Cochlear implant simulator with independent representation of the full spiral ganglion

Jacques A. Grange, John F. Culling, and Naomi S. L. Harris
School of Psychology, Cardiff University, CF103AT, Cardiff, United Kingdom
grangeja@cardiff.ac.uk, cullingj@cardiff.ac.uk, naomislharris@gmail.com

Sven Bergfeld
Department of Cognitive Neuroscience, Bielefeld University, 33615 Bielefeld, Germany
svenbergfeld@posteo.de

Abstract: In cochlear implant simulation with vocoders, narrow-band carriers deliver the envelopes from each analysis band to the cochlear positions of the simulated electrodes. However, this approach does not faithfully represent the continuous nature of the spiral ganglion. The proposed "SPIRAL" vocoder simulates current spread by mixing all envelopes across many tonal carriers. SPIRAL demonstrated that the classic finding of reduced speech-intelligibility benefit with additional electrodes could be due to current spread. SPIRAL produced lower speech reception thresholds than an equivalent noise vocoder. These thresholds are stable for between 20 and 160 carriers.

[QJF]
Date Received: May 10, 2017 **Date Accepted:** October 17, 2017

1. Introduction

All vocoders used to simulate hearing with a cochlear implant (CI) mimic the bandpass filtering and envelope extraction of CI processors. However, various approaches have been taken to delivering the envelopes to appropriate places in a normal-hearing ear via narrow-band carriers. In Shannon *et al.* (1995), a broadband noise is first modulated by each envelope, then bandpass filtered. Alternatively, one can modulate a set of tones (Dorman *et al.*, 2002). The frequencies of the narrow-band carriers correspond to the cochlear places of the simulated electrodes.

Electric fields generated by individual CI electrodes can overlap substantially, producing a mixture of information from different channels at each point on the spiral ganglion. This spectral smearing phenomenon is due to current spread from each electrode and along the spiral ganglion. Many studies (e.g., Friesen *et al.*, 2001; Dorman *et al.*, 2002) have used vocoders that do not simulate current spread and have one carrier per analysis channel. To simulate current spread in a noise vocoder, Fu and Nogaki (2005) used noise carriers with triangular spectra, featuring exponential decay slopes to either side of the center frequency, thereby creating an output mixture of envelopes that varies continuously across frequency. However, Whitmal *et al.* (2007) and Oxenham and Kreft (2014) argued that a tone vocoder may better reflect characteristics of electric hearing than a noise vocoder, because narrow-band noise carries intrinsic modulation that interferes with that of the speech envelope. Crew *et al.* (2012) and Oxenham and Kreft (2014) proposed methods of incorporating current spread into tone-vocoded CI simulations. These schemes involve modulating the carriers by mixed envelopes, each carrier receiving a weighted contribution from the envelopes extracted at the analysis stage. The weighting assumes exponential decay of current along the spiral ganglion and neglects the effect of the distance between electrode array and spiral ganglion.

Although these tone vocoders can represent current spread without adding spurious modulations, they still have two potential problems. First, the number of carriers is yoked to the number of analysis bands, making it impossible to independently examine the effect of changing the number of analysis bands (i.e., electrodes); if performance is affected by the spectral density of the vocoder, this effect may mask the influence of the number of processor channels. Second, they employ a single tone per processor channel, whereas in a real implantation, all parts of the spiral ganglion would receive unique envelope mixes. Combining (1) the continuous envelope mixtures of the Fu and Nogaki approach with (2) the absence of carrier-induced fluctuations of the Oxenham and Kreft approach should produce a vocoder that more faithfully models electrode-nerve coupling and the effects of current spread.

Current spread has two potential effects: it is thought to elevate speech reception thresholds (SRTs), but is also thought to limit the effective number of channels. Srinivasan *et al.* (2013) showed that tripolar stimulation, designed to reduce current spread, does indeed lower SRTs. Threshold elevation has been demonstrated using controlled degrees of simulated current spread by Fu and Nogaki (2005) and Bingabr *et al.* (2008), amongst others. The hypothesis that current spread limits the effective number of channels comes from Friesen *et al.* (2001). They found that performance reached a plateau above ~8 channels for CI users, but continued to improve with additional channels in simulations with no current spread. They proposed, and it has since been widely believed, that this plateau reflects a limit to the number of effective electrodes and that current spread is most likely responsible. Some support for this interpretation was provided by Bingabr *et al.* (2008), who observed a drop in performance for 8 and 16 simulated electrodes when current spread was increased to the highest level that they used, but not for 4 electrodes. However, the plausibility of the hypothesis of Friesen *et al.* is yet to be supported by simulations that employ a tone vocoder.

