**Figure 11.** (a) Spectrogram of a short extract from a saxophone piece with corresponding electrodograms for (b) ACE 8 maxima and (c) ECC respectively. Note that the spectrogram's y-axis (200 - 8000 Hz) is logarithmically scaled. The 1<sup>st</sup> and 3<sup>rd</sup> harmonics around e22 and e16 are missing with ACE, as indicated by the two dark arrows. Some of the other missing harmonics are also indicated by the lighter arrows. Resolving the higher harmonics could possibly also be hindered by smudging due to the clustered higher stimulus intensities used in ACE. ECC, in contrast, is able to represent even more of the input signal spectrum, and the slightly lower stimulation levels may help to reduce this smudging.

generally reported in the literature (e.g. [33] [38, 39]). As explained in [33], this may in part be due to the use of ripple amplitudes that were sinusoidal on a logarithmic scale rather than a linear scale as used in [32]. Another possible explanation could be the presence of ripple-edge cues as argued by [40]. It would be interesting to consider if their modified spectral-ripple test would show better performance with ECC over ACE given the spectral sharpening inherent in ACE.

The results from the OLSA adaptive sentences in noise test are interesting in that ECC appears to be able to yield better performance with speech in noise than ACE. Here, a possible explanation is that the greater representation of the input signal by ECC resulted in more of both the target signal as well as the noise being presented. The increased representation of the target signal then allowed the listener to better extract it from the accompanying noise. By comparison, ACE selects only a limited number of maxima from the noisy input signal, which may either contain the test signal or noise. This would generally result in a reduced representation of not only the noise but also the target signal. The corresponding reduction in the amount of target signal presented would then in turn lead to greater difficulties in separating it from the accompanying noise. It is unclear from the results here whether the reported softer overall loudness with ECC as tested could have affected the results as well. This is not expected to be the case, since both target signal and accompanying noise are equally softer with ECC. Nevertheless, as with the spectral ripple discrimination test above, the effect of using the original unweighted input signal amplitudes for the stimulation levels should also to be investigated further. It is also not clear from the results here whether rate loudness cues have contributed to these results or not.

The last discussion point here suggests that there are potential merits in a coding strategy which presents as much of the input signal spectrum as possible such as ECC, compared to one with more limited representation such as ACE. This would be particularly more so with complex input signals such as speech in noise.

Due to the pilot nature of these preliminary assessments, ECC parameters such as  $\delta$ , the recovery function timing and the SOE function extent used in the pilot tests reported here have not been optimized. It is possible that optimized ECC parameters would have yielded different or possibly more pronounced results. However, these test results provide an insight into the general perceptual differences that can be expected between ECC and ACE.

## 7. CONCLUSIONS

A novel Excitability Controlled Coding (ECC) strategy based on the neural excitability of stimulated auditory neurons, especially their refractory behaviour, is presented here. By also taking into account the electric field spread, a more efficient representation of the input signal activity can be expected. ECC also encodes the input signal intensity into the corresponding stimulation rate of a particular frequency channel, potentially augmenting the intensity information already present in the stimulation pulse intensity. Pilot test results from 4 CI listeners suggest that ECC may be advantageous with complex input signals such as speech in noise.

## DECLARATIONS

### Acknowledgements

This work was supported by a research grant from Cochlear AG, Basel, Switzerland.

### Ethics Approval

Approval of the Ethics Committee of the University of Zurich was obtained. All participants gave written informed consent after a comprehensive explanation of the procedures.

### Conflict of Interests

Author Matthijs Killian is employed by Cochlear Technology Centre, Mechelen, Belgium. The re-

maintaining authors declare that there are no conflicts of interests.

