Vowel identification by cochlear implant users: Contributions of duration cues and dynamic spectral cues

Gail S. Donaldson,[a] Catherine L. Rogers, Lindsay B. Johnson, and Soo Hee Oh
Department of Communication Sciences and Disorders, University of South Florida, PCD 1017,
4202 East Fowler Avenue, Tampa, Florida 33620, USA

A recent study from our laboratory assessed vowel identification in cochlear implant (CI) users, using full /dVd/ syllables and partial (center- and edges-only) syllables with duration cues neutralized [Donaldson, Rogers, Cardenas, Russell, and Hanna (2013). J. Acoust. Soc. Am. **134**, 3021–3028]. CI users' poorer performance for partial syllables as compared to full syllables, and for edges-only syllables as compared to center-only syllables, led to the hypotheses (1) that CI users may rely strongly on vowel duration cues; and (2) that CI users have more limited access to dynamic spectral cues than steady-state spectral cues. The present study tested those hypotheses. Ten CI users and ten young normal hearing (YNH) listeners heard full /dVd/ syllables and modified (center- and edges-only) syllables in which vowel duration cues were either preserved or eliminated. The presence of duration cues significantly improved vowel identification scores in four CI users, suggesting a strong reliance on duration cues. Duration effects were absent for the other CI users and the YNH listeners. On average, CI users and YNH listeners demonstrated similar performance for center-only stimuli and edges-only stimuli having the same total duration of vowel information. However, three CI users demonstrated significantly poorer performance for the edges-only stimuli, indicating apparent deficits of dynamic spectral processing.
© 2015 Acoustical Society of America. [http://dx.doi.org/10.1121/1.4922173]

[TCB]   Pages: 65–73

I. INTRODUCTION

When vowels are embedded in syllables, phonetic cues to vowel identity exist both in the center portion of the vowel and in the formant transitions that lead into and out of the vowel. The *vowel center* contains quasi-static spectral cues to vowel identity, including the target frequencies of the vowel formants ($F1$, $F2$, and $F3$) and slowly varying changes in those frequencies [i.e., vowel inherent spectral change (VISC)]. The *duration of the vowel center* provides an additional cue that helps listeners to disambiguate pairs of vowels with similar formant patterns. Finally, the *vowel edges*, or formant transitions, contain dynamic spectral cues in the form of rapid changes in formant frequency that reflect movements of the articulators toward and away from the vowels' nominal target positions, in the context of adjacent phonemes.

Normal hearing (NH) listeners are able to maintain high levels of vowel identification performance when listening to either center-only stimuli (excised vowels) or edges-only stimuli (so-called "silent-center" vowels), even when vowel duration cues are not available (Strange *et al.*, 1983; Jenkins *et al.*, 1983; Kirk *et al.*, 1992; Donaldson *et al.*, 2010). The ability to identify vowels when only a portion of the spectro-temporal cues are available likely helps such listeners to maintain high levels of speech perception in background noise, i.e., compared to listeners who receive a degraded signal due to hearing loss and/or cochlear implantation.

Relatively little is known about cochlear implant (CI) users' ability to make use of the dynamic spectral cues that exist in vowel edges as compared to the quasi steady-state spectral cues that exist in vowel centers. A study by Kirk *et al.* (1992) found that CI users with early-generation devices achieved better vowel recognition when listening to vowel centers, as compared to vowel edges, when duration cues were removed from both types of stimuli. Even when both vowel center and duration cues were available, the addition of vowel edges did not substantially improve subjects' scores. These findings suggested that the CI users had limited access to the dynamic spectral cues that exist in formant transitions. Nonetheless, a positive correlation was observed between vowel identification scores for the edges-only stimuli and scores on a word recognition test, suggesting that access to dynamic spectral cues supported better word recognition in some subjects.

Recently, we completed a study modeled after that of Kirk *et al.* (1992) in order to examine the relative contributions of vowel centers (quasi steady-state spectral cues) and vowel edges (dynamic spectral cues) to vowel identification by post-lingually deafened CI users with modern-day devices (Donaldson *et al.*, 2013). We undertook this study for two main reasons: First, because implant technology and sound processing strategies have evolved substantially since 1992, it was unclear whether the findings of Kirk *et al.* would generalize to contemporary CI users. Second, Kirk *et al.* tested the edges-only condition using an unusual stimulus type in which the initial and final edges of the syllable were abutted, rather than preserving a silent segment in place

[a] Electronic mail: gdonalds@usf.edu

of the vowel center. This use of abutted edges may have limited subjects' performance for the edges-only stimulus condition.

Stimuli in the Donaldson et al. (2013) study were naturally-spoken /dVd/ syllables ("deed, did, Dade, dead, dud, dad" and "Dodd") and modified versions of those syllables that preserved 80 ms of the excised vowel center (center-only stimuli), or deleted the vowel center, retaining only the edges of the syllables (edges-only stimuli). The initial and final segments of the edges-only stimuli were restricted to very brief durations (20 ms each) to ensure that steady-state cues were completely eliminated from all tokens. Because we were interested primarily in listeners' use of spectral cues, the silent centers of the edges-only stimuli were neutralized in duration. As a result, vowel duration cues were eliminated from both the center- and edges-only stimuli. Vowel-identification was tested using a seven-alternative forced choice (7AFC) procedure. As expected, considerable variability in performance was observed among the individual CI users. Most CI users could make use of steady-state spectral cues in the center-only condition (mean 42% correct), but scores for the center-only stimuli were significantly poorer than scores for the unmodified syllables (72%). Scores for the edges-only stimuli (29%) were lower than those for the center-only stimuli and many CI users performed at chance levels for the edges-only condition. Consistent with several previous reports (Strange et al., 1983; Jenkins et al., 1983; Donaldson et al., 2010), a comparison group of young normal hearing (YNH) listeners achieved near-perfect performance for the full and center-only stimuli and maintained relatively high levels of performance (70%) for the edges-only stimuli.

