

Speech recognition as a function of the number of channels in perimodiolar electrode recipients

Katelyn A. Berg

Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, 1215 21st Avenue South, Nashville, Tennessee 37232, USA

Jack H. Noble and Benoit M. Dawant

Department of Electrical Engineering & Computer Science, Vanderbilt University, 2201 West End Avenue, Nashville, Tennessee 37235, USA

Robert T. Dwyer

Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, 1215 21st Avenue South, Nashville, Tennessee 37232, USA

Robert F. Labadie

Department of Otolaryngology, Vanderbilt University Medical Center, 1215 21st Avenue South, Nashville, Tennessee 37232, USA

René H. Gifford^{a)}

Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, 1215 21st Avenue South, Nashville, Tennessee 37232, USA

(Received 10 August 2018; revised 4 February 2019; accepted 9 February 2019; published online 26 March 2019)

This study investigated the number of channels needed for maximum speech understanding and sound quality in 30 adult cochlear implant (CI) recipients with perimodiolar electrode arrays verified via imaging to be completely within scala tympani (ST). Performance was assessed using a continuous interleaved sampling (CIS) strategy with 4, 8, 10, and 16 channels and n-of-m with 16 maxima. Listeners were administered auditory tasks of speech understanding [monosyllables, sentences (quiet and +5 dB signal-to-noise ratio, SNR), vowels, consonants], spectral modulation detection, as well as subjective estimates of sound quality. Results were as follows: (1) significant performance gains were observed for speech in quiet (monosyllables and sentences) with 16- as compared to 8-channel CIS, (2) 16 channels in a 16-of-m strategy yielded significantly higher outcomes than 16-channel CIS for sentences in noise (percent correct and subjective sound quality) and spectral modulation detection, (3) 16 channels in a 16-of-m strategy yielded significantly higher outcomes as compared to 8- and 10-channel CIS for monosyllables, sentences (quiet and noise), consonants, spectral modulation detection, and subjective sound quality, (4) 16 versus 8 maxima yielded significantly higher speech recognition for monosyllables and sentences in noise using an n-of-m strategy, and (5) the degree of benefit afforded by 16 versus 8 maxima was inversely correlated with mean electrode-to-modiolus distance. These data demonstrate greater channel independence with perimodiolar electrode arrays as compared to previous studies with straight electrodes and warrant further investigation of the minimum number of maxima and number of channels needed for maximum auditory outcomes. © 2019 Acoustical Society of America.

<https://doi.org/10.1121/1.5092350>

[DDO]

Pages: 1556–1564

I. INTRODUCTION

One of the goals of cochlear implant (CI) programming is to maximize the number of independent neural populations that are stimulated. A number of studies have suggested, however, that no more than 4–8 independent sites may be available, even for arrays with as many as 22 electrodes (e.g., Fishman *et al.*, 1997; Friesen *et al.*, 2001; Shannon *et al.*, 2011). Most likely, the number of independent sites is limited by substantial overlaps in the electric fields from adjacent electrodes, commonly termed channel interaction.

Devices using non-simultaneous, interleaved stimulation experience channel interaction caused by neural spread of excitation, possibly due to the proximity of electrode contacts or the stimulation rate. Channel interaction is unavoidable for intracochlear electrical stimulation as the electrodes are in a highly conductive fluid and are relatively far away from the target neurons in the modiolus. In fact, the electrical pulses from a CI may result in spread of excitation functions spanning one third or more of the array (e.g., Hughes *et al.*, 2013; Padilla and Landsberger, 2016), and additionally are relatively far away from the target neurons in the modiolus.

Researchers have investigated the number of electrodes needed to maximize speech understanding performance,

^{a)}Electronic mail: rene.gifford@Vanderbilt.edu

both in normal hearing (NH) subjects listening to vocoded CI simulations and in CI recipients. For CI simulations with NH subjects, four noise-vocoded channels are sufficient for recognition of consonants, vowels, and simple sentences (Shannon *et al.*, 1995). Loizou *et al.* (2000a) found that, for CI simulations using an “n-of-m” strategy, performance plateaued beyond 2–6 channels for more difficult speech measures such as NU-6 words, CUNY sentences, and TIMIT sentences. For speech in noise, NH performance for noise-vocoded CI simulations continued to improve up to 20 channels (Dorman *et al.*, 1998; Friesen *et al.*, 2001).

Focusing on studies using CI recipients, Fishman *et al.* (1997) found asymptotic speech recognition with five channels for consonants, eight channels for vowels and monosyllables, and four channels for CUNY sentences. Friesen *et al.* (2001) further investigated the number of channels required for speech recognition both in quiet and in noise for both Clarion and N22 CI recipients and found no significant improvement beyond eight channels for vowels, consonants, and marginally significant improvements were found between seven and ten electrodes for monosyllables and sentences.

In summary, these studies found no improvement in CI performance for recognition of consonants, vowels, monosyllabic words, sentences, and speech in noise beyond 4–8 channels, with NH performance with CI simulations continuing to improve up to 20 channels for speech recognition in noise. This performance plateau could be explained by limited independent neural populations, channel interaction, ceiling effects on some tasks, and limitations of envelope-based speech coding. Though not explicitly stated in earlier studies, an implicit assumption in this experimental design is that the electrode is located in scala tympani (ST) and has a uniform electrode-to-modiolus interface across the array. However, these assumptions are not accurate for many CI recipients. Research has shown that only 89% of the lateral wall and 58% of perimodiolar electrodes are completely in ST (Wanna *et al.*, 2014), with perimodiolar electrodes and cochleostomy approaches most commonly resulting in an electrode crossing from ST to scala vestibuli (SV) (O’Connell *et al.*, 2016). Additionally, there is growing evidence that a uniform electrode-to-modiolus distance is not achieved along the array, even for lateral wall electrodes (Noble *et al.*, 2012, 2014). Thus it is the case that all channels are likely not created equal as the number of independent channels available to a CI recipient may be impacted by variable electrode-to-modiolus distance across the array. Further, many of the previous studies were completed with either first generation (Fishman *et al.*, 1997; Friesen *et al.*, 2001) or second generation (Shannon *et al.*, 2011) CI technology implanted using more traumatic surgical approaches in patients meeting prior, more conservative implant indications. A recent study was completed by Croghan *et al.* (2017) with newer generation CI recipients. For sentence recognition at various signal-to-noise ratios (SNRs), they reported higher outcomes with 22 active electrodes with 8 maxima versus 12 active electrodes with 8 maxima. However, this study kept the number of maxima constant at 8 maxima in an n-of-m strategy irrespective of the number of active electrodes and did not have image-based