## REFERENCES

1. Tong, Y.C., Black, R.C., Clark, G.M., Forster, I.C., Millar, J.B., O'Loughlin, B.J. and Patrick, J.F. (1979) A Preliminary Report on a Multiple-Channel Cochlear Implant Operation. *The Journal of Laryngology & Otology*, **93**, 679-695. <https://doi.org/10.1017/S0022215100087545>
2. Spillmann, T., Dillier, N. and Guntensperger, J. (1982) Electrical Stimulation of Hearing by Implanted Cochlear Electrodes in Humans. *Applied Neurophysiology*, **45**, 32-37. <https://doi.org/10.1159/000101574>
3. Shannon, R.V. (1983) Multichannel Electrical Stimulation of the Auditory Nerve in Man. II. Channel Interaction. *Hearing Research*, **12**, 1-16. [https://doi.org/10.1016/0378-5955\(83\)90115-6](https://doi.org/10.1016/0378-5955(83)90115-6)
4. Cohen, L.T., Saunders, E. and Richardson, L.M. (2004) Spatial Spread of Neural Excitation: Comparison of Compound Action Potential and Forward-Masking Data in Cochlear Implant Recipients. *International Journal of Audiology*, **43**, 346-355. <https://doi.org/10.1080/14992020400050044>
5. Wilson, B.S., Finley, C.C., Farmer, J.C., Lawson, D.T., Weber, B.A., Wolford, R.D., Kenan, P.D., White, M.W., Merzenich, M.M. and Schindler, R.A. (1988) Comparative Studies of Speech Processing Strategies for Cochlear Implants. *Laryngoscope*, **98**, 1069-1077.
6. von Wallenberg, E.L., Hochmair, E.S. and Hochmairdesoyer, I.J. (1990) Initial Results with Simultaneous Analog and Pulsatile Stimulation of the Cochlea. *Acta Oto-Laryngologica*, **469**, 140-149.
7. Wilson, B.S., Finley, C.C., Lawson, D.T., Wolford, R.D., Eddington, D.K. and Rabinowitz, W.M. (1991) Better Speech Recognition with Cochlear Implants. *Nature*, **352**, 236-238. <https://doi.org/10.1038/352236a0>
8. Wilson, B.S., Finley, C.C., Lawson, D.T., Wolford, R.D. and Zerbi, M. (1993) Design and Evaluation of a Continuous Interleaved Sampling (Cis) Processing Strategy for Multichannel Cochlear Implants. *Journal of Rehabilitation Research and Development*, **30**, 110-116.
9. Skinner, M.W., Holden, L.K., Whitford, L.A., Plant, K.L., Psarros, C. and Holden, T.A. (2002) Speech Recognition with the Nucleus 24 SPEAK, ACE, and CIS Speech Coding Strategies in Newly Implanted Adults. *Ear and Hearing*, **23**, 207-223. <https://doi.org/10.1097/00003446-200206000-00005>
10. Vandali, A.E., Sucher, C., Tsang, D.J., McKay, C.M., Chew, J.W.D. and McDermott, H.J. (2005) Pitch Ranking Ability of Cochlear Implant Recipients: A Comparison of Sound-Processing Strategies. *Journal of the Acoustical Society of America*, **117**, 3126-3138. <https://doi.org/10.1121/1.1874632>
11. Laneau, J., Wouters, J. and Moonen, M. (2006) Improved Music Perception with Explicit Pitch Coding in Cochlear Implants. *Audiology and Neuro-Otology*, **11**, 38-52. <https://doi.org/10.1159/000088853>
12. Drennan, W.R. and Rubinstein, J.T. (2008) Music Perception in Cochlear Implant Users and Its Relationship with Psychophysical Capabilities. *Journal of Rehabilitation Research and Development*, **45**, 779-789. <https://doi.org/10.1682/JRRD.2007.08.0118>
13. Morton, K.D., Torriano, P.A., Throckmorton, C.S. and Collins, L.M. (2008) Mandarin Chinese Tone Identification in Cochlear Implants: Predictions from Acoustic Models. *Hearing Research*, **244**, 66-76. <https://doi.org/10.1016/j.heares.2008.07.008>
14. Vandali, A.E. and van Hoesel, R.J.M. (2011) Development of a Temporal Fundamental Frequency Coding Strategy for Cochlear Implants. *Journal of the Acoustical Society of America*, **129**, 4023-4036. <https://doi.org/10.1121/1.3573988>
15. Blamey, P.J., Dowell, R.C., Brown, A.M., Clark, G.M. and Seligman, P.M. (1987) Vowel and Consonant Recognition of Cochlear Implant Patients Using Formant-Estimating Speech Processors. *Journal of the Acoustical Society of America*, **82**, 48-57. <https://doi.org/10.1121/1.395436>