The present study was undertaken to address two issues raised by our earlier study. The first issue involved CI users' reliance on *vowel duration cues*, and the second involved CI users' access to *dynamic spectral cues*.

A. Vowel duration cues

In our previous study, an analysis of CI users' errors for the center-only condition suggested that at least some CI users relied on vowel duration cues to identify vowels in the full /dVd/ syllables. This finding further suggested that the absence of duration cues in the partial (center- and edges-only) conditions may have contributed to reduced performance for those conditions.

Two previous studies have addressed the influence of vowel duration cues on vowel identification by CI users. A study by Iverson et al. (2006) described findings from two relevant experiments. Their first experiment examined the contributions of formant movement during the center portion of the vowel (VISC), and vowel duration. Subjects were post-lingually deafened CI users and NH listeners. Subjects listened to /hVd/ syllables that (1) contained both formant movement and duration cues, (2) contained no formant movement cues, but retained their natural durations, (3) retained formant movement without duration cues, and (4) retained neither formant movement nor duration cues. The NH listeners heard the unprocessed stimuli for each condition, as well as the same stimuli passed through two-, four-, and eight-channel noise vocoders to simulate CI processing with a range of spectral resolution. Findings revealed that the best performance was always achieved with stimuli that contained both formant movement and duration cues, while the poorest performance was achieved with stimuli that contained neither formant movement nor duration cues. Vowel identification scores were significantly poorer when duration cues were removed; however, there was no clear evidence that subjects placed more perceptual weight on duration cues when there was poorer spectral resolution. That is, vowel duration cues appeared to be weighted similarly by CI and NH listeners and across the two-, four-, and eight-channel vocoder conditions.

A second experiment in the Iverson et al. (2006) study required subjects to select the parameters of formant movement and vowel duration that generated the best exemplars of each target vowel. Findings indicated that the vowel spaces generated by CI and NH subjects were similar, and supported the earlier finding that both groups used formant movement and duration cues to a similar extent. The authors attributed the lack of differences between groups to the fact that CI subjects in their study were post-lingually deafened, and therefore had some exposure to phonemes before they lost normal auditory input.

In contrast to the findings of Iverson et al. (2006), a report by Winn et al. (2012) suggested that CI users place more weight on vowel duration cues when spectral cues are degraded. Winn et al. used a cue-trading paradigm to evaluate listeners' weighting of phonetic cues when labeling synthetic vowels as either /i/ (in the syllable "heat") or /I/ (in the syllable "hit"). Subjects were CI users and NH listeners. As in the study by Iverson et al., the stimuli heard by the NH listeners were either unprocessed, or were spectrally degraded using a four- or eight-band noise vocoder. The synthetic stimuli varied in formant structure ($F1$ and $F2$ frequency), VISC, and vowel duration. Findings indicated that as spectral resolution was increasingly degraded, NH listeners relied less on the formant structure and VISC cues and more on the vowel duration cue. Actual CI users demonstrated a similar pattern of results as the spectrally degraded NH listeners; this finding suggested that CI users weighted duration cues more heavily than NH listeners (in the unprocessed condition) due to their impaired spectral resolution.

To further investigate the possibility that CI users rely more strongly on duration cues than NH listeners, the present study directly compared vowel identification performance for partial (center- or edges-only) /dVd/ syllables in which vowel duration cues were either neutralized or preserved. Of interest was whether CI users' performance for the partial syllable conditions would improve significantly when duration cues were present, compared to when they were absent, and whether the effect of restoring duration cues would be greater for CI users than for YNH listeners.

B. Dynamic spectral cues

The second issue raised by our previous study's outcomes concerned the ability of CI users to access the more

dynamic spectral cues in vowel edges. As noted earlier, CI users' average performance for the edges-only condition was quite low, even compared to performance for the center-only condition. However, because the durations of initial and final edges were limited to 20 ms, the total duration of acoustic information contained in the edges-only stimuli (40 ms) was only half that contained in the center-only stimuli (80 ms). Further, for some vowels, the 20-ms edges excluded portions of the formant transitions. Thus, it seemed possible that CI users' scores for the edges-only condition were limited by the brief durations of vowel edges retained in those stimuli and that performance in that condition underestimated the contribution of vowel edges to subjects' identification of vowels in full syllables.

In the present study, we compared performance for the original edges-only stimuli (20-ms edges) to performance for edges-only stimuli that retained all (or nearly all) of the initial and final formant transitions for all vowels (40-ms edges). Our goals were to determine (1) whether CI users would demonstrate better use of dynamic spectral cues when more complete representations of the vowel edges were provided; and (2) whether center- and edges-only stimuli would yield similar levels of performance when matched for total duration of acoustic information. We have shown previously that center- and edges-only /bVb/ stimuli that are matched in total duration of vowel information produce similar levels of vowel identification in NH listeners (Donaldson et al., 2010).

II. METHODS

A. Subjects

Ten CI users (13 ears) and 10 YNH listeners (22–27 yrs of age) served as subjects. All subjects were native speakers of American English. CI subjects were adults, 24–82 yrs of age, who had used unilateral or bilateral CIs for at least 1 year (see Table I). We have previously shown (Donaldson et al., 2010) that older NH listeners perform similarly to their younger counterparts when tested on a vowel identification task similar to the one used in the present study. In that study, younger and older subjects achieved similar performance for full syllables, as well as partial (center- and edges-only) stimuli similar to those used here; the only exception

TABLE I. Description of CI subjects.