confirmation of electrode location for the 9 perimodiolar electrode recipients—an electrode commonly documented to result in an ST-SV location (Wanna *et al.*, 2014).

The objective of this study was to investigate the number of electrodes needed for asymptotic speech understanding, sound quality, and spectral resolution using CI recipients implanted under current labeled indications, atraumatic surgical techniques, and perimodiolar electrodes completely in ST. We hypothesized that CI recipients with perimodiolar electrodes in ST would be able to take advantage of a greater number of channels (>8) for speech recognition, sound quality, and spectral resolution.

II. EXPERIMENT I

A. Study participants

Eleven postlingually deafened adult CI users with perimodiolar electrodes were recruited for participation. All participants included in data analysis had all 22 electrical contacts within ST as confirmed by postoperative CT scans and image analysis (Noble *et al.*, 2012). Inclusion criteria required at least 6 months of CI experience and ≥ 18 active electrodes in their clinical map. Table I provides demographic information.

B. Methods

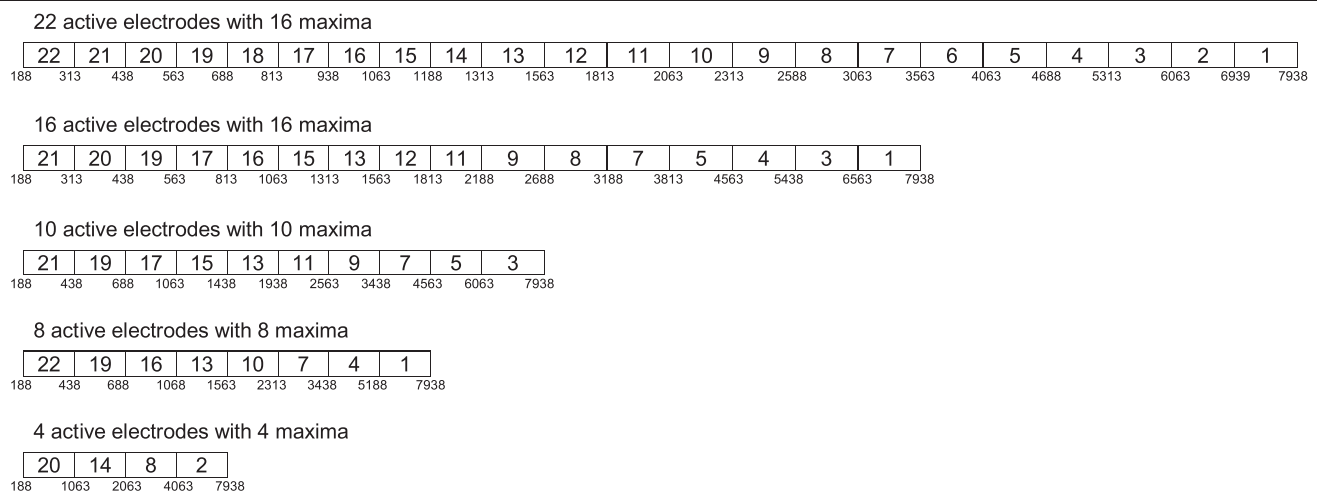
All experimental activities were completed in accordance with IRB approved protocols at Vanderbilt University and Vanderbilt University Medical Center. Five CI programs using 4, 8, 10, 16, and “all on” active electrodes were created to replicate the spatially selective programs described by Friesen *et al.* (2001). Refer to Table II for specific electrodes activated to achieve the spatially selective maps. Similar to Friesen *et al.* (2001), the input frequency range was held constant across the experimental CI programs. All parameters were left as programmed in the participants’ own maps, except the number of maxima was changed to be equal in number to the active electrodes consistent with continuous interleaved sampling (CIS; Wilson *et al.*, 1991). The one exception was for the all on condition, for which the number of maxima was set to 16 due to clinical software limitations. For the all on condition, the participants had anywhere from 19 to 22 active electrodes, consistent with their everyday map. An important methodological difference between the current and previous studies is the use of CIS and the ACE n-of-m processing strategies instead of previous generation strategies (e.g., SPEAK, SAS). All participants were already using a channel stimulation rate of 900 Hz with 25- μ s pulse duration with their everyday map and this was kept constant for this study. Upper stimulation levels were globally adjusted using the participants’ own maps to achieve equivalent loudness across all experimental maps. Threshold levels were not adjusted from the participant’s own map. All front-end processing features were deactivated, with the exception of Autosensitivity Control (ASC) and adaptive dynamic range optimization (ADRO).

Electrode condition and assessment measure order were both randomized using a Latin Square design. All testing

TABLE I. Experiments I and II demographics including age, sex, electrode type, CI experience (months), number of active electrodes (deactivated electrodes in parentheses), scalar location, mean electrode-to-modiolus distance (mm), and CNC and AzBio +5 scores with 8 and 16 maxima, respectively. Subjects participating in experiment 1 are unshaded and subjects participating only in experiment 2 are shaded. Unknown information is indicated via —.