This report presents a new "SPIRAL" vocoder, designed to address all the issues reported above. We present below a description of the vocoder and two experiments that demonstrate that SPIRAL can simulate a reduction of number of effective electrodes by current spread, and also that stable results can be obtained with a range of parameter settings.

2. The SPIRAL vocoder

Figure 1 shows a schematic representation of SPIRAL. The analysis stage passes the signal through a bank of rectangular 512-point finite impulse response (FIR) filters that span 120–8658 Hz, with center frequencies equally spaced on an ERB scale (Moore and Glasberg, 1983). Envelopes are extracted by half-wave rectification and low-pass filtering. The default modulation cut-off frequency is 50 Hz, based on the finding of Drullman *et al.* (1994) that modulation frequencies above 30 Hz contribute little to speech-in-noise intelligibility, and to remove potentially resolved sidebands. The frequencies corresponding to the simulated electrode positions are then determined. At the reconstruction stage, an exponential decay function represents current spread. In contrast with the vocoder of Oxenham and Kreft (one tonal carrier per analysis channel), SPIRAL employs a number of tonal carriers that is independent of the analysis stage to represent the full extent and the continuous nature of the spiral ganglion. Excitation spreads from the simulated electrodes to an ERB-spaced array of carriers in the 20–20 000 Hz frequency range. Each random-phase carrier is modulated by a mixed envelope, with a weighted contribution from each of the envelopes extracted at the analysis stage. The modulated carriers are summed to form the vocoded signal. Following Oxenham and Kreft, current spread is expressed as a negative current decay slope S (in dB/oct.). For analysis band i center frequency $CF(i)$ and carrier tone frequency $F(j)$, a weight $W(i,j)$ is computed using Eq. (1). With a set of n extracted envelopes $E(i)$, the mixed envelope $M(j)$ used to modulate tonal carrier j is computed using Eq. (2),

$$W(i,j) = 10^{[(S/10)\times \mathrm{abs}(\log 2(CF(i)/F(j)))]}, \tag{1}$$

$$M(j) = \sqrt{\sum_{i=1}^{n} \left(W(i,j) * E(i)^2 \right)}. \tag{2}$$

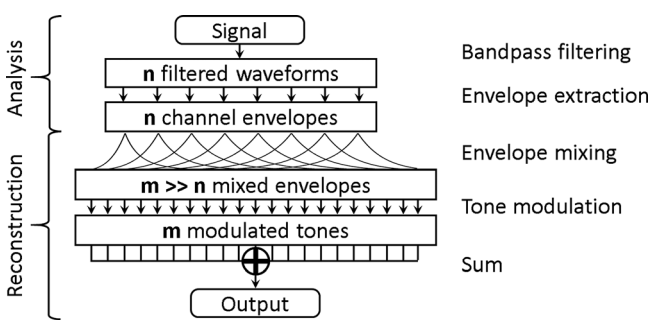

Fig. 1. SPIRAL vocoder flow diagram (see text for description).

SPIRAL can bridge the gap between a traditional tone vocoder and a noise vocoder, because using many tones will produce a densely-sampled spectrum, but without the intrinsic modulation of a noise source. SPIRAL is similar, in its concept, to the Churchill *et al.* (2014) harmonic-complex vocoder in that it uses many tonal (20+) carriers, but the energy distribution of the SPIRAL carriers is spread along the spectrum in a manner more consistent with the even distribution of ganglion cells along the spiral ganglion. SPIRAL does not incorporate the pulsatile nature of CI stimulation because such added complexity is yet to be proven essential in CI simulations. SPIRAL was programmed in MATLAB (The MathWorks, Natick, MA, USA) and is made available as supplemental material.[1]

3. Experimental rationale

The logic of the hypothesis of Friesen *et al.* (2001) that current spread reduces the number of effective electrodes, dictates that performance should plateau at a lower number of electrodes when simulated current spread is increased. Experiment 1 attempted to demonstrate such an effect using controlled simulation of current spread. Both the SPIRAL and Oxenham and Kreft (O&K) vocoders were used in order to test whether the independent representation of the spiral ganglion, provided by SPIRAL, is necessary to produce the effect.

An inherent property of SPIRAL is that when a high spectral density of tonal carriers is used, beats between carriers occur that could be perceived as amplitude modulation. Experiment 2 tested whether such beats are detrimental to speech intelligibility, within the operating range of SPIRAL. With a very high number of carriers, SPIRAL should approximate the Fu and Nogaki (2005) noise vocoder, which was consequently used for comparison.