Subject[a]	M/F	Age (yrs)	HL onset (yrs)[b]	Duration CI use (yrs)	Device/strategy
CI-2	F	59	30	4.0	Cochlear Freedom/ACE
CI-6 (bil)	F	67	51R; 43L	5.7R, 8.5L	AB Harmony/Fidelity120
CI-17	M	56	10	6.2	AB Harmony/Fidelity120
CI-44 (bil)	F	24	3	4.0R, 4.0L	AB Harmony/Fidelity120
CI-49	M	82	50	1.7	Med-El OPUS2/FSP
CI-50	M	82	80	1.0	AB Harmony/HiRes P
CI-51	F	59	21	4.3	Cochlear Freedom/ACE
CI-52 (bil)	F	27	birth	0.8R, 0.8L	Med-El OPUS2/FSP
CI-53	F	28	birth	5.5	AB Harmony/Fidelity120
CI-55	F	18	birth	5.4	AB Harmony/HiRes

[a]bil = bilateral CI user tested in each ear separately.
[b]Age at onset of bilateral severe-to-profound hearing loss.

occurred for center-only stimuli with durations (≤40 ms) substantially shorter than those used in the present study. Thus, it was not expected that age differences alone would generate differences in outcomes for the YNH and CI groups in the present study.

Three of the 10 CI users were prelingually deafened (CI-52, CI-53, and CI-55) and 6 were post-lingually deafened. The remaining CI subject (CI-44) was diagnosed with significant hearing loss at 3 years of age, but was able to benefit from bilateral hearing aids during her childhood years and demonstrated excellent speech and language abilities; thus, she was classified as being post-lingually deafened. The bilateral CI users (CI-6, CI-44, and CI-52) completed vowel identification testing for each ear separately. Demographic information for the CI users is provided in Table I.

Research procedures were approved by the University of South Florida Institutional Review Board and all subjects gave informed consent. Subjects were paid for their participation.

B. Stimuli

Stimuli consisted of /dVd/ syllables containing seven target vowels, as in deed, did, Dade, dead, dud, dad and Dodd. The unmodified ("Full") syllables were recorded from three female monolingual speakers of American English as described in Donaldson et al. (2013). Three tokens of each syllable were recorded per talker; one token was used for practice testing and the remaining two tokens were used in final testing. Six additional sets of stimuli were created from the Full stimuli using sound-editing software (Cool Edit Pro, 2000, Adobe, Inc., San Jose, CA) as described below. Two of these stimulus sets (Center-80 and Edges-20 conditions) were identical to those used in the earlier study. In all cases, a 10-ms linear ramp was applied to the edited edges of stimuli, to prevent acoustic transients.

1. Center-only, duration fixed (Center-80)

The initial and final formant transitions were deleted, preserving 80 ms of the vowel center. The vowel midpoint was located and portions of the vowel 40 ms preceding and following the vowel midpoint were preserved.

2. Center-only, duration preserved (Center-DP)

The initial and final formant transitions were deleted, preserving 50% of the vowel as measured from the release of the initial /d/ to the closure of the final /d/. Specifically, the vowel midpoint plus 25% of the vowel duration preceding and following the midpoint were retained. Because stimulus duration was proportional to the duration of the vowel in the original syllable, this condition preserved a cue related to vowel duration.

3. Edges-only (40 ms), duration fixed (Edges-40)

The vowel center was attenuated to silence, preserving 40 ms of the syllable following the vowel onset and 40 ms of the syllable preceding the final consonant closure. The silent

gap in the center of the syllable was increased or decreased in duration until the total duration of the vowel was equal to the average vowel duration of the Full syllables (259 ms). The duration of dynamic segments (40 ms) was selected to preserve the majority of dynamic information for all stimuli; for some stimuli, these segments may have included small portions of the quasi-static portions of the vowel centers.

4. Edges-only (40 ms), duration preserved (Edges-40DP)

These stimuli were identical to the Edges-40 syllables except that the original duration of the silent-gap in the center of the syllables was preserved.

5. Edges-only (20 ms), duration fixed (Edges-20)

The vowel center was attenuated to silence, preserving 20 ms of the syllable following the vowel onset and 20 ms of the syllable preceding the final consonant closure. The silent gap in the center of the syllable was increased or decreased in duration until the total duration of the vowel was equal to the average vowel duration of the Full syllables (259 ms). The duration of dynamic segments (20 ms) was selected to ensure that steady-state portions of the vowel were always excluded.

6. Edges-only (20 ms), duration preserved (Edges-20DP)

These syllables were identical to the Edges-20 syllables except that the original duration of the silent gap in the center of the syllable was preserved.

Figure 1 shows temporal waveforms for all seven stimulus conditions for each of two stimuli produced by the same female talker. The syllables "dead" and "dad" were selected for this illustration because they have contrasting vowel durations.

The 63 Full stimuli (unmodified syllables) were equated in root-mean-square amplitude based on measurements taken over the center portion of the vowel; this resulted in all syllables having approximately equal loudness. During vowel identification testing, the Full stimuli were presented at an average level of 66.5 dBA, with level-roving applied randomly over a 6 dB range (64–70 dBA) across presentations to eliminate the possibility that subjects could identify specific stimuli on the basis of residual loudness cues. The partial stimuli (Center-DP, Center-80, Edges-40, Edges-40DP, Edges-20, and Edges-20DP conditions) were presented using the same amplification and attenuation settings as used for the Full stimuli, including the random level rove. This procedure resulted in the center-only syllables (Center-80 and Center-DP conditions) having similar loudness as the Full stimuli, but caused the edges-only syllables (Edges-40, Edges-40DP, Edges-20, and Edges-20DP conditions) to be softer, due to the naturally lower intensity levels that occur at syllable edges.

C. Procedures

CI users were tested with their personal speech processors using the sound-mapping program that they would normally use in a quiet listening condition. While listening to sample Full stimuli, each subject adjusted the CI volume control to a setting that resulted in a loudness percept of "slightly loud but comfortable." Using the same volume setting, the subject then listened to sample Edges-20 stimuli and rated their loudness using an eight-point loudness scale. The loudness ratings obtained across listeners and ears were: "medium loud" ($n = 3$), "medium" ($n = 6$) and "medium soft" ($n = 5$). In all cases, the CI user indicated that the Edges-20 stimuli were easily audible.