ID	Age	Sex	Device/ electrode	CI exp	# Active electrodes in clinical map	Scalar location	Mean electrode-to-modiolus distance	CNC 8 max, 16 max	AzBio +5 8 max, 16 max
1	70	M	CI24RE(CA)	121	22	ST	0.49	96, 90	93, 91
2	62	M	CI24RE(CA)	32	20 (1–2)	ST	0.20	50, 46	0, 39
3	62	M	CI24RE(CA)	39	20 (1–2)	ST	0.32	60, 76	20, 74
4	64	F	CI24RE(CA)	26	21 (1)	ST	0.69	66, 60	33, 10
5	77	M	CI24RE(CA)	50	22	ST	0.80	52, 62	51, 49
6	72	F	CI24RE(CA)	20	19 (1–3)	ST	0.48	58, 64	16, 28
7	24	F	CI24RE(CA)	122	22	ST	0.39	92, 96	48, 47
8	87	F	Profile CI532	10	20 (14–15)	ST	0.42	70, 74	2, 24
9	62	M	Profile CI532	21	22	ST	0.56	78, 72	27, 61
10	79	M	Profile CI532	21	19 (1–3)	ST	0.39	64, 76	34, 45
11	78	F	Profile CI532	19	22	ST	0.44	64, 70	25, 56
12	70	M	CI24RE(CA)	121	22	ST	0.51	70, 90	63, 91
13	74	M	CI24RE(CA)	19	22	ST	0.46	30, 32	0, 20
14	56	M	CI512	81	19 (1–3)	ST	0.50	60, 56	21, 23
15	72	F	Profile CI512	17	19 (1–3)	ST	0.41	70, 60	24, 63
16	84	F	CI24RE(CA)	36	22	ST	0.65	80, 84	53, 47
17	54	F	CI24RE(CA)	27	22	ST	0.40	80, 86	69, 76
18	78	M	Profile CI512	3	20 (1,2)	ST	0.36	60, 62	25, 21
19	43	F	CI24RCA	44	22	ST-SV	0.30	78, 72	33, 45
20	53	F	CI512	84	22	ST-SV	0.52	64, 74	62, 69
21	76	F	CI24RE(CA)	54	21 (1)	ST-SV	0.52	34, 56	4, 40
22	54	F	CI24RE(CA)	79	22	ST-SV	0.58	62, 78	28, 25
23	54	F	CI24RE(CA)	61	22	ST-SV	0.63	62, 70	0, 1
24	78	F	CI512	75	22	—	—	60, 61	15, 19
25	32	M	CI24RE(CA)	48	22	—	—	68, 74	30, 48
26	32	M	CI24RE(CA)	48	22	—	—	74, 82	41, 61
27	80	F	CI24RE(CA)	68	22	—	—	72, 61	17, 15
28	63	M	CI24RE(CA)	39	22	—	—	48, 54	26, 31
29	78	M	Profile CI512	3	19 (1–3)	—	—	76, 80	1, 4
30	71	F	CI24RCA	154	20 (1–2)	—	—	56, 80	33, 78
mean	64.6	N/A	N/A	51.4	21.1	N/A	0.49	64.0, 67.3	28.3, 41.9

TABLE II. Electrode deactivation methods and associated frequency allocations for all channel conditions. *In cases for which the participant had electrode(s) deactivated in their clinical map, we chose to activate the closest available electrode to maintain the greatest spatial separation between activated electrodes. For example, if E1 elicited a non-auditory percept, E2 would be activated instead for the 8, 16, and all on conditions.



was completed acutely. Each of the five CI programs was tested using a loudspeaker at 0-degrees azimuth and 1 m from the participant in a single walled sound booth using: Consonant Nucleus Consonant (CNC) 6 words, AzBio sentences in quiet, and +5 SNR using 20-talker babble noise, vowels (closed set), consonants (closed set), and spectral modulation detection (SMD) using the quick spectral modulation detection (QSMD) test (Gifford *et al.*, 2014). Subjective sound quality judgments were assessed using a visually presented 10-point scale (1 = very poor; 10 = very good), in which the participant rated the overall sound quality of the list of CNC words, AzBio sentences in quiet and +5 SNR for each condition. Vowel stimuli consisted of 13 vowels in /bVt/ format (“bait, Bart, bat, beet, Bert, bet, bit, bite, boat, boot, bought, bout, but”). Vowel formants were equal duration (90 ms) so that vowel length could not serve as a cue. Consonants were 16 male consonant tokens in the /aCa/ context. The QSMD task as used in this study employs a method of constant stimuli using a single modulation rate (1 cyc/oct) with 9 modulation depths (4 to 22 dB, in 2-dB steps) for which performance is expressed as the overall score, in percent correct, across all modulation depths.¹

Target stimuli were presented at a calibrated level of 60 dB sound pressure level (SPL).

C. Results

One-way repeated measures analysis of variance was completed with the number of channels as the independent variable and speech/auditory perception scores and sound quality ratings as the dependent variable. *Post hoc* analyses were completed with all-pairwise, multiple comparisons using a Holm–Sidak statistic. Figure 1 displays mean scores for speech recognition (panel A), sound quality judgments (panel B), as well as vowels, consonants, consonant features, and QSMD (panel C) for each of the channel conditions. In an attempt to minimize the influence of floor and ceiling effects, CNC and AzBio sentence recognition scores were converted from percent correct to rationalized arcsine units or RAU (Studebaker, 1985) prior to analysis.

