Optimization of a Spectral Contrast Enhancement Algorithm for Cochlear Implants Based on a Vowel Identification Model

Waldo Nogueira, Thilo Rode and Andreas Büchner

Abstract Speech intelligibility achieved with cochlear implants (CIs) shows large variability across different users. One reason that can explain this variability is the CI user's individual electrode nerve interface which can impact the spectral resolution they can achieve. Spectral resolution has been reported to be related to vowel and consonant recognition in CI listeners. One measure of spectral resolution is the spectral modulation threshold (SMT), which is defined as the smallest detectable spectral contrast in a stimulus. In this study we hypothesize that an algorithm that improves SMT may improve vowel identification, and consequently produce an improvement in speech understanding for CIs. With this purpose we implemented an algorithm, termed spectral contrast enhancement (SCE) that emphasizes peaks with respect to valleys in the audio spectrum. This algorithm can be configured with a single parameter: the amount of spectral contrast enhancement entitled "SCE factor". We would like to investigate whether the "SCE factor" can be individualized to each CI user. With this purpose we used a vowel identification model to predict the performance produced by the SCE algorithm with different "SCE factors" in a vowel identification task.

In five CI users the new algorithm has been evaluated using a SMT task and a vowel identification task. The tasks were performed for SCE factors of 0 (no enhancement), 2 and 4. In general it seems that increasing the SCE factor produces a decrease in performance in both the SMT threshold and vowel identification.

W. Nogueira (✉) · A. Büchner
Dept. of Otolaryngology and Hearing4all, Medical University Hannover, Hannover, Germany
e-mail: nogueiravazquez.waldo@mh-hannover.de

A. Büchner
e-mail: buechner@hoerzentrum-hannover.de

T. Rode
HörSys GmbH, Hannover, Germany
e-mail: rode.thilo@hzh-gmbh.de

P. van Dijk et al. (eds.), *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing,* Advances in Experimental Medicine and Biology 894,
DOI 10.1007/978-3-319-25474-6_11

Keywords Cochlear implant · Model · Vowel identification · Spectral contrast enhancement

1 Introduction

Cochlear implants (CIs) can restore the sense of hearing in profound deafened adults and children. CI signal processing strategies have been developed for speech understanding in quiet, such that many post-lingually deafened adults with CIs can recognize 60–80 % of sentences presented in quiet (Friesen et al. 2001). However, speech intelligibility in noise and music perception, although very variable, remain generally poor for CI listeners.

For example it is still challenging for many CI users to discriminate vowels and phonemes in a closed set identification task without background noise (Sagi et al. 2010; Svirsky et al. 2011). These difficulties might be produced by the limited spectral resolution delivered by CI devices. Spectral resolution may be degraded by the broad electrical fields created in the cochlea when the electrodes are stimulated.

In a recent study, the identification of spectrally smeared vowels and consonants was improved by spectral contrast enhancement (SCE) in a group of 166 normal hearing listeners (Alexander et al. 2011). Spectral contrast is defined as the level difference between peaks and valleys in the spectrum. In CIs, spectral contrast is degraded because of the limited number of stimulation electrodes and overlapping electric fields activating the nervous system through the bony structure of the cochlea. This might reduce the differences in amplitudes between peaks and valleys in the input making it more difficult to locate spectral dominance (i.e., formants) which provide crucial cues to speech intelligibility and instrument identification. Loizou and Poroy 2001 showed that CI users need a higher spectral contrast than normal hearing listeners in vowel identification tasks.

In this study we propose a new sound coding strategy that uses SCE for CI users. The working principle of the coding strategy can affect speech intelligibility. For example "NofM" strategies such as ACE were developed in the 1990s to separate speech signals into M sub-bands and derive envelope information from each band signal. N bands with the largest amplitude are then selected for stimulation (N out of M). One of the consequences here is that the spectral contrast of the spectrum is enhanced, as only the N maxima are retained for stimulation. In this work, we want to investigate whether additional spectral enhancement can provide with improved speech intelligibility.