Vowel identification testing was performed using a custom script written for the Eprime version 1.1 software (Psychology Software Tools, Inc., 2002, Sharpsburg, PA). Stimuli were played-out from a personal computer through a
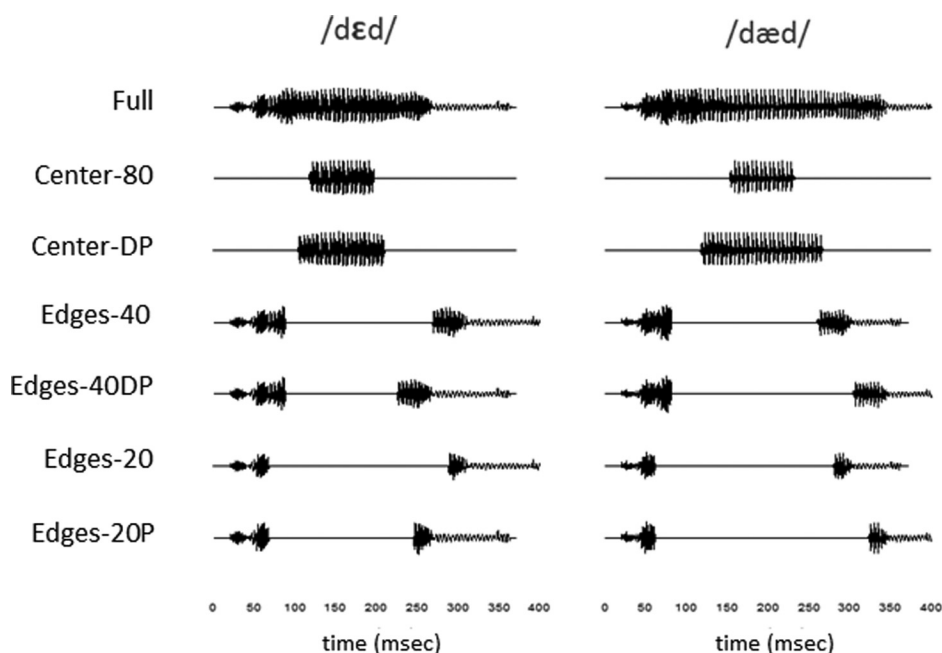

FIG. 1. Temporal waveforms demonstrating the seven stimulus conditions (indicated along the left side of the figure) for two /dVd/ syllables produced by the same talker.

Lynx I sound card (Lynx Studio Technology, Costa Mesa, CA), attenuated [Tucker Davis PA-5 attenuator (Tucker Davis Technologies, Alachua, FL) in passive mode] and routed to a high quality speaker (Spendor S3/5se, Spendor Audio Systems, Ltd., East Sussex, UK) inside a double-walled sound booth. A 7AFC paradigm, without feedback, was used for the vowel identification task. Stimuli were presented in blocks of 42 tokens (3 talkers × 7 vowels × 2 tokens per talker) with each block including the complete set of tokens for a single stimulus condition. The Full condition was presented first, followed by the remaining six conditions in random order. This process was repeated four times so that each subject completed four test blocks for each stimulus condition.

Percent-correct scores were converted into rationalized arcsine units (RAUs; Studebaker, 1985) for statistical analysis. A two-way (group × stimulus condition) repeated measures analysis of variance with one repeated factor (stimulus condition) was performed on the RAU-transformed scores to evaluate differences between groups and across conditions within groups. *Post hoc* comparisons (paired *t*-tests) were completed using the Holm-Sikak method. The significance (*p*) values shown below have been adjusted to account for multiple comparisons. Critical difference scores were used to compare individual subjects' performance between pairs of stimulus conditions.

III. RESULTS

Table II lists the mean data for both YNH listeners and CI users across stimulus conditions in both percent correct and RAU scores. YNH subjects demonstrated near-perfect levels of performance for both the Full and partial stimuli, with mean scores ranging from 99.6% for the Full syllables to 79.2% for the Edges-20 syllables. The CI users performed more poorly than the YNH listeners, overall, with scores ranging from 74.1% for the Full syllables to 34.0% for the Edges-20 syllables.

RAU scores from the right side of Table II are plotted in Fig. 2. RAU scores reduce the impact of ceiling effects on the YNH data, leading to a relatively similar pattern of performance across stimulus conditions for the two listener groups. For both groups, Full stimuli support the highest performance,
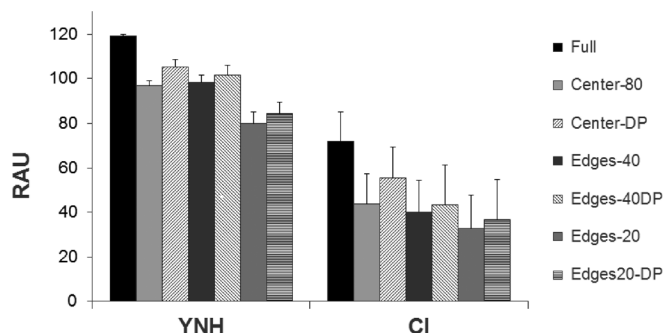

FIG. 2. Group mean data for YNH and CI subjects.

followed by a somewhat lower performance for the center conditions and 40-ms edges conditions, with the poorest performance observed for the 20-ms edges conditions. Within each duration pair (i.e., Center-80 vs Center DP; Edges-40 vs Edges-40DP; Edges-20 vs Edges-20DP) there is a consistent trend for performance to be higher when duration cues were preserved than when they were removed. Compared to the YNH listeners, the CI users demonstrated similar absolute decreases but larger percentage decreases for the partial syllable conditions relative to the Full condition. For example, mean performance decreased by 39 RAU in both groups from the Full to the Edges-20 condition; however, this change represented a 32% decrease in performance for the YNH listeners as compared to a 54% decrease for the CI users.