4. Materials and method

30 young adults (aged 19–23, mean 20), self-reported as normal-hearing, were recruited from the Cardiff University undergraduate population. Each was briefed prior to providing consent to perform an experiment, debriefed after completion and compensated for their time, in accordance with our institutional review board rules. Twenty participated in the first experiment and ten in the second. Each participant was tested in a single-walled audiometric booth. A monitor and keyboard enabled participants to follow instructions and enter transcripts of the target speech material. Stimuli were prepared by a custom MATLAB program and presented diotically via Sennheiser HD650 headphones. Sentences from the IEEE corpus were employed as target speech, each mixed with stationary speech-shaped noise. The vocoded output was scaled such that total sound level would be approximately 65 dBA, regardless of speech-to-noise ratio (SNR).

A 50-Hz modulation cut-off frequency was used in both experiments. Experiment 1 measured percent-correct word recognition ("performance") as a function of number of electrodes. We simulated current spread with the O&K vocoder for which the number of carrier tones equaled the number of analysis channels and the SPIRAL vocoder using 80 carrier tones. Pilot experiments exhibited ceiling or floor effects at any fixed SNR within the range of simulated current spread. To address this issue, a new methodology enabled comparison of psychometric functions within the same intelligibility range by normalizing maximum performance. We measured SRTs, (see procedure common to both experiments below) for 70% performance with all 20 electrodes activated, at spread levels ordered in increasing severity. For each participant, and within a spread condition, the mean of three SRTs set the SNR for the main performance measure. With a fixed material order, each participant was allocated a random permutation of 21 conditions: three spread levels (−200 dB/oct. for no spread and −16 or −8 dB/oct. for mild and severe spread) by seven numbers of electrodes (3, 4, 6, 8, 10, 15, and 20). With 15 sentences per condition and 9 prior SRT runs, there were 405 sentences in all.

Experiments 2 measured SRTs for 50% performance with 20 electrodes. Experiment 2 had ten conditions: two current spread levels (−8 dB/oct. and −16 dB/oct.) × five carrier conditions (SPIRAL with 20, 40, 80, or 160 carriers and noise vocoder). The noise vocoder was similar to the Fu and Nogaki (2005) vocoder, except that the analysis stage used the same FIR filters as SPIRAL. The sentence list order was fixed and the randomly ordered conditions rotated against material. After four practice runs, two SRTs were measured per condition and averaged.

Throughout, SRTs followed a one-up, one-down adaptive procedure converging on the target performance (see Culling, 2016). Each SRT used a list of ten sentences. Initially, sentences were randomly selected from the list and presented at −8 dB

SNR, such that no word would be intelligible. Sentences were presented at 4 dB increments in SNR, until participants recognized more than one or two of the five keywords for the 50% or 70% performance measures, respectively. This triggered the adaptive phase, in which the remaining sentences were presented only once and, depending on performance (above or below the target), SNR increased or decreased in 2 dB steps. The last eight computed SNRs were averaged to calculate the SRT.

5. Results

Experiment 1 performance measures for the O&K and SPIRAL vocoders are shown in the left and right panels of Fig. 2, respectively. The different psychometric functions are largely superimposed with the O&K vocoder, but are clearly separated with SPIRAL. In order to quantify this difference, logistic functions were fitted for each level of spread. Analyses of variance (ANOVAs) operated on the inflection point positions for each vocoder showed no significant effect of current spread for the O&K vocoder [$F(2,18) = 0.73$, $p = 0.50$], but a significant shift to a lower number of electrodes with increased current spread for SPIRAL [$F(2,18) = 6.48$, $p < 0.01$]. Comparing means across all conditions revealed an interaction between level of current spread and vocoder type [$F(2,36) = 3.53$, $p < 0.05$]. The mean SNRs used to test participants, in the key of Fig. 2, show the significant adjustment required to normalize performance [$F(2,36) = 308$, $p < 0.001$]. The SNRs required for normalization did not differ significantly between vocoders [$F(1,18) = 0.11$, $p = 0.74$].

Experiment 2 mean SRTs as a function of number of carriers are displayed in Fig. 3. A 2×4 ANOVA, computed across the eight SPIRAL conditions, confirmed the detrimental main effect of current spread [$F(1,9) = 169.5$, $p < 0.001$]. There was no effect of number of carriers [$F(3,27) = 0.132$, $p = 0.94$]. Following an averaging of SRTs across numbers of carriers for the SPIRAL conditions, the noise vocoder consistently led to significantly higher thresholds than SPIRAL [$F(1,9) = 90.7$, $p < 0.001$]. An interaction between current spread and vocoder type further revealed a significantly larger increase in SRT with current spread for the noise vocoder [$F(1,9) = 22.1$, $p < 0.001$].