D. CNC word recognition and sound quality

For CNC word recognition, there was a significant main effect of number of channels [$F_{(4,40)} = 34.69, p < 0.0001$,

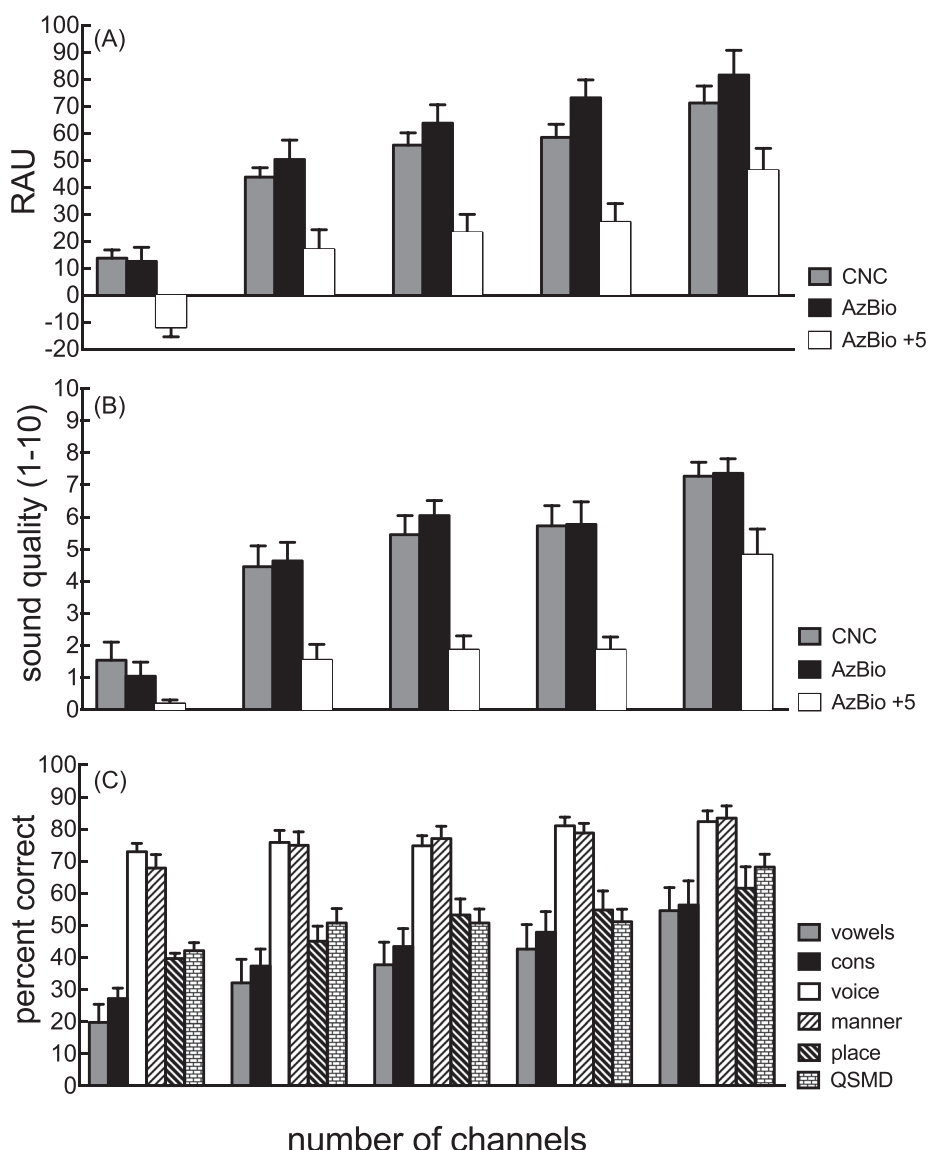


FIG. 1. Mean outcomes for 11 listeners with ST perimodiolar electrode arrays across all tested channel conditions for scores for CNC words, AzBio sentences in quiet, AzBio sentences at +5 dB SNR (Panel A), sound quality ratings for CNC words and AzBio sentences in quiet and noise (Panel B), and vowels, consonants, consonant features, and QSMD (Panel C). Error bars are +1 standard error measurement.

$\eta_p^2=0.78$]. *Post hoc* analyses revealed significant performance differences between four electrodes and all other electrode conditions ($p < 0.0001$ for all comparisons). Additionally, there was a significant difference between performance obtained with all on versus 8 channels ($t = 5.26$, $p < 0.0001$), all on versus 10 channels ($t = 3.00$, $p = 0.024$), and 16 versus 8 channels ($t = 2.82$, $p = 0.03$). For CNC sound quality, statistical analysis revealed a significant main effect of number of channels [$F_{(4,40)} = 17.29$, $p < 0.0001$, $\eta_p^2 = 0.63$]. *Post hoc* analyses revealed significant performance differences between four electrodes and all other electrode conditions ($p < 0.002$ for all comparisons). There were also significant qualitative differences between all on and 8 channels ($t = 2.82$, $p = 0.002$).

E. AzBio sentence recognition in quiet and sound quality

For AzBio sentence recognition in quiet, there was a significant effect of number of channels [$F_{(4,40)} = 31.68$, $p < 0.0001$, $\eta_p^2 = 0.76$]. *Post hoc* analyses revealed significant performance differences between 4 channels and all other electrode conditions ($p < 0.0001$ in all cases) as well as between all on and 8 channels ($t = 4.61$, $p < 0.001$), all on and 10 channels ($t = 2.62$, $p = 0.048$), and 16 versus 8 channels ($t = 3.37$, $p = 0.008$). For quiet AzBio sound quality, there was a significant effect of number of channels [$F_{(4,40)} = 22.43$, $p < 0.0001$, $\eta_p^2 = 0.69$]. *Post hoc* analyses revealed significant performance differences between 4 channels and all other electrode conditions ($p < 0.0001$ in all cases). Additionally, the all on condition was found to yield significantly higher qualitative judgments than 8 channels ($t = 3.81$, $p = 0.003$).