When designing speech coding strategies, the large variability in speech intelligibility outcomes has to be considered. For example, two sound coding strategies can produce opposite effects in speech performance, even when the CI users are post-locutive adults and have enough experience with their CIs. One possible reason that might explain this variability is the electrode nerve interface of each individual which can impact the spectral resolution they can achieve. Spectral resolution has been reported to be closely related to vowel and consonant recognition in cochlear implant (CI) listeners (Litvak et al. 2007). One measure of spectral resolution is the spectral modulation threshold (SMT), which is defined as the smallest detectable spectral contrast in the spectral ripple stimulus (Litvak et al. 2007). In this study we hypothesize that an SCE algorithm may be able to improve SMT and therefore may also be able to improve vowel recognition.

Recently a relatively simple model of vowel identification has been used to predict confusion matrices of CI users. Models of sound perception are not only beneficial in the development of sound coding strategies to prototype the strategy and create hypotheses, but also to give more robustness to the results obtained from an evaluation in CI users. Evaluations with CI users are time consuming and results typically show large variability. In this study we use the same model developed by (Sagi et al. 2010) and (Svirsky et al. 2011) to show the potential benefits of SCE in "NofM" strategies for CIs.

2 Methods

2.1 The Signal Processing Method: SCE in NofM Strategies for CIs

The baseline or reference speech coding strategy is the advanced combinational encoder (ACE, a description of this strategy can be found in Nogueira et al. 2005). The ACE strategy can be summarized in five signal processing blocks: (1) The Fast Fourier Transform (FFT); (2) The envelope detector; (3) The NofM band selection; (4) The loudness growth function (LGF) compression and (5) The channel mapping. The new SCE strategy incorporates a new processing stage just before the NofM band selection. The goal of this stage is to enhance spectral contrast by attenuating spectral valleys while keeping spectral peaks constant. The amount of spectral contrast enhancement can be controlled by a single parameter termed SCE factor. A more detailed description of the algorithm will be published elsewhere (Nogueira et al. 2016).

2.2 Hardware Implementation

All the stimuli were computed in Matlab© using the ACE and the SCE strategies. The stimuli were output from a standard PC to the Nucleus Cochlear© implant using the Nucleus Interface Communicator (NIC). The Matlab toolbox was used to process the acoustic signals and compute the electrical stimuli delivered to the CI. For each study participant we used their clinical map, i.e., their clinical stimulation rate, comfort and threshold levels, number of maxima and frequency place allocation table. For the experiments presented in the report we used three different configurations of the SCE strategy which only differed in the amount of spectral contrast enhancement applied. The three strategies are denoted by SCE0, SCE2 and SCE4. SCE0 means no spectral enhancement and is exactly the same strategy as the clinical ACE strategy.

2.3 Experiments in Cochlear Implant Users

2.3.1 Participants

Five CI users of the Freedom/System5 system participated in this study. The relevant details for all subjects are presented in Table 1. All the test subjects used the ACE strategy in daily life and all had a good speech performance in quiet.

Each CI user participated in a study to measure SMTs and vowel idenitification performance.

2.3.2 Spectral Modulation Threshold

We used the spectral ripple test presented by (Litvak et al. 2007) to estimate the spectral modulation thresholds of each CI user. This task uses a cued two interval, two-alternative, forced choice procedure. In the first interval the standard stimulus was always presented. The standard stimulus had a flat spectrum with bandwidth extending from 350 to 5600 Hz. The signal and the second standard were randomly presented in the other two intervals. Both signals were generated in the frequency domain assuming a sampling rate of 44,100 Hz. The spectral shape of the standard and the signal were generated using the equation:

$$\left|F(f)\right| = \begin{cases} 10^{\frac{c}{2}\sin(2\pi(log_2(\frac{f}{350})f_c+\theta_0))}, \text{ for } 350 < f < 5600 \\ 0, \text{ otherwise} \end{cases}$$

where F is the amplitude of a bin with center frequency f (in Hertz), f_c is the spectral modulation frequency (in cycles/octave), and θ_0 is the starting phase. Next, noise