6. Discussion

Our aim was to devise a vocoder that simulates current spread across a continuous spiral ganglion without the detrimental effects on speech intelligibility of random modulations inherent to noise vocoders. SPIRAL employs many (20+) tonal carriers spread over the entire cochlear frequency range, which provides an improved representation of the spiral ganglion compared to the sparse representation of a conventional tone vocoder. Up to 160 carriers can be employed in time-efficient SRT experiments with on-the-fly processing of short sentences.

The main point of interest in this report is that current spread may not simply affect the effective signal-to-noise ratio, it may also reduce the benefit of increasing the number of electrodes. Experiment 1, for which performance with 20 electrodes was normalized, examined performance as a function of the number of activated electrodes

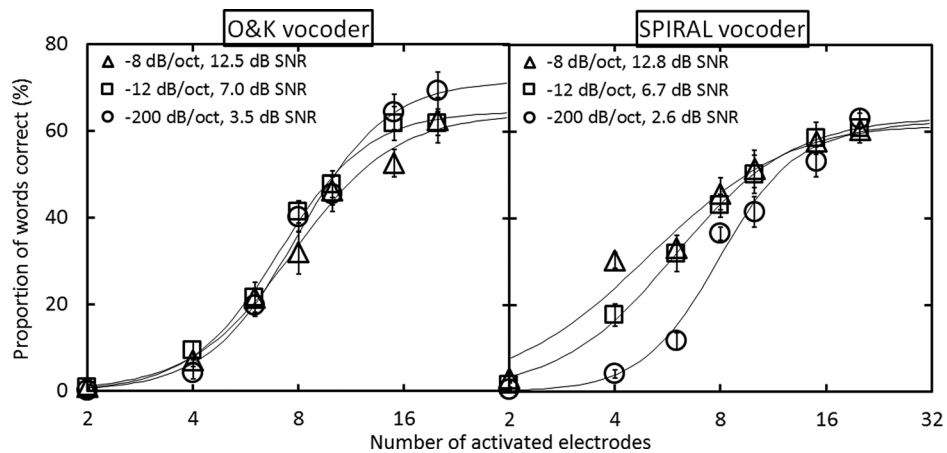

Fig. 2. Experiment 1 performance (proportion of IEEE-sentence words correctly understood), normalized for 20 activated electrodes, as a function of number of electrodes and current decay slope in one of two simulations with normal-hearing listeners: Oxenham and Kreft (2014) tone vocoder (left panel) and SPIRAL (right panel). Continuous lines represent the best logistic-function fit for each psychometric function. Error bars are the standard error of the mean.

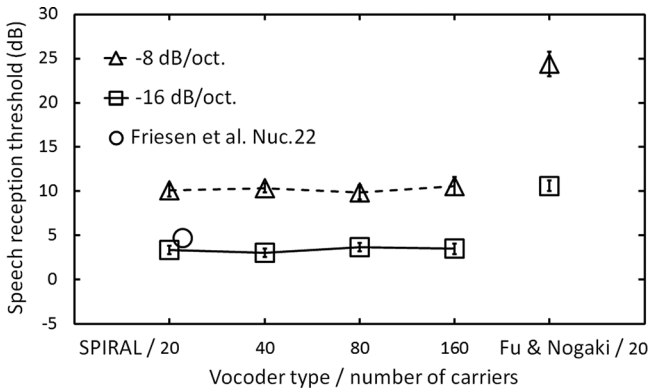

Fig. 3. Experiment 2 SRTs for IEEE sentences in speech-shaped noise as a function of number of SPIRAL carriers and for noise vocoding (Fu and Nogaki, 2005), for two current decay slopes. An SRT extracted from Friesen *et al.* (2001) for Nucleus 22 Cochlear CI users (HINT sentences) is also displayed.

and degree of current spread simulated with the O&K and SPIRAL vocoders. For both vocoders, the 10-dB elevation of SRTs with simulated current spread is in agreement with the data of Friesen *et al.* and consistent with the elevation of thresholds found with noise vocoders. For the simulation to support the hypothesis of Friesen *et al.* that current spread reduces the number of effective electrodes, two things could happen with increasing current spread, after performance normalization: (1) the logistic function slopes could remain the same with functions shifted down the number-of-electrode axis or (2) the logistic function slope could decrease. In both cases, the plateauing of performance would occur at a lower number of electrodes. The O&K vocoder showed little effect of current spread on the shape of the logistic functions. In contrast, SPIRAL produced shallower logistic functions at higher current spread, implying a reduced benefit of larger numbers of electrodes. This reduction in slope, which appears consistent with the hypothesis of Friesen *et al.*, must be related (1) to SPIRAL varying only the number of simulated electrodes (with a fixed number of carriers) and/or (2) to the use of a larger number of carriers by SPIRAL than by the O&K vocoder.