F. AzBio sentence recognition in +5 dB SNR and sound quality

For AzBio sentence recognition at +5 dB SNR, there was a significant effect of number of channels [$F_{(4,40)} = 19.32$, $p < 0.0001$, $\eta_p^2 = 0.66$]. *Post hoc* analyses revealed significant performance differences between 4 versus 10 channels ($t = 3.09$, $p = 0.018$), 4 versus 16 channels ($t = 3.09$, $p = 0.022$), and 4 channels versus all on ($t = 8.51$, $p < 0.0001$). Further, the all on condition yielded significantly higher scores than 8 channels ($t = 6.01$, $p < 0.001$), 10 channels ($t = 5.42$, $p < 0.001$), and 16 channels ($t = 5.42$, $p < 0.001$). For sound quality judgments of AzBio sentences at +5 dB, there was a significant effect of channels [$F_{(4,40)} = 21.64$, $p < 0.0001$, $\eta_p^2 = 0.68$]. *Post hoc* analyses revealed significant qualitative differences between 4 channels versus all other conditions ($p < 0.0001$), as well as all on versus 8 ($t = 4.58$, $p < 0.001$), all on versus 10 ($t = 3.61$, $p = 0.004$), and all on versus 16 ($t = 3.00$, $p = 0.018$).

G. Vowel recognition

For vowels, there was a significant effect of channels [$F_{(4,40)} = 6.75$, $p = 0.0003$, $\eta_p^2 = 0.40$]. *Post hoc* analyses revealed significant performance differences between 4 versus 16 channels ($t = 3.26$, $p = 0.02$), 4 channels versus the all

on condition ($t = 4.98$, $p < 0.001$), and the all on condition versus 8 channels ($t = 3.21$, $p = 0.021$).

H. Consonant recognition

For consonant recognition, there was a significant effect of channels [$F_{(4,40)} = 14.35$, $p < 0.0001$, $\eta_p^2 = 0.59$]. *Post hoc* analyses revealed significant performance differences between 4 versus 10 channels ($t = 3.96$, $p = 0.002$), 4 versus 16 channels ($t = 5.04$, $p < 0.001$), 4 channels versus the all on condition ($t = 7.11$, $p < 0.001$), all on versus 8 channels ($t = 4.64$, $p < 0.001$), and all on versus 10 channels ($t = 3.16$, $p = 0.018$). For consonant place of stimulation, there was a significant effect of channels [$F_{(4,40)} = 9.63$, $p < 0.0001$, $\eta_p^2 = 0.49$]. *Post hoc* analyses revealed significant performance differences between 4 and 10 channels ($t = 3.46$, $p = 0.009$), 4 and 16 channels ($t = 3.86$, $p = 0.003$), 4 channels versus the all on condition ($t = 5.60$, $p < 0.001$), as well as the all on condition versus 8 channels ($t = 4.23$, $p = 0.001$). For consonant manner, there was a significant effect of the number of channels [$F_{(4,40)} = 4.53$, $p = 0.004$, $\eta_p^2 = 0.31$]. *Post hoc* analyses revealed significant performance differences between the all on condition and 4 channels for manner ($t = 4.10$, $p = 0.002$). For consonant voicing, there was a significant effect of the number of channels [$F_{(4,40)} = 4.27$, $p = 0.006$, $\eta_p^2 = 0.30$]. *Post hoc* analyses revealed significant performance differences between the all on condition and 4 channels for voicing ($t = 3.36$, $p = 0.017$).

I. QSMD

For spectral resolution via QSMD, there was a significant effect of channels [$F_{(4,40)} = 11.67$, $p < 0.0001$, $\eta_p^2 = 0.32$]. *Post hoc* analyses revealed significant performance differences between 4 channels and the all on condition ($t = 6.62$, $p < 0.0001$), 8 channels versus all on ($t = 4.42$, $p < 0.001$), 10 channels versus all on ($t = 4.42$, $p < 0.001$), and 16 channels versus the all on condition ($t = 4.33$, $p < 0.001$).

J. Discussion

CI recipients with perimodiolar electrodes completely within ST demonstrated significantly higher outcomes with 16 channels over 8 channels on CNC words and AzBio sentences in quiet, a result not found in previous studies. These CI recipients also demonstrated higher outcomes in the all on condition using n-of-m as compared to 16 channels using CIS for AzBio +5 sentence recognition and sound quality ratings as well as for spectral resolution via QSMD (1 cyc/oct). However, it is unknown if this improvement with the all on condition over 16-channel CIS is due to increasing the number of electrodes, n-of-m signal processing strategy, and/or different input frequency range between the conditions. For other measures, we found a significant difference between the all on condition and 10 channels for CNC word recognition, AzBio sentence recognition both in quiet and in noise, consonant recognition, and spectral resolution via QSMD. Further the all on condition yielded significantly

higher performance than 8 channels for CNC word recognition and sound quality, AzBio sentence recognition in quiet and sound quality, AzBio sentences in noise and sound quality, vowel and consonant recognition, consonant place, and spectral resolution via QSMD. Consistent with our first hypothesis, modern-day CI recipients with perimodiolar electrodes fully inserted in ST show significant increases in performance using up to 22 electrodes and 16 maxima with ACE compared to maps using 4–10 channels with CIS. This finding is in contrast to previous work, which found no further improvements beyond 10 channels using SPEAK (Fishman *et al.*, 1997; Friesen *et al.*, 2001) and CIS processing strategies (Shannon *et al.*, 2011). However, the current study's finding is largely consistent with the results of Croghan and colleagues (2017) who documented higher sentence recognition in noise with 22 vs 12 active electrodes using 8 maxima and ACE—though stimulation strategies were not consistent across the two studies.

We theorize that the lower electrode-to-modiolus distance afforded by perimodiolar electrodes affords continued performance gains beyond 8 channels as perimodiolar arrays can result in a lower charge required for upper stimulation levels (Davis *et al.*, 2016) and lower stimulation levels yields less channel interaction (Chatterjee and Shannon, 1998). Reduced channel interaction with perimodiolar electrodes in ST could allow better spectral resolution. This is evidenced by performance gains beyond 8 channels on tasks that are highly dependent upon peripheral spectral resolution, such as monosyllabic words, vowels, consonant place, and QSMD. In contrast, consonant voicing and manner—tasks for which high levels are possible solely via temporal processing (e.g., van Tasell *et al.*, 1987; Shannon *et al.*, 1995)—showed little effect of the number of channels with differences observed only between the all on condition and 4 channels.