Table 1 Patients details participating in the study

Id.	Age	Side	Cause of deafness	Implant experience in years
1	48	Right	Sudden Hearing Loss	1
2	38	Left	Antibiotics	1.8
3	46	Left	Unknown	7.5
4	25	Right	Ototoxika	7
5	22	Left	Unkwown	3.2

was added to the phase of each bin prior to computing the inverse Fourier transform. The standard was generated using a spectral contrast *c* equal to 0. The amplitude of each stimulus was adjusted to an overall level of 60 dB sound pressure level (SPL). Independent noise stimuli were presented on each observation interval. The stimulus duration was 400 ms. A 400 ms pause was used between the stimuli.

Thresholds were estimated using an adaptive psychophysical procedure employing 60 trials. The signal contrast level was reduced after three consecutive correct responses and increased after a single incorrect response. Initially the contrast was varied in a step size of 2 dB, which was reduced to 0.5 dB after three reversals in the adaptive track (Levitt 1971). Threshold for the run was computed as the average modulation depth corresponding to the last even number of reversals, excluding the first three. Using the above procedure, modulation detection thresholds were obtained for the modulation frequency of 0.5 cycles/octave which is the one that correlates best with vowel identification (Litvak et al. 2007).

2.3.3 Vowel Identification Task

Speech understanding was assessed using a vowel identification task. Vowel stimuli consisted of eight long vowels 'baat', 'baeaet', 'beet', 'biit', 'boeoet', 'boot', 'bueuet', 'buut'. All vowels had a very similar duration of around 180 ms. The stimuli were uttered by a woman. An 8-alternative forced choice task procedure 8-AFC was created where 2 and 4 repetitions of each vowel were used for training and testing respectively. The vowels were presented at the same 60 dB SPL level as the spectral ripples with a loudness roving of+/- 1.5 dB.

2.3.4 The standard Multidimensional Phoneme Identification Model

We used a model of vowel identification to select the amount of spectral contrast enhancement SCE factor. The model is based on the multidimensional phoneme identification (MPI) model (Sagi et al. 2010; Svirsky et al. 2011). A basic block diagram of the model is presented in Fig. 1.

The model estimates relevant features from electrodograms generated by the CI sound processor. Because we are modelling a vowel identification task it makes sense to extract features related to formant frequencies. Moreover, because we are analyzing the effect of enhancing spectral contrast it seems logical to use spectral contrast features between formants. In this study the number of features was limited to two formants and therefore the MPI model is two dimensional. Next, the MPI model adds noise to the features. The variance of the noise is set based on the individual abilities of a CI user to perceive the features extracted from the electrodogram. In our implementation we used the results of the SMT task to set the noise variance. The obtained jnd from the SMT task was scaled between 0.001 and 0.5. This number was applied as the variance of the Gaussian noise applied in the MPI model and for this reason is termed jnd noise in Fig. 1.

