Spring 4-2011

Vowel Identification by Cochlear Implant Users: Comparison of Vowel Edges and Vowel Centers

Nada Hanna
University of South Florida

Vowel Identification by Cochlear Implant Users:

Comparison of Vowel Edges and Vowel Centers

Nada Hanna

Honors College Thesis
April, 2011

Gail S. Donaldson, Ph.D., Director
Catherine L. Rogers, Ph.D., Reader

Department of Communication Sciences and Disorders
College of Behavioral and Community Science
University of South Florida

Introduction

Individuals diagnosed with severe or profound hearing loss often utilize a

cochlear implant (CI) as a means of treatment for their hearing impairment. The

prosthesis directly stimulates the auditory nerve by introducing electrical currents

into the cochlea; this triggers a sequence of neural events that results in sound

perception by the brain (Wilson, 2008).   A key limitation of cochlear implants is

that they support considerably poorer spectral resolution than the resolution that

exists for normal hearing (NH) listeners (Nelson et al., 2008).   This can lead to

reduced speech perception among CI users as compared to normal-hearing (NH)

listeners.

The present study focuses on individual differences in CI users' vowel

identification. When vowels are produced in isolation, the vocal tract maintains a

constant shape over the duration of the vowel, and the acoustic "signature" of each

vowel reflects the positions of the articulators.   Specifically, several bands of energy

("formants") are generated, with the frequency patterns of the formants varying

systematically for different vowels. A NH listener can discriminate isolated vowels

on the basis of the frequencies of the first, second and third formants (F1, F2 and F3,

respectively), aided by cues related to the duration of the vowel nucleus. Duration

cues help to differentiate pairs of vowels with relatively similar F1 and F2

frequencies.   CI users can also identify isolated vowels on the basis of formant-

frequency and duration cues; however, their use of F1 and F2 cues may be limited

by their poorer spectral resolution.

When vowels are produced in continuous speech, rapid changes in formant

frequency occur between consonant and vowel segments (formant transitions).

These formant transitions are brief in duration (typically 30-50 ms) but may be

particularly important to speech perception, since they provide cues to the identity

of both the vowel and consonant portions of consonant-vowel-consonant (CVC)

syllables.   As described below, NH listeners can make use of both formant

transitions (dynamic cues) as well as the steady state formant frequencies described

above (static cues) to identify vowels. It is less clear whether CI users can make use

of dynamic cues in addition to static cues.

A series of studies by Strange and her colleagues have evaluated the ability of

NH listeners to identify vowels in CVC syllables on the basis of their static

component (quasi steady-state formants in the vowel nucleus) and dynamic

components (formant transitions).   In an early, seminal study, Strange et al. (1983)

modified 10 naturally produced /bVb/ syllables to contain different combinations of

steady state, dynamic and vowel duration cues. The vowels tested were /i, ɪ, e, ə, ɛ,

ɑ, æ, o, u, ʊ/.   A single adult male produced two repetitions of each syllable, and the

syllable was then divided into three segments (dynamic-static-dynamic). Seven

modified syllables were then created from each natural (control) syllable. Table 1

summarizes the eight stimulus conditions and the information contained in each

(i.e., duration cues, dynamic formant transition information and/or quasi-static

information from the vowel nucleus).

The subjects tested by Strange et al. were native speakers of American

English who reported no hearing loss. Nineteen subjects were assigned randomly to

Table 1. Summary of the 8 syllable conditions tested by Strange et al. (1983).

Condition	Description	Cues included
Control Syllables	No modifications	Duration, dynamic, static
Silent Center Syllables (SC)	Center attenuated to silence	Duration, dynamic
Variable Center Syllables (VC)	Initial and Final components attenuated to silence	Duration, static
Fixed Center Syllables (FC)	Trimmed to length of shortest stimulus	Static
Shortened Silent Center Syllables (SHSC)	Trimmed center silence to shortest original silent interval	Dynamic
Lengthened Silent Center Syllables (LOSC)	Longest original interval of silence used for each	Dynamic
Initials Only (I)	Only initial component	½ Dynamic
Finals Only (F)	Only final component	½ Dynamic

 each of the eight conditions. Excluding the control condition that produced nearly

perfect performance (1% errors), the SC syllables produced the fewest errors (6%).

Thus, although the SC stimuli were essentially "vowel-less," their identification was

not much worse than the unmodified syllables.   This led Strange et al. to conclude

that static cues are not essential for the identification of vowels in syllables.   In

contrast to the SC condition (which contained both initial and final formant

transitions), the Initials Only and Finals Only conditions yielded the highest

percentage of errors (>45% error).   Strange et al. interpreted this to mean that

both the initial and final portions of the "vowel gesture" are needed for high levels of

syllable identification.   Duration cues appeared to be less important, since

performance on the two duration-neutralized SC conditions (SHSC and LOSC) were similar to the silent center condition that included vowel duration cues (SC).

Strange et al. also included a multiple-talker experiment to evaluate the effect of talker variability. Findings showed the same pattern of results as the single-talker experiment; however, there were more errors produced across all conditions. Based on the results of this experiment and their initial, single-talker experiment, Strange et al. reached the general conclusion that dynamic spectral information contained in the initial and final transitional segments of the syllables is sufficient for NH listeners to maintain accurate vowel identification, even when the vowel nucleus is attenuated to silence.

Jenkins and Strange (1983) performed a similar experiment that supported findings in the Strange et al. (1983) study. Stimuli were the same /bVb/ syllables used by Strange et al; however, two new stimulus conditions replaced the SHSC and LOSC conditions.   These conditions were the Hiss Centered (HC) and Abutted Syllables (AS) conditions described in Table 2.

Table 2. Additional stimulus conditions used by Jenkins and Strange (1983).