Gains in performance beyond 8 channels in a CIS strategy could also be related to an increased number of maxima. Previous studies have not systematically varied the number of maxima to understand its effect on speech recognition. Increasing the number of spectral peaks chosen within each 1 ms timeframe could be providing additional usable spectral information that could help explain why gains in performance beyond 8 channels in the current study were greatest on tasks that are highly spectral dependent. Future studies should investigate whether further improvements in speech recognition and sound quality increase with channels using other manufacturers' devices and speech coding strategies. Due to software limitations, the current study investigated a combination of CIS in the 4- to 16-channel conditions and n-of-m in the all-on condition with 16 maxima.

There was potential bias for the all on condition as the number of active electrodes and frequency table were consistent with the users' everyday program whereas the other electrode conditions did not preserve the tonotopic representation of frequency consistent with the listeners' everyday programs (Fu and Shannon, 1999). However, what was not familiar was the fact that the all on program used 16 maxima which doubled the overall stimulation rate per frame as compared to the listeners' everyday maps using 8 maxima.

Further research is needed to investigate the impact of map familiarity with respect to overall stimulation rate as well as number of active electrodes and frequency allocation.

In the current study, the input frequency range was held constant across the different experimental conditions, as occurs with the clinical fitting system. However, as the number of electrodes was reduced, the channel specific frequency band was broader for each active electrode (Table II), potentially resulting in some degree of frequency mismatch (relative to patients' clinical frequency allocations). As such, performance decrements with reduced-channel maps may be related to the number of electrodes, increased bandwidths for each electrode, and/or potential frequency mismatch for active electrodes relative to clinical allocations. It is also possible that channel interaction and frequency mismatch may have been reduced in the current study as compared to Friesen *et al.* (2001), due to known ST placement of the electrode array and lower electrode-to-modiolus distances. For QSMD, listeners may have been attending to the spectral resolution within or across discrete spectral regions, rather than the full spectral bandwidth. In future studies, additional tasks such as spectral ripple discrimination may provide better estimates of functional spectral resolution. In summary, additional work is needed to control for a number of variables including overall stimulation rate (pulse integration), frequency allocation per channel, and strategy (e.g., CIS vs n-of-m) as well as the potential interactions between these variables.

III. EXPERIMENT II

The primary aim of experiment II was to examine the effect of increasing the number of maxima from the clinical software default of 8 to 16 in the all on condition, given that spectral resolution was found to be significantly higher in the all on condition as compared to conditions containing 4–16 channels in experiment I. We hypothesized that participants with a close electrode-to-modiolus distance would achieve higher speech recognition with 16 versus 8 maxima due to greater channel independence and availability of more stimulated electrodes each frame, which could afford better spectral resolution.

A. Study participants

Thirty postlingually deafened adult CI users (mean age = 64.6 yr, range 24 to 87) with perimodiolar electrodes were recruited for participation. The 11 participants from experiment I were also included in this sample as well as 7 perimodiolar recipients with confirmed ST electrode location, 5 perimodiolar with transcalar displacement (ST-SV), and 7 perimodiolar recipients with unknown electrode placement as these 7 participants did not have postoperative CT. Data for participants who completed both experiment I and II were collected during the same study visit. Implanted devices are shown in Table I. Inclusion criteria required at least 3 months of CI experience and at least 18 active electrodes in the clinical map.

B. Methods

All participants were tested with two maps: (1) all on (19–22 active electrodes) with 8 maxima (consistent with the everyday map), and (2) all on with 16 maxima. Electrode condition and assessment order were randomized using a Latin Square design. All testing was done acutely following the addition of a 16-maxima map, without making any changes to the patient’s upper or lower stimulation levels. Thus the number of active electrodes and frequency allocation tables were equivalent across the two programs. It is possible that by controlling these parameters, the 16-maxima program could have been perceived as louder than the 8-maxima program due to a higher overall rate; however, there were no subjective reports of loudness differences between conditions. CNC word recognition and AzBio sentences at +5 dB SNR were presented at 60 dB SPL in the sound field. All participants were using a 900-Hz rate with 25 μ s pulse phase duration, ASC, and ADRO with their everyday map which was kept constant for the study. Prior to statistical analysis, all scores were converted to RAUs.

C. Results

Figure 2 displays speech recognition scores, in percent correct, for CNC words in quiet and AzBio sentence recognition at +5 dB for the 16-maxima condition as a function of the 8-maxima condition. Circles represent participants with confirmed ST electrode locations, circles with cross represent participants with transcalar displacement (ST-SV), and diamonds represent participants with unknown scalar location. The dotted diagonal lines represent the lines of identity and the dashed lines represent the 95% confidence interval for each of the measures (Thornton and Raffin, 1978; Spahr et al., 2012). Paired *t*-tests were completed comparing speech recognition scores, in RAU, with 16 vs 8 maxima for CNC words as well as AzBio sentences in noise. For CNC words, there was a significant difference between scores obtained with 16 maxima (mean 69.5 RAU) as compared to 8 maxima (mean 64.8 RAU) ($t_{29}=2.79$, two-tailed $p=0.009$), though this difference was not statistically significant when completing the analysis for scores expressed in percent correct. For AzBio sentences at +5 dB, 16 maxima

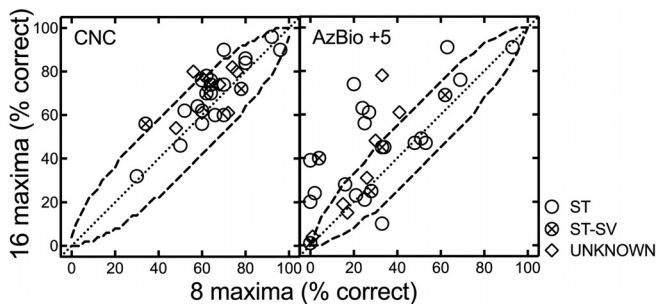


FIG. 2. Individual speech recognition for CNC monosyllabic words in quiet and AzBio sentences at +5 dB with 16 maxima as a function of 8 maxima. The dashed lines represent the 95% confidence interval for each of the measures. Circles represent participants with confirmed ST electrode location, circles with cross represent participants with transcalar displacement (ST-SV), and diamonds represent participants with unknown scalar location.