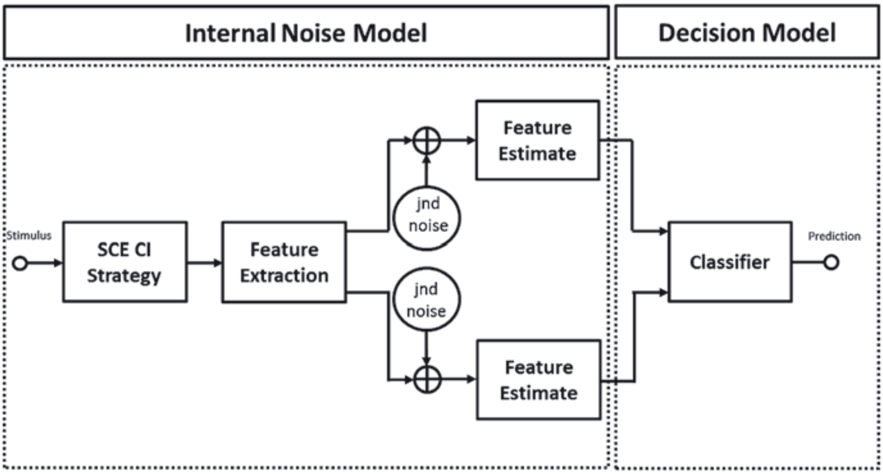

Fig. 1 Two dimensional implementation of the MPI model from Sagi et al. 2010 and Svirsky et al. 2011

3 Results

3.1 Results from the MPI model

The MPI model has been used to model the performance of the SCE strategy for five virtual CI users. All virtual CI user differs from each other in their most comfortable and threshold levels used in the map. That means that the speech processor of each virtual CI user will generate different electrodograms for the same vowels. Next, formant features were extracted from the electrodograms based on the spectral contrast between formants 1 and 2. Noise was added to the spectral contrast features (jnd noise). Three amounts of noise were added 0.01, 10 and 50% of the magnitude of the formants extracted. Figure 2 presents the results predicted by the model in percentage of correct vowels identified for five different SCE factors (0, 0.5, 1, 2 and 4).

From the modelling results it can be observed that maximum performance is achieved for SCE factors 2 and 4. The other interesting aspect is that there is no difference in performance across the five different virtual CI users for "Noise Factor" 0.001 (Fig. 2a). That means that the feature extraction (i.e. the extraction of the formants) is robust to the different electrodograms of each virtual CI user. Differences in performance between the five virtual CI users can only be observed for "Noise Factor" 0.1 and 0.5, meaning that the "jnd noise" is the only parameter explaining the differences.

From the MPI modelling results we decided to use SCE factors 2 and 4 (equivalent to increasing the original spectral contrast of the spectrum by 3 and by 5 in a dB scale) to be investigated in CI users.

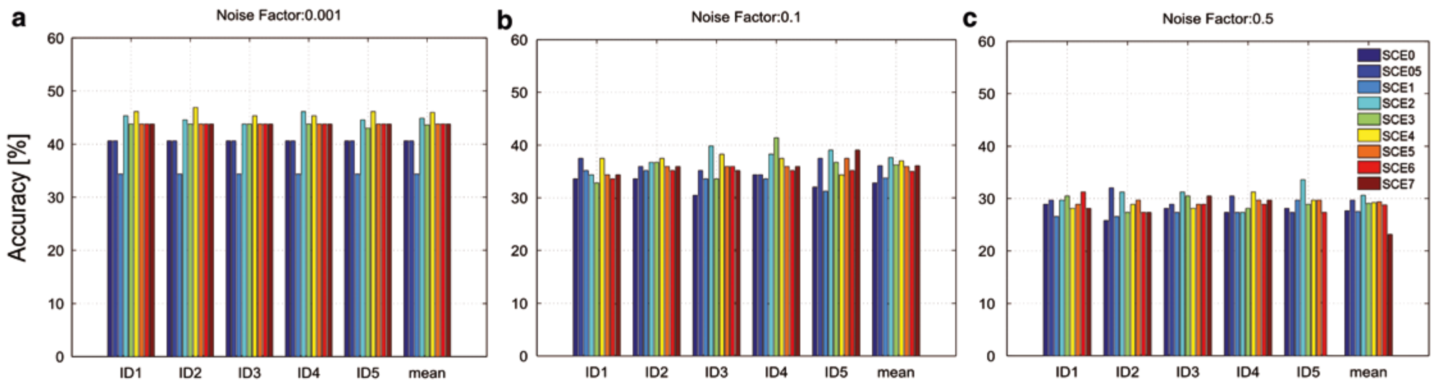

Fig. 2 Percentage of correctly identified vowels predicted by the MPI model for 5 virtual CI users using 5 different amounts of SCE (0, 0.5, 1, 2 and 4) and for different amounts of internal noise. (Noise Factors 0.001, 0.1 and 0.5)

3.2 Results Vowel Identification in CI users

Figure 3 presents the averaged results of the vowel identification task for the five subjects participating in this study.