16. Patrick, J.F. and Clark, G.M. (1991) The Nucleus 22-Channel Cochlear Implant System. *Ear and Hearing*, **12**, S3-S9. <https://doi.org/10.1097/00003446-199108001-00002>
17. Riss, D., Arnoldner, C., Baumgartner, W.D., Kaider, A. and Hamzavi, J.S. (2008) A New Fine Structure Speech Coding Strategy: Speech Perception at a Reduced Number of Channels. *Otology & Neurotology*, **29**, 784-788. <https://doi.org/10.1097/MAO.0b013e31817fe00f>
18. Hu, H., Krasoulis, A., Lutman, M. and Bleeck, S. (2013) Development of a Real Time Sparse Non-Negative Matrix Factorization Module for Cochlear Implants by Using xPC Target. *Sensors (Basel)*, **13**, 13861-13878. <https://doi.org/10.3390/s131013861>
19. Zeng, F.G. and Shannon, R.V. (1995) Loudness of Simple and Complex Stimuli in Electric Hearing. *The Annals of Otology, Rhinology & Laryngology*, **166**, 235-238.
20. Theelen-van den Hoek, F.L., Boymans, M., van Dijk, B. and Dreschler, W.A. (2016) Adjustments of the Amplitude Mapping Function: Sensitivity of Cochlear Implant Users and Effects on Subjective Preference and Speech Recognition. *International Journal of Audiology*, **55**, 674-687. <https://doi.org/10.1080/14992027.2016.1202454>
21. Noguiera, W., Büchner, A., Lenarz, T. and Edler, B. (2005) A Psychoacoustic NofM-Type Speech Coding Strategy for Cochlear Implants. *EURASIP Journal on Advances in Signal Processing*, **18**, 3044-3059. <https://doi.org/10.1155/ASP.2005.3044>
22. Buechner, A., Beynon, A., Szyfter, W., Niemczyk, K., Hoppe, U., Hey, M., Brokx, J., Eyles, J., Van de Heyning, P., Paludetti, G., Zarowski, A., Quaranta, N., Wesarg, T., Festen, J., Olze, H., Dhooge, I., Muller-Deile, J., Ramos, A., Roman, S., Piron, J.P., Cuda, D., Burdo, S., Grolman, W., Vaillard, S.R., Huarte, A., Frachet, B., Morera, C., Garcia-Ibanez, L., Abels, D., Walger, M., Muller-Mazotta, J., Leone, C.A., Meyer, B., Dillier, N., Steffens, T., Gentine, A., Mazzoli, M., Rypkema, G., Killian, M. and Smoorenburg, G. (2011) Clinical Evaluation of Cochlear Implant Sound Coding Taking into Account Conjectural Masking Functions, MP3000. *Cochlear Implants International*, **12**, 194-204. <https://doi.org/10.1179/1754762811Y0000000009>
23. Laneau, J., Wouters, J. and Moonen, M. (2004) Relative Contributions of Temporal and Place Pitch Cues to Fundamental Frequency Discrimination in Cochlear Implantees. *The Journal of the Acoustical Society of America*, **116**, 3606-3619. <https://doi.org/10.1121/1.1823311>
24. Morsnowski, A., Charasse, B., Collet, L., Killian, M. and Muller-Deile, J. (2006) Measuring the Refractoriness of the Electrically Stimulated Auditory Nerve. *Audiology and Neuro-Otology*, **11**, 389-402. <https://doi.org/10.1159/000095966>
25. Killian, M.J.P. (2009) Sound Processing Method and System. WIPO.
26. Irwin, C. (2006) NIC v2 Software Interface Specification E11318RD (Technical Report). Cochlear Ltd., Lane Cove.
27. Swanson, B.A. and Mauch, H. (2006) Nucleus Matlab Toolbox 4.20 Software User Manual. Cochlear Ltd., Lane Cove.
28. Greenwood, D.D. (1990) A Cochlear Frequency-Position Function for Several Species—29 Years Later. *The Journal of the Acoustical Society of America*, **87**, 2592-605. <https://doi.org/10.1121/1.399052>
29. Dillier, N., Bogli, H. and Lai, W.K. (1995) Electrographic Analysis and Field Evaluation of the Speak Coding Strategy. *The Annals of Otology, Rhinology & Laryngology*, **166**, 354-356.
30. Grasmeder, M.L. and Lutman, M.E. (2006) The Identification of Musical Instruments through Nucleus Cochlear Implants. *Cochlear Implants International*, **7**, 148-158. <https://doi.org/10.1179/cim.2006.7.3.148>
31. Fu, Q.J. and Shannon, R.V. (2002) Frequency Mapping in Cochlear Implants. *Ear and Hearing*, **23**, 339-348. <https://doi.org/10.1097/00003446-200208000-00009>
32. Henry, B.A. and Turner, C.W. (2003) The Resolution of Complex Spectral Patterns by Cochlear Implant and