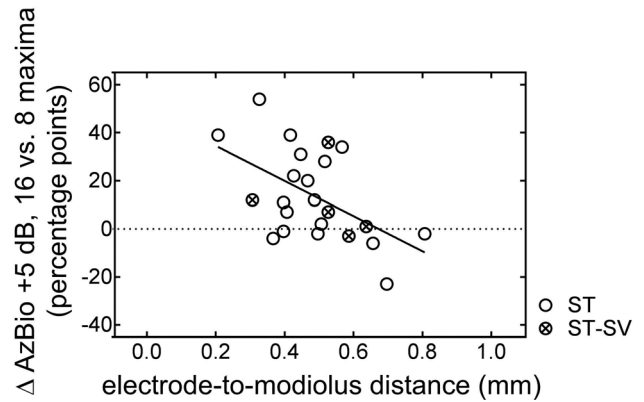


FIG. 3. Individual gain for AzBio sentence recognition at +5 dB SNR, in percentage points, with 16 versus 8 maxima is plotted as a function of mean electrode-to-modiolus distance, in mm.

(mean 42.6 RAU) yielded significantly higher outcomes than 8 maxima (mean 27.0 RAU) ($t_{29}=4.19$, two-tailed $p=0.0002$); unlike CNC word recognition, this difference remained statistically significant when analyzing raw scores in percent correct.

Figure 3 displays improvement with 16 versus 8 maxima, in percentage points, as a function of mean electrode-to-modiolus distance for the 23 participants for whom we had postoperative imaging data. We found a statistically significant correlation between degree of improvement on AzBio sentence recognition at +5 dB with 16 maxima and mean electrode-to-modiolus distance ($r=-0.52$, $p=0.0104$). Thus subjects with closer electrode-to-modiolus proximity demonstrated a greater benefit from increasing maxima for speech in noise. This correlation was not statistically significant for CNC word recognition ($r=-0.08$, $p=0.72$).

D. Discussion

Speech recognition was significantly better with 16 maxima as compared to 8 maxima for perimodiolar electrode recipients. Of this sample, only one listener exhibited a significant decrement in performance with 16 maxima—based on 95% confidence interval for test–retest variability—suggesting that clinicians should program perimodiolar recipients with 16 maxima to acutely assess performance benefit (particularly in noise). The largest gains in performance for speech in noise were those with the smallest electrode-to-modiolus distances, while those with the largest electrode-to-modiolus distances experienced no benefit or even a decrement in performance. Monosyllabic word recognition was also significantly higher with 16 maxima at the group level following RAU transform and no subjects exhibited a significant decrement in performance beyond that expected by test–retest variability. We hypothesize that this effect is likely due to perimodiolar CI users taking advantage of less channel interaction afforded by a closer electrode-to-modiolus distance allowing for a greater across-frequency resolution of the information transmitted across 16 versus 8 maxima. However, we cannot rule out that this effect could be due to a doubling of overall stimulation rate per frame which would

presumably provide better temporal representation of incoming stimuli (e.g., Loizou *et al.*, 2000b; Rubinstein and Hong, 2003). That is, increasing the maxima in an n-of-m strategy increases the overall stimulation rate (per frame) which stands to provide better representation of temporal envelope and spectrotemporal contrasts. Research is ongoing investigating whether CI recipients with greater electrode-to-modiolus distances—such as those with lateral wall electrodes—exhibit a similar benefit from more channels and/or greater maxima.

E. Summary

Current CI recipients with perimodiolar electrodes and using ACE (n-of-m) processing strategy achieve asymptotic performance with a higher number of channels than previously documented with older generation CI technology (e.g., Fishman *et al.*, 1997; Friesen *et al.*, 2001; Shannon *et al.*, 2011). The findings can be summarized as follows:

- The all on condition with 19 to 22 active electrodes and 16 maxima resulted in significantly higher outcomes than even 16 channels for AzBio +5 sentence recognition and sound quality ratings as well as for spectral resolution via QSMD.
- The all on condition with 19 to 22 active electrodes and 16 maxima resulted in significantly higher outcomes than even 10 channels for CNC word recognition, AzBio sentences in quiet and noise, consonant recognition, spectral resolution via QSMD, as well as subjective sound quality for AzBio sentences in noise.
- The all on condition yielded a significantly higher performance than 8 channels for CNC word recognition and sound quality, AzBio sentence recognition in quiet and sound quality, AzBio sentences in noise and sound quality, vowel and consonant recognition, consonant place, and spectral resolution via QSMD.
- 16-channel CIS yielded a significantly higher performance than 8-channel CIS for both CNC words and AzBio sentences in quiet.
- Increasing maxima from 8 to 16 in the all on condition resulted in a significantly higher monosyllabic word recognition and sentence recognition in noise.
- There was a significant negative correlation between electrode-to-modiolus distance and degree of benefit afforded by 16 maxima for sentences in noise.

This study did not control for local frequency mismatch or overall stimulation rate across experimental conditions; these variables or some interaction may have contributed to the present pattern of results. Additional research is needed to further investigate the effects of electrode scalar location, electrode-to-modiolus distance, overall stimulation rate, processing strategy (n-of-m versus CIS), and familiarization effects.