Individual YNH subjects tended to show a consistent pattern of performance across stimulus conditions; in contrast, individual CI users showed several different patterns of performance across conditions. Figure 3 shows the percent-correct scores of individual CI subjects, ordered according to their performance on the Full syllable condition (poorest performance on the left to best performance on the right). Identification scores ranged from 44.8% to 88.9% for the Full syllables, reflecting the broad range of speech recognition ability known to exist among the CI population. Unlike the YNH listeners, almost every CI user showed a substantial drop in performance from the Full to the partial syllable conditions. Exceptions to this pattern were observed for subject CI-51, whose performance levels for some of the partial-syllable conditions were nearly as good as the level she achieved for the Full condition, and subjects CI-55 and CI-6R, who achieved similar performance for the Center-80DP condition as for the Full condition. Several other CI users (CI-44L, CI-50, and CI-2) achieved relatively good scores (ranging from 43% to 70%) for the partial (center- and edges-only) conditions; however, their scores for these conditions were clearly depressed compared to their scores for the Full syllables. A few CI subjects, including two prelingually deafened subjects (CI-53 and CI-52), achieved scores for the edges-only conditions that fell near, or within, the range of chance performance (i.e., below the dashed horizontal line in Fig. 3).

A. Effect of preserving vowel duration cues

To address our first hypothesis, each group's mean performance was compared for pairs of stimuli that differed only in the presence or absence of vowel duration cues (i.e.,

TABLE II. Mean vowel identification performance across groups (YNH, CI) and listening conditions, expressed as percent-correct scores and RAU scores.

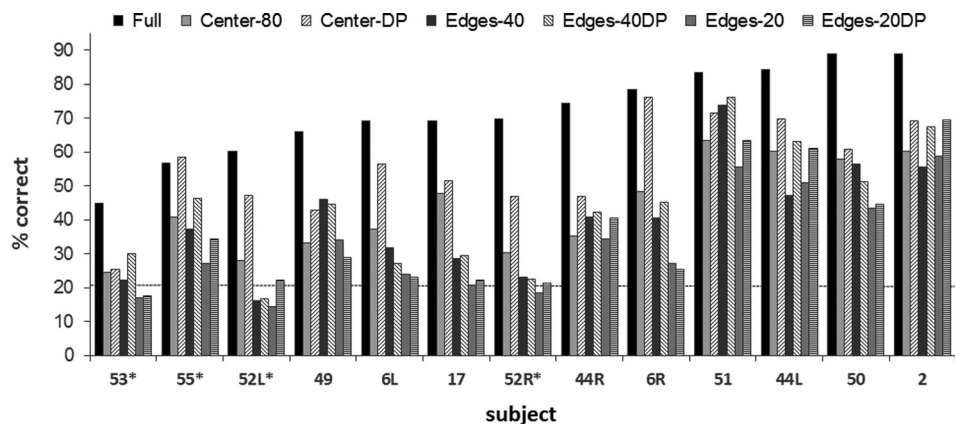|	% correct		RAU	
Condition	YNH mn (s.d.)[a]	CI mn (s.d.)	YNH mn (s.d.)	CI mn (s.d.)
Full	99.6 (0.6)	70.6 (13.7)	119.3 (4.8)	71.9 (13.9)
Center-80	92.3 (2.2)	41.6 (14.5)	97.0 (3.9)	43.9 (12.9)
Center-DP	95.6 (3.3)	54.3 (16.1)	105.2 (8.9)	55.4 (13.7)
Edges-40	92.0 (5.0)	38.0 (16.8)	98.4 (8.4)	40.2 (15.9)
Edges-40DP	94.0 (4.4)	39.9 (18.1)	101.7 (8.51)	43.3 (17.7)
Edges-20	79.2 (5.0)	31.8 (13.4)	80.0 (5.81)	32.8 (5.81)
Edges-20DP	82.6 (5.4)	35.2 (15.6)	84.3 (6.6)	36.6 (17.5)

[a]s.d. = standard deviation.

FIG. 3. Individual data for CI subjects. The dashed horizontal line indicates the upper limit of the 95% confidence interval for chance performance. Prelingually-deafened subjects are identified by asterisks next to the subject number.

Center-80 vs Center-DP; Edges-40 vs Edges-40DP; and Edges-20 vs Edges-20DP) (refer to Fig. 2). For the center-only comparison (Center-80 vs Center-DP), CI users demonstrated a significant effect of duration cues (paired t-test, $p < 0.021$) but YNH listeners did not. The CI users' mean scores increased from 45.3% in the Center-80 condition to 58.1% in the Center-DP condition (12.8 percentage points; 28.2% improvement) when duration cues were restored. Somewhat surprisingly, neither group of subjects demonstrated a significant benefit of duration cues for the edges-only conditions (Edges-40 vs Edges-40DP, or Edges-20 vs Edges-20DP).

Individual YNH subjects showed a consistent pattern of results, with scores being slightly higher for the duration-preserved stimuli compared to the corresponding duration-removed stimuli (individual data not shown). With one exception, however, none of the individual YNH listeners showed significant duration effects for either the center- or edges-only stimuli.

Individual CI users were more variable in their patterns of results. The mean CI users' data showed a significant benefit of duration for the center-only syllables, but, only 3 of 10 individual CI subjects (5 of 13 ears: C-6L, CI-6R, CI-52L, CI-52R, CI-55) demonstrated a significant benefit of duration for the center-only stimuli (i.e., scores that differed by more than the critical difference). Notably, however, each of these individuals (ears) demonstrated a large effect size (Cohen's $d > 3.9$). In addition, one CI user (CI-44L) showed a significant effect of duration cues with a large effect size (Cohen's $d = 3.7$) for the Edges-40 stimuli (Edges-40 vs Edges-40DP comparison) even though the group data showed no significant effect. Consistent with the group data, none of the CI users showed a significant benefit of duration cues for the Edges-20 stimuli (Edges-20 vs Edges-20DP comparison).