3.3 Results Spectral Modulation Threshold in CI users

Figure 4 presents the individual and averaged results for the spectral ripple task.

Unexpectedly, additional spectral contrast, which in turn increases the spectral modulation depth, could no produce an improvement in jnd SMT.

3.4 Correlation Between Spectral Modulation Threshold and Vowel Identification

An important question for the analysis was whether the results obtained from the spectral ripple task could be used to predict the outcome of the vowel identification

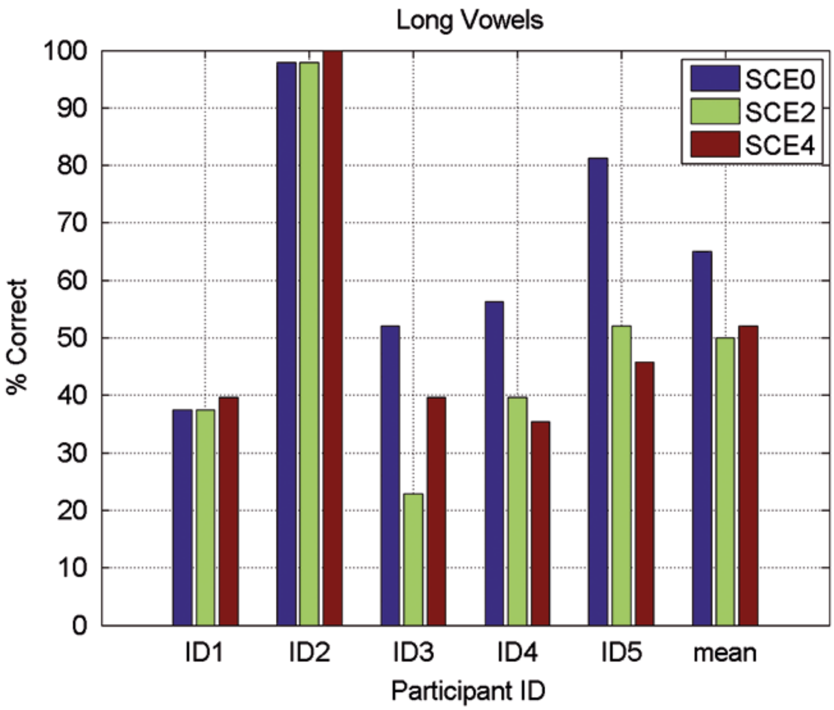

Fig. 3 Results of the vowel identification task for the three strategies (SCE0, SCE2 and SCE4) for 5 CI users and averaged as % correct