Normal-Hearing Listeners. *The Journal of the Acoustical Society of America*, **113**, 2861-2873.

<https://doi.org/10.1121/1.1561900>

33. Won, J.H., Drennan, W.R. and Rubinstein, J.T. (2007) Spectral-Ripple Resolution Correlates with Speech Reception in Noise in Cochlear Implant Users. *Journal of the Association for Research in Otolaryngology*, **8**, 384-392. <https://doi.org/10.1007/s10162-007-0085-8>
34. Wagener, K.C., Brand, T. and Kollmeier, B. (1999) Entwicklung und Evaluation eines Satztests für die deutsche Sprache Teil III: Evaluation des Oldenburger Satztests. *Zeitschrift für Audiologie*, **38**, 86-95.
35. Fredelake, S. and Hohmann, V. (2012) Factors Affecting Predicted Speech Intelligibility with Cochlear Implants in an Auditory Model for Electrical Stimulation. *Hearing Research*, **287**, 76-90. <https://doi.org/10.1016/j.heares.2012.03.005>
36. Zhu, Z., Tang, Q., Zeng, F.G., Guan, T. and Ye, D. (2012) Cochlear-Implant Spatial Selectivity with Monopolar, Bipolar and Tripolar Stimulation. *Hearing Research*, **283**, 45-58. <https://doi.org/10.1016/j.heares.2011.11.005>
37. Fielden, C.A., Kluk, K., Boyle, P.J. and McKay, C.M. (2015) The Perception of Complex Pitch in Cochlear Implants: A Comparison of Monopolar and Tripolar Stimulation. *The Journal of the Acoustical Society of America*, **138**, 2524-2536. <https://doi.org/10.1121/1.4931910>
38. Jeon, E.K., Turner, C.W., Karsten, S.A., Henry, B.A. and Gantz, B.J. (2015) Cochlear Implant Users' Spectral Ripple Resolution. *The Journal of the Acoustical Society of America*, **138**, 2350-2358. <https://doi.org/10.1121/1.4932020>
39. Horn, D.L., Won, J.H., Rubinstein, J.T. and Werner, L.A. (2017) Spectral Ripple Discrimination in Normal-Hearing Infants. *Ear and Hearing*, **38**, 212-222. <https://doi.org/10.1097/AUD.0000000000000373>
40. Aronoff, J.M. and Landsberger, D.M. (2013) The Development of a Modified Spectral Ripple Test. *The Journal of the Acoustical Society of America*, **134**, EL217-EL222. <https://doi.org/10.1121/1.4813802>