ACKNOWLEDGMENTS

The authors would like to thank Ashudee Kirk, MS; Anotnio Schefano, Shelia Lewis, and Linsey Sunderhaus, AuD for their contributions and support toward this project. Portions of this dataset were presented at the 2017

Conference on Implantable Auditory Prostheses, the 2018 American Auditory Society, and the 2018 Maximizing Performance in CI Recipients: Programming Concepts meetings. The project described is supported by Award Nos. R01 DC008408 (PI: Labadie), R01 DC009404 (PI: Gifford), DC014037 (PI: Noble), T35-DC008763 (PI: Hood), and R01 DC014462 (PI: Dawant) from the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Institutes of Health.

¹The QSMD version used here is commonly termed “acoustic QSMD” as it was designed for and intended for use with acoustic hearing ears. The reason is that the modulation rate and range of depths render the task more difficult than the typical “electric QSMD” as described in Gifford *et al.* (2014).

- Chatterjee, M., and Shannon, R. V. (1998). “Forward masked excitation patterns in multielectrode electrical stimulation,” *J. Acoust. Soc. Am.* **103**, 2565–2572.
- Croghan, N. B. H., Duran, S. I., and Smith, Z. (2017). “Re-examining the relationship between number of cochlear implant channels and maximal speech intelligibility,” *J. Acoust. Soc. Am.* **142**, EL537–EL543.
- Davis, T. J., Zhang, D., Gifford, R. H., Dawant, B. M., Labadie, R. F., and Noble, J. H. (2016). “Relationship between electrode-to-modiolus distance and current levels for adults with cochlear implants,” *Otol. Neurotol.* **37**(1), 31–37.
- Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z. (1998). “The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels,” *J. Acoust. Soc. Am.* **104**, 3583–3585.
- Fishman, K. E., Shannon, R. V., and Slattery, W. H. (1997). “Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor,” *J. Speech Lang. Hear. Res.* **40**, 1201–1215.
- Friesen, L. M., Shannon, R. V., 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 Shannon, R. V. (1999). “Effects of electrode location and spacing on phoneme recognition with the Nucleus-22 cochlear implant,” *Ear Hear.* **20**, 321–331.
- Gifford, R. H., Hedley-Williams, A., and Spahr, A. J. (2014). “Clinical assessment of spectral modulation detection for adult cochlear implant recipients: A non-language based measure of performance outcomes,” *Int. J. Audiol.* **53**, 159–164.
- Hughes, M. L., Stille, L. J., Baudhuin, J. L., and Goehring, J. L. (2013). “ECAP spread of excitation with virtual channels and physical electrodes,” *Hear. Res.* **306**, 93–103.
- Loizou, P. C., Dorman, M. F., Tu, Z., and Fitzke, J. (2000a). “Recognition of sentences in noise by normal-hearing listeners using simulations of speak-type cochlear implant signal processors,” *Ann. Otol. Rhinol. Laryngol. Suppl.* **185**, 67–68.
- Loizou, P. C., Poroy, O., and Dorman, M. (2000b). “The effect of parametric variations of cochlear implant processors on speech understanding,” *J. Acoust. Soc. Am.* **108**(2), 790–802.
- Noble, J. H., Gifford, R. H., Hedley-Williams, A., Sunderhaus, L., Labadie, R. F., and Dawant, B. M. (2014). “Clinical evaluation of an image-guided cochlear implant programming strategy,” *Audiol. Neurotol.* **19**, 400–411.
- Noble, J. H., Gifford, R. H., Labadie, R. F., and Dawant, B. M. (2012). “Statistical shape model segmentation and frequency mapping of cochlear implant stimulation targets in CT,” *Med. Imag. Comput. Asst. Intervention* **15**(Pt 2), 421–428.
- O’Connell, B. P., Cakir, A., Hunter, J. B., Francis, D. O., Noble, J. H., Labadie, R. F., Zuniga, G., Dawant, B. M., Rivas, A., and Wanna, G. B. (2016). “Electrode location and angular insertion depth are predictors of audiologic outcomes in cochlear implantation,” *Otol. Neurotol.* **37**(8), 1016–1023.
- Padilla, M., and Landsberger, D. M. (2016). “Reduction in spread of excitation from current focusing at multiple cochlear locations in cochlear implant users,” *Hear. Res.* **333**, 98–107.
- Rubinstein, J. T., and Hong, R. (2003). “Signal coding in cochlear implants: Exploiting the stochastic effects of electrical stimulation,” *Ann. Otol. Rhinol. Laryngol.* **191**, 14–19.

- Shannon, R. V., Cruz, R. J., and Galvin, J. J. (2011). "Effect of stimulation rate on cochlear implant users' phoneme, word and sentence recognition in quiet and in noise," *Audiol. Neurotol.* **16**, 113–123.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Spahr, A. J., Dorman, M. F., Litvak, L. L., Van Wie, S., Gifford, R. H., Loizou, P. C., Loiseau, L. M., Oakes, T., and Cook, S. (2012). "Development and validation of the AzBio sentence lists," *Ear Hear.* **33**, 112–117.
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech Hear. Res.* **28**(3), 455–462.
- Thornton, A. R., and Raffin, M. J. M. (1978). "Speech discrimination scores modeled as a binomial variable," *J. Speech Hear. Res.* **21**(3), 507–518.
- van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. (1987). "Speech waveform envelope cues for consonant recognition," *J. Acoust. Soc. Am.* **82**, 1152–1161.
- Wanna, G. B., Noble, J. H., Carlson, M. L., Gifford, R. H., Dietrich, M. S., Haynes, D. S., Dawant, B. M., and Labadie, R. F. (2014). "Impact of electrode design and surgical approach on scalar location and cochlear implant outcomes," *Laryngoscope* **124**(Sup 6), S1–S7.
- Wilson, B. S., Finley, C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (1991). "Better speech recognition with cochlear implants," *Nature Pub. Group* **352**(18 July 1991), 236–238.