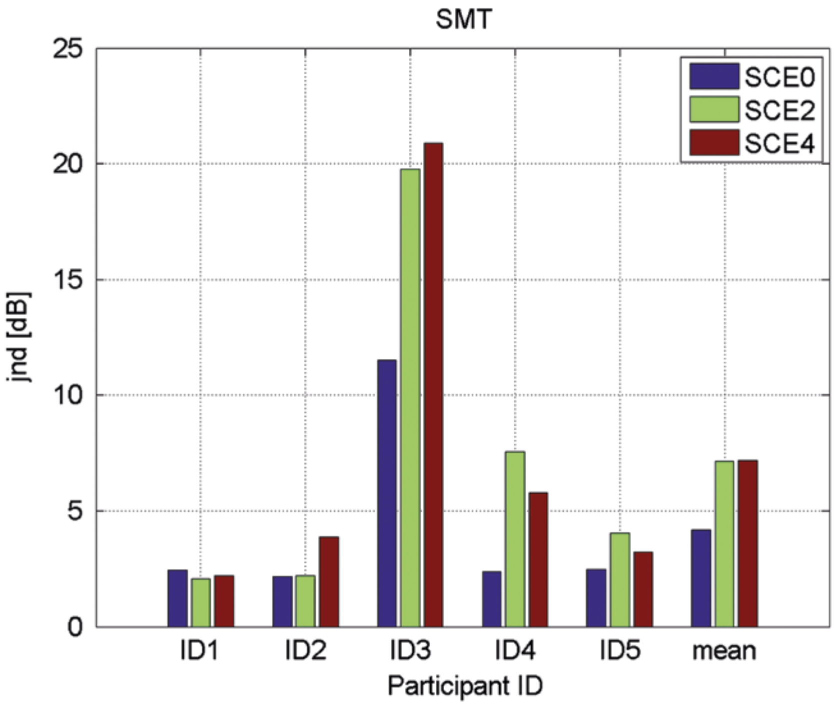

Fig. 4 Results of the spectral modulation threshold task for the three strategies (SCE0, SCE2 and SCE4) given as just noticeable differences (jnd SMT) in decibels. The lower is the jnd the better is the result

task. This can be seen in the left plot of Fig. 5 using an SCE factor of 0. Probably because of the low number of participants only a relatively weak correlation between the two measures was observed.

In the same manner, the middle and right plots in Fig. 5 show the relationship between the improvements of the two tasks comparing the results using an SCE factor 0 to those using SCE factors 2 and 4 respectively. Again, the correlation observed is weak but still a trend for the relationship can be seen. It seems that for the SCE factors used the decline in performance in the SMT is somewhat connected to a decline in performance in the vowel identification task. It remains unclear if an increase of the number of participants would confirm this trend.

4 Discussion

A new sound coding strategy that enhances spectral contrast has been designed. The amount of spectral contrast enhancement can be controlled by a single parameter termed SCE factor.