Experiment 2 examined the effect of the number of carriers, comparing results with the Fu and Nogaki noise vocoder. The SPIRAL data shows that SPIRAL can make use of 20 to 160 carriers with no effect on SRTs, thereby establishing an operating range within which beats between carriers do not interfere with speech intelligibility. The Fu and Nogaki noise vocoder consistently led to higher SRTs than SPIRAL, an effect that increased with greater current spread. This may have been caused not only by noise-carrier fluctuations, but by their interaction with current spread; the combination seemed to result in a sensation of "grittiness" in the finished stimulus. SRTs from SPIRAL simulation with a current decay slope of −16 dB/oct. (representative of current spread with a Nucleus 22 CI according to Nelson *et al.*, 2011) are a much closer match to the CI data of Friesen *et al.* than noise-vocoded simulations. However, the validity of such a comparison is limited by differing degrees of adaptation and material difficulty.

In SPIRAL, individual carriers can be conceived as representing subpopulations of ganglion cells. This opens the opportunity to simulate variations in spiral ganglion health. Dead or partially depleted regions can be simulated by the removal or dampening of the corresponding carriers. With the reconstruction stage decoupled from the analysis stage, variations in insertion depth and frequency-place mapping can easily be represented by application of the Greenwood formula (Greenwood, 1990). Thus, SPIRAL provides a means of CI simulation with an improved representation of electrode to spiral ganglion coupling and high experimental flexibility.

Acknowledgments

This research was supported by the Oticon Foundation.

References and links

[1]See supplementary material at http://dx.doi.org/10.1121/1.5009602 for the SPIRAL MATLAB code (R2017a).

Bingabr, M., Espinoza-Varas, B., and Loizou, P. (**2008**). "Simulating the effect of spread of excitation in cochlear implants," Hear. Res. **241**, 73–79.

Churchill, T., Kan, A., Goupell, M., and Litovsky, R. (**2014**). "Spatial hearing benefits demonstrated with presentation of acoustic temporal fine structure cues in bilateral cochlear implant listeners," J. Acoust. Soc. Am. **136**, 1246–1256.

Crew, J. D., Galvin, J. J., and Fu, Q.-J. J. (**2012**). "Channel interaction limits melodic pitch perception in simulated cochlear implants," J. Acoust. Soc. Am. **132**, EL429–EL435.

Culling, J. F. (**2016**). "Speech intelligibility in virtual restaurants," J. Acoust. Soc. Am. **140**, 2418–2426.

Dorman, M. F., Loizou, P. C., Spahr, A. J., and Maloff, E. (**2002**). "A comparison of the speech understanding provided by acoustic models of fixed-channel and channel-picking signal processors for cochlear implants," J. Speech. Lang. Hear. Res. **45**, 783–788.

Drullman, R., Festen, J., and Plomp, R. (**1994**). "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053–1064.

Friesen, L., Shannon, R., Baskent, D., and 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**, 1150–1163.

Fu, Q. J., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19–27.

Greenwood, D. D. (**1990**). "A cochlear frequency position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Moore, B., and Glasberg, B. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.

Nelson, D. A., Kreft, H. A., Anderson, E. S., and Donaldson, G. S. (**2011**). "Spatial tuning curves from apical, middle, and basal electrodes in cochlear implant users," J. Acoust. Soc. Am. **129**, 3916–3933.

Oxenham, A., and Kreft, H. (**2014**). "Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing," Trends Hear. **18**, 1–14.

Shannon, R., Zeng, F., and Kamath, V. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Srinivasan, A., Padilla, M., Shannon, R., and Landsberger, D. (**2013**). "Improving speech perception in noise with current focusing in cochlear implant users," Hear. Res. **299**, 29–36.

Whitmal, N., Poissant, S., Freyman, R., and Helfer, K. (**2007**). "Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience," J. Acoust. Soc. Am. **122**, 2376–2788.