One possible reason that some CI users were unable to benefit from duration cues in the edges-only syllables is that poor spectral resolution limited their ability to identify the correct region of vowel space. Thus, even if a listener attended to the vowel duration cues provided in duration-preserved stimuli, those cues may have had little impact on performance. To assess this possibility, we examined error matrices for the Edges-40, Edges-40DP, Edges-20, and Edges-20DP stimulus conditions (combined across CI subjects and ears) to determine whether errors related to vowel duration were reduced when duration cues were preserved. Target vowels were categorized as having intrinsic durations that were short ("did, dead, dud"), medium ("deed") or long ("Dade, Dodd, dad") based on measurements of vowel duration for the 21 Full syllables. Table III shows the distribution of correct and incorrect responses for each stimulus condition, with errors divided according to whether the incorrect response had the *same intrinsic duration*, or a *different intrinsic duration*, as the stimulus. It can be seen that the distribution of CI users' errors did not change according to whether duration cues were retained or neutralized. This finding suggests that the CI users were unable to make use of vowel-duration cues in the edges-only listening conditions, perhaps because duration cues were coded in the silent gaps of the stimuli, rather than in the duration of the entire vowel (edges plus center). Thus, our speculation that these listeners selected vowels in the correct vowel-duration category, but incorrect region of vowel space was not supported.

Table III also includes an analysis of CI users' errors in the Center-80 and Center-DP conditions. This analysis confirms that listeners' improved performance in the Center-DP condition, relative to the C-80 condition, reflects their ability to make use of vowel duration information. It also rules out the possibility that average differences in vowel duration (i.e., Center-DP stimuli having longer average durations than Center-80 stimuli) can explain this finding.

B. Effect of lengthening vowel edges

To assess the effects of lengthening vowel edges from 20 to 40 ms, performance was compared for the Edges-20 and

TABLE III. Distribution of CI subjects' responses for the edges-only stimulus conditions.

Condition	Correct responses (%)	Error responses	
		Correct duration (%)	Incorrect duration (%)
Edges-40	38.5	19.3	42.2
Edges-40DP	41.1	17.7	41.2
Edges-20	30.2	22.1	47.8
Edges-20DP	33.1	19.4	47.5
Center-80	43.0	13.8	43.2
Center-DP	55.2	17.4	27.4

Edges-40 conditions. The YNH listeners achieved significantly higher scores for the Edges-40 condition than for the Edges-20 condition (paired t-test, $p < 0.021$), indicating that longer edges supported higher levels of performance. Their mean scores increased from 79.3% for the Edges-20 condition to 92.0% for the Edges-40 condition (12.7 percentage points; 16.0% improvement). CI users' scores did not differ significantly for the Edges-20 and Edges-40 conditions, although their mean scores increased from 34.0% for the Edges-20 condition to 41.4% for the Edges-40 condition (7.4 percentage points; 21.8% improvement). Individual YNH listeners showed a consistent pattern of results with nine of ten subjects having scores for the Edges-40 condition that were more than one critical difference higher than scores for the Edges-20 condition (individual data not shown). In contrast to the YNH listeners, the CI users exhibited variable effects of edge length. Four of the 10 CI subjects (4 of 13 ears: CI-49, CI-6R, CI-51, and CI-50) demonstrated significantly higher performance for the Edges-40 stimuli than for the Edges-20 stimuli, even though the group differences failed to reach statistical significance.

C. Comparison of performance for center-only and edges-only syllables

Based on a comparison of CI users' data for the Center-80 and Edges-20 syllables in our previous study, we speculated that CI users were less able to use dynamic vowel information present in the edges of the vowel than steady-state information present in the vowel centers (Donaldson *et al.*, 2013). It was of interest to revisit this issue in light of the present data, by comparing CI users' performance for center-only (Center-80) and edges-only (Edges-40) stimuli, which contained the same total duration of vowel information.

Figure 4 shows individual CI users' performance for the Center-80 and Edges-40 conditions, as well as the Full condition. Mean performance for the Center-80 stimuli (45.3%) was not significantly different than for the Edges-40 stimuli (41.4%); differences in individual scores for these conditions were significant in 4 subjects (4 of 13 ears), however. This finding suggests that the Center-80 and Edges-40 stimuli, which provide equal total durations of (non-silent) vowel information, support similar levels of vowel identification performance in most CI users. As discussed below, this result leads us to revise the tentative conclusion reached in our earlier study (Donaldson *et al.*, 2013), i.e., that dynamic spectral cues are inherently more difficult for CI users to process than static spectral cues. Instead, the data in Fig. 4 suggest that the majority of CI users are able to process dynamic spectral cues equally well as static spectral cues, but some CI users (i.e., CI-52L, CI-17, and CI-44L in Fig. 4) have relatively greater difficulty with dynamic cues, compared to steady state. It is possible that the reduced intensity level of the edges-only stimuli, relative to the center-only stimuli, contributed to poorer performance for the edges-only stimuli in some subjects. However, it seems unlikely that such intensity differences can explain this finding because only one of three CI users who demonstrated greater difficulty with

dynamic spectral cues (CI-52L) reported a strong reduction in loudness (i.e., more than one step on the loudness rating scale) for the Edges-20 stimuli as compared to the Full stimuli.

It is important to note that CI subjects' scores for both the Center-80 and Edges-40 syllables were substantially poorer, in most cases, than their scores for the Full syllables. This finding suggests that CI users have difficulty identifying vowels on the basis of either quasi-static or dynamic spectral cues alone, and contrasts with the ability of YNH listeners to successfully identify vowels on the basis of a single type of spectral cue (i.e., static or dynamic), as reflected by the mean data in Fig. 2.

IV. DISCUSSION

A. Vowel duration cues

One purpose of this study was to examine the hypothesis that CI users depend more strongly on duration cues than YNH listeners when identifying vowels in syllables. To this end, vowel-identification performance was measured in CI users and in a comparison group of YNH listeners, using /dVd/ stimuli that either preserved or eliminated duration cues. Overall, CI users varied in their reliance on vowel duration cues, with a subset of four CI users showing a strong reliance on duration cues (for either center- or edges-only stimuli), and the remainder showing a weaker reliance on these cues. Notably, however, all but two of the CI users showed duration effects that were larger than those for the average YNH listener.