Condition	Description	Information Contained
Hiss Centered Syllables (HC)	Filled SC with naturally produced speech noise /ʃ/	Duration, dynamic
Abutted Syllables (AS)	Adjoined initial and final, no center component	Dynamic

Jenkins and Strange's rationale for using the HC stimuli was that speech noise inserted into the silent portion of the SC syllables could give those syllables an

improved sense of continuity without providing any new information related to

vowel identity.   The speech noise was adjusted to an amplitude that was lower than

the vowel nucleus it replaced, but was still easily audible.   The AS stimuli were

intended to avoid the listener's use of durational information and dynamic spectral

information as an advantage by deleting the center component altogether.

However, a potential problem with these stimuli was that the initial and final

components were juxtaposed, reducing overall syllable duration to an unnaturally

short value.   Similar to the SC and HC stimuli, this condition contained information

from the formant transitions but excluded information about the target frequencies

that the formants reach in the vowel center.

Jenkins and Strange tested 15 to 20 native American English speakers for

each condition. Error patterns across conditions were similar to those described by

Strange et al. (1983). Performance for the HC syllables did not differ significantly

from performance for the SC syllables; thus, these stimuli were not considered to be

useful for future experiments. The AC syllables produced similarly poor results,

likely due to the unnaturally short overall duration of the vowels and the acoustic

discontinuity between the initial and final transitions.

Taken together, findings from the Strange et al. (1983) and Jenkins and

Strange (1983) studies indicate that information related to vowel identification is

present throughout the vowel gesture. In the case of CVC syllables, this information

may be found not only in the quasi-steady state portion of the vowel, which reflects

the target frequencies of vowel formants, but also in the initial and final formant

transitions.   This view of vowel identification differs from earlier theories of vowel

perception that emphasize listeners' perception of canonical targets in the vowel

nucleus.   An important implication of the "vowel gesture" view of vowel

identification is that different acoustic cues to vowel identification (i.e., dynamic,

static and duration cues) represent overlapping sources of information that can be

used to identify vowels under difficult listening conditions. This phenomenon may

at least partly explain why NH listeners are able to maintain relatively high levels of

speech recognition under difficult listening conditions, e.g., when speech occurs at

soft levels or in the presence of background noise.

Although many studies have shown that NH listeners have the ability to

identify vowels using only dynamic formant transitions and quasi-state formants,

there was, until recently, no information concerning the effects of aging on this

ability.   Donaldson et al. (2010) addressed the issue of aging effects by assessing

vowel identification in 12 young (18-28 yrs) and 15 older (56-78 yrs) NH listeners

using /bVb/ syllables that were modified to contain varied durations of static cues

(vowel centers) and dynamic cues (formant transitions).   The stimuli consisted of 6

syllables: "beeb, bib, babe, beb, bab, and bob" selected from a larger set of stimuli

recorded previously by Rogers and Lopez (2008). The original syllables (FULL)

were modified to create three new types of syllables: Gap20, silent-center (SC) and

center-only (CO). Gap20 syllables were created by attenuating a 20 ms segment of

the stimulus at its temporal center, to determine whether a discontinuity in the

acoustic stimulus (without substantial loss of information) would degrade

performance. The SC syllables were created by attenuating to silence varying

durations of the vowel centers, thereby maintaining 80, 60, 40, 30, 20, or 10 ms of

the initial and final formant transitions.   Finally, the CO syllables were created by

eliminating the initial and final formant transitions, preserving 80, 60, 40, or 20 ms

of the vowel centers. Given the above modifications, the Full syllables contained

durational, dynamic and static cues, the SC contained only static cues, and the CO

contained only dynamic cues.

As expected, the Full and Gap20 syllables provided near perfect scores for

both groups of subjects.   However, performance worsened with decreasing

duration for the SC and CO syllables.   Older NH listeners performed similarly to

younger listeners for most of the stimulus conditions; their performance was

slightly poorer for the shortest (20, 40 and 60 ms) CO stimuli. Both younger and

older listeners demonstrated similar performance for static and dynamic conditions

that included equal total durations of acoustic information (e.g., SC40 vs. CO80).

This was interpreted to indicate that NH listeners can extract similar amounts of

information related to vowel identity from the centers and edges of CVC syllables.

To date, only one study (Kirk et al., 1992) has evaluated CI users' ability to

make use of static and dynamic cues to vowel identification. This study used a

paradigm modeled after Jenkins et al. (1983); however, silent center syllables were

replaced by syllables in which the initial and final transitions were abutted. In

addition, /wVb/ stimuli were used in addition to the /bVb/ syllable context used in

earlier studies.   The nine stimulus conditions used by Kirk et al. are summarized in

Table 3.

Subjects in the Kirk et al. study consisted of 10 Nucleus 22 and 10 Ineraid CI

users and 12 NH listeners ranging from 24-60 years of age.   It is important to note

Table 3. Stimulus conditions used by Kirk et al (1983).

Condition	Description	Information contained
b-Vowel-b	Contained amplitude-scaled syllables	Duration, dynamic, static
w-Vowel-b	Contained amplitude-scaled syllables	Duration, dynamic, static
Excised Vowel	Formant transitions deleted and steady state vowel components presented alone. /bVb/ duration	Duration, static
b-Mean Duration Vowel-b	Equated all steady state duration	Static, dynamic
w-Mean Duration Vowel-b	Equated all steady state duration	Static, dynamic
Excised Mean Duration Vowel	Formant transitions deleted and steady state vowel components presented alone. Duration of b-MDV-b	Dynamic
Abutted-bb Transitions	Adjoined initial and final, no center component	Dynamic
Abutted-wb Transitions	Adjoined initial and final, no center component	Dynamic
Excised Longest Duration Vowel	Formant transitions deleted and steady state vowel