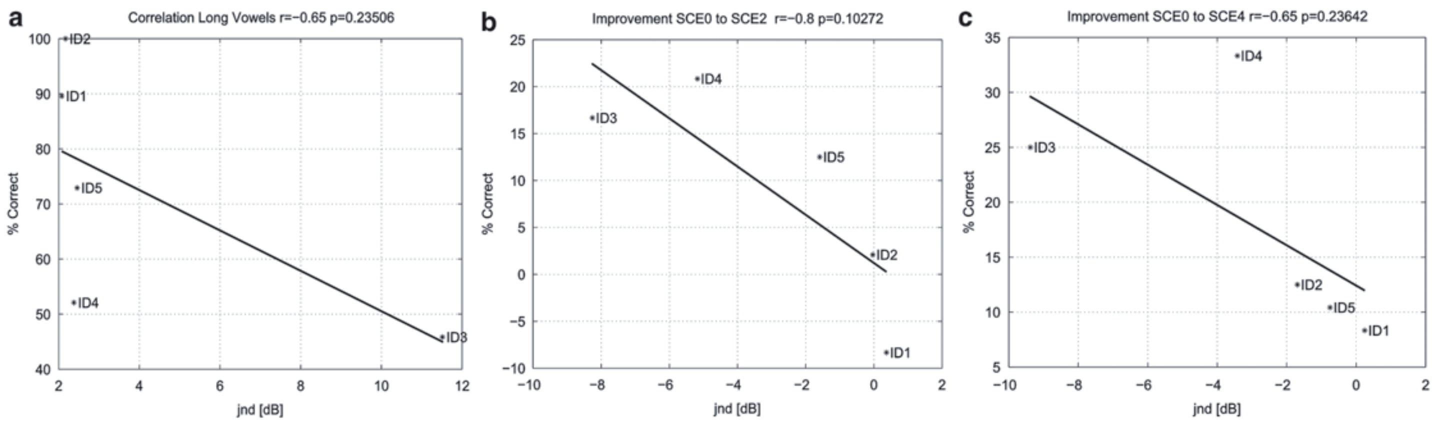

Fig. 5 **a** Correlation between vowel identification performance and jnd SMT in dB. **b** Correlation between the difference (SCE0–SCE2) in vowel identification performance and the difference in jnd SMT (SCE0–SCE2). **c** Correlation between the difference (SCE0–SCE4) in vowel identification performance and the difference in jnd SMT (SCE0–SCE4)

A model of vowel identification has been used to investigate the effect of SCE on vowel identification. The model predicts that increasing the amount of SCE increases vowel identification accuracy. Based on these results we decided to use SCE factors 2 and 4 (equivalent to increasing the original spectral contrast of the spectrum by 3 and by 5 in a dB scale).

The new SCE strategy has been evaluated in CI users. Results from a vowel identification task and a SMT task in five CI users show differences in vowel identification scores for different SCE factors. In general, it seems that SCE produces a detrimental effect in spectral modulation detection and vowel identification in CI users. These results are contrary to the model predictions. Previous studies in the literature give reasons to believe that spectral contrast enhancement would result in a benefit for the chosen tasks. It is possible that spectral valleys are attenuated too much and relevant information required by the CI users to understand speech is lost. These effects are not taken into account by the MPI model, and this could explain the contradictory results between experiments and CI users and modelling results. Still, it is possible that the SCE factors selected where too high, for this reason we think that a follow-up study should investigate whether lower amounts SCE can provide improvements in CI users.

Acknowledgments The authors would like to thank the subjects who have participated in the experiments and the two anonymous reviewers for their comments on different versions of this manuscript. This work was supported by the DFG Cluster of Excellence EXC 1077/1 "Hearing4all" and Cochlear.

References

Alexander JM, Jenison RL, Kluender KR (2011) Real-time contrast enhancement to improve speech recognition. PLoS One 6(9):e24630. http://doi.org/10.1371/journal.pone.0024630

Friesen LM, Shannon RV, Baskent D, Wang X (2001) Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. J Acoust Soc Am 110(2):1150. http://doi.org/10.1121/1.1381538

Levitt H (1971) Transformed up-down methods in psychoacoustics. J Acoust Soc Am 49(2B):467–477. http://doi.org/ http://dx.doi.org/10.1121/1.1912375

Litvak LM, Spahr AJ, Saoji AA, Fridman GY (2007) Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners. J Acoust Soc Am 122(2):982–991. http://doi.org/ http://dx.doi.org/10.1121/1.2749413

Loizou PC, Poroy O (2001) Minimum spectral contrast needed for vowel identification by normal hearing and cochlear implant listeners. J Acoust Soc Am 110(3):1619–1627. http://doi.org/ http://dx.doi.org/10.1121/1.1388004
Nogueira W, Rode T, Büchner A (2016) Spectral contrast enhancement improves speech intelligibility in 1 noise in NofM strategies for cochlear implants. J Acoust Soc Am
Nogueira W, Büchner A, Lenarz T, Edler B (2005) A psychoacoustic "NofM"-type speech coding strategy for cochlear implants. Eurasip J Appl Signal Process 2005(18):3044–3059
Sagi E, Meyer TA, Kaiser AR, Teoh SW, Svirsky MA (2010) A mathematical model of vowel identification by users of cochlear implants. J Acoust Soc Am 127(2):1069–1083. http://doi.org/ http://dx.doi.org/10.1121/1.3277215
Svirsky MA, Sagi E, Meyer TA, Kaiser AR, Teoh SW (2011) A mathematical model of medial consonant identification by cochlear implant users. J Acoust Soc Am 129(4):2191–2200. http://doi.org/ http://dx.doi.org/10.1121/1.3531806