Iverson *et al.* (2006) and Winn *et al.* (2012) examined the weighting of vowel duration cues by CI users and NH listeners. Findings from the study by Iverson *et al.* suggested that CI users weight vowel duration cues to a similar extent as NH listeners. In contrast, findings from Winn *et al.* suggested that CI users weight duration cues more heavily than NH listeners. Findings from the current study fall somewhere between the findings of these two previous studies, suggesting that some CI users rely more strongly than NH listeners on duration cues, while the remainder show a milder reliance on duration cues. Specifically, three of ten CI users exhibited a significant benefit of duration cues for center-only syllables and one additional CI subject showed a significant benefit of duration cues for the Edges-40 syllables. Our data also suggest that poorer-performing CI users may rely more heavily on duration cues than better-performing CI users; however, the sample size of the present study is too small to support any strong conclusions in this regard.

It should be noted that differences in phonetic context could have contributed to differences in findings across studies. In the present study, duration effects were examined using isolated vowel centers (i.e., Center-80 vs Center-DP conditions) and edges-only stimuli derived from seven /dVd/ syllables. Iverson *et al.* (2006) made use of naturally-spoken /hVd/ syllables that incorporated 13 target vowels, and Winn *et al.* (2012) used a synthesized continuum of /hit/-/hɪt/ stimuli. In addition, both Iverson *et al.* and Winn *et al.* reported only group findings, and not data for individual CI users. Thus, it is possible that some of the individual CI users

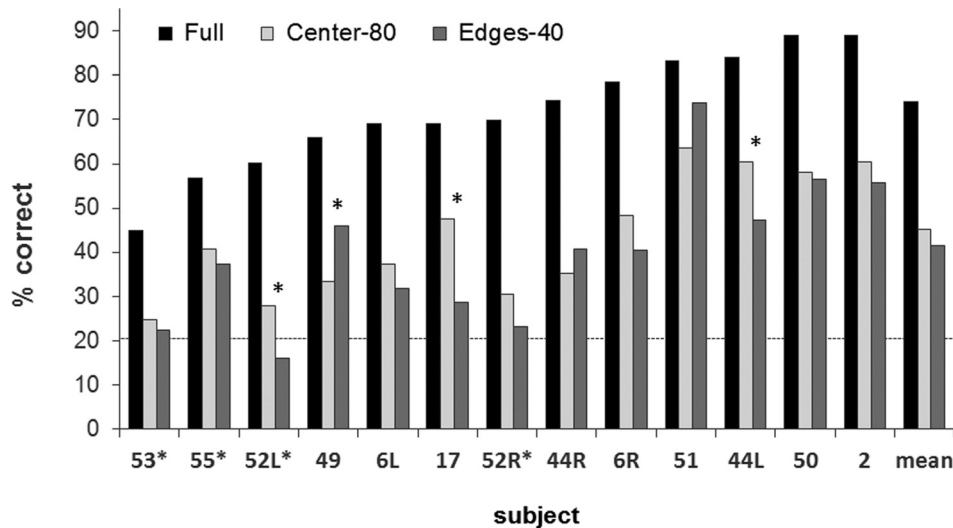J. Acoust. Soc. Am. **138** (1), July 2015

Donaldson *et al.* 71

FIG. 4. Individual CI users' data for the Full, Center-80, and Edges-40 syllables. Asterisks indicate significant differences between scores for the Center-80 and Edges-40 conditions.

in their studies demonstrated effects that were different than the average findings, similar to what was found for CI users in the present study.

There are two possible reasons why some CI users may rely more heavily than NH listeners on vowel duration cues. First, even though the CI gives listeners improved access to spectral cues, they are nonetheless likely to have reduced spectral resolution, compared to NH listeners, forcing them to depend on duration cues to distinguish vowels with similar formant structures (Winn *et al.*, 2012). Second, CI users may have learned to rely heavily on duration cues prior to implantation when spectral cues were less available, and failed to adjust their cue weights after receiving a CI. That is, some CI users may place greater perceptual weight on vowel duration cues, relative to spectral cues, because spectral cues were less available to them prior to implantation (i.e., due to the magnitude of hearing loss in the mid- to high-frequency regions). In the latter scenario, targeted auditory training could potentially facilitate an increased weighting of spectral cues, and concomitant decreased weighting of duration cues, to support better vowel perception. Such training could make use of duration-neutralized vowels (similar to our C80 syllables) as training stimuli.

It is not entirely clear why neither the YNH listeners nor the CI users showed a significant effect of duration cues for the edges-only stimuli. One possibility is that whatever benefit was conveyed by the duration cues in the edges-only syllables was negated by the silent center itself. That is, the temporal discontinuity in these stimuli may have disrupted listeners' ability to make use of duration differences in the silent centers. Our previous study (Donaldson *et al.*, 2013) provides some support for this notion. Specifically, that study included a stimulus condition (Gap20) in which Full stimuli were modified to contain a short temporal interruption by attenuating to silence a 20-ms segment of the vowel at its temporal center. YNH listeners achieved near perfect performance for both the Gap20 and Full conditions; thus, any detrimental effect of the 20-ms gap may have been obscured by a ceiling effect. Most CI users demonstrated a small decrement for the Gap20 condition compared to the Full condition (<5 percentage points); however, 4 of 11 CI

users showed larger decrements (>10 percentage points). This finding suggests that the presence of a temporal interruption *per se* may reduce performance in some CI users, even when minimal acoustic information is removed during the gap.

B. Dynamic spectral cues

The second purpose of this study was to determine whether performance for the edges-only conditions would improve if longer segments of the edges were retained. Findings showed that lengthening vowel edges from 20 to 40 ms improved the identification of edges-only stimuli by some CI users (4 of 10 subjects in the present study) but not for the group as a whole.

In our previous study (Donaldson *et al.*, 2013), CI users' performance was substantially poorer for brief, edges-only (Edges-20) syllables as compared to center-only (Center-80) syllables. This finding led us to speculate that dynamic spectral cues are inherently more difficult for CI users to perceive than static spectral cues. In the present study, Edges-40 syllables produced higher average performance than Edges-20 syllables, and scores for the Edges-40 syllables were similar to scores for the Center-80 syllables for most CI users. This finding leads us to revise our previous speculation that dynamic cues are inherently more difficult for CI users to process than static spectral cues. It suggests instead that dynamic spectral information is more difficult to access for some CI users, but not all.

Although the Edges-40 and Center-80 stimuli produced nearly equal levels of performance for the CI users, both types of partial syllables were identified much more poorly than the Full syllables. A less severe reduction was observed for the YNH listeners. CI users' relatively poorer ability to identify vowels on the basis of partial cues likely contributes to an increased difficulty understanding speech in the presence of background noise.

C. Clinical testing of vowel recognition

Clinically, vowel recognition is typically assessed using naturally-spoken /hVd/ syllables that include all available
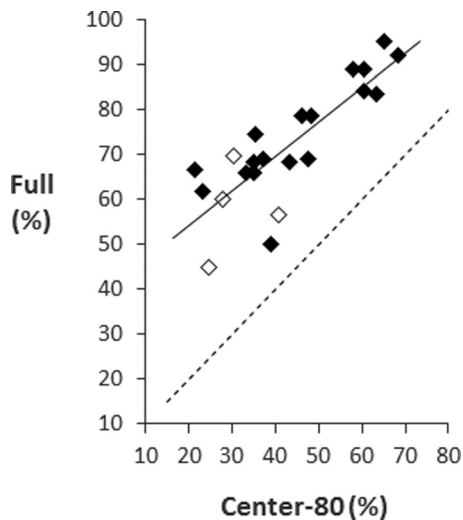
FIG. 5. Comparison of scores for Full and Center-80 syllables, for 22 ears from 18 subjects who participated in the present study or our earlier study (Donaldson et al., 2013). Only data from the present study were included for subjects who participated in both studies. Data points representing prelingually deafened subjects are indicated with open symbols.

cues to vowel identity, i.e., static spectral cues, dynamic spectral cues, and duration cues. Scores on this test are sometimes viewed as providing a rough index of CI users' spectral resolution for speech signals. However, findings from the present study, as well as our previous study (Donaldson et al., 2013), suggest that identification scores obtained with full syllables may overestimate listeners' ability to utilize static spectral cues in speech.

To examine the relationship between CI users' vowel identification performance for Full syllables (which provide all possible cues to vowel identity) and Center-80 syllables (which provide only static spectral cues), Fig. 5 plots both Full and Center-80 scores for the 18 CI users (22 ears) who were tested in the current study and/or the previous study (Donaldson et al., 2013). It can be seen that vowel identification performance for the Full syllables is generally proportionate to scores for the Center-80 syllables, except at low performance levels where there is some scatter in the data. The absolute difference between Center-80 scores and Full scores decreases as performance level increases, perhaps reflecting less reliance on vowel duration cues among the better performing subjects. Nonetheless, the Full syllables appear to provide a similar ranking of subjects' performance as the Center-80 syllables, suggesting that clinical tests of vowel recognition provide a reasonable index of individual CI users' static spectral acuity for speech, except perhaps among poorer performing listeners.

V. CONCLUSIONS

(1) Many CI users show duration effects that are similar to the average effects for YNH listeners; however, some CI users demonstrate a stronger reliance on duration cues, likely reflecting a perceptual strategy to compensate for reduced spectral acuity.

(2) Most CI users appear to have similar access to dynamic spectral cues and static spectral cues. For these individuals, the ability to make use of both types of spectral cues in speech is likely limited by the same factor; namely, reduced spectral resolution. However, some CI users (e.g., three of ten subjects in the present study) appear to have significantly more difficulty processing dynamic spectral cues as compared to static spectral cues.

(3) Further research is needed to confirm whether apparent deficits in dynamic spectral processing observed in some CI users are indicative of true deficits in dynamic processing or reflect a disruptive influence of the temporal interruptions in our edges-only stimuli. To this end, future studies in our laboratory will investigate CI users' processing of static and dynamic spectral cues in psychophysical tasks that make use of uninterrupted stimuli.

Donaldson, G., Rogers, C., Cardenas, E., Russell, B., and Hanna, N. H. (2013). "Vowel identification by cochlear implant users: Contributions of static and dynamic spectral cues," J. Acoust. Soc. Am. 134, 3021–3028.

Donaldson, G., Talmage, E., and Rogers, C. (2010). "Vowel identification by younger and older listeners: Relative effectiveness of vowel edges and vowel centers," J. Acoust. Soc. Am. 128, EL105–EL110.

Iverson, P., Smith, C., and Evans, B. (2006). "Vowel recognition via cochlear implants and noise vocoders: Effects of formant movement and duration," J. Acoust. Soc. Am. 120, 3998–4006.

Jenkins, J., Strange, W., and Edman, T. (1983). "Identification of vowels in 'vowelless' syllables," Percept. Psychophys. 34, 441–450.

Kirk, K., Tye-Murray, N., and Hurtig, R. (1992). "The use of static and dynamic vowel cues by multichannel cochlear implant users," J. Acoust. Soc. Am. 91, 3487–3497.

Strange, W., Jenkins, J., and Johnson, T. (1983). "Dynamic specification of coarticulated vowels," J. Acoust. Soc. Am. 74, 695–705.

Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," J. Speech Lang. Hear. Res. 28, 455–462.

Winn, M., Chatterjee, M., and Idsardi, W. (2012). "The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing," J. Acoust. Soc. Am. 131, 1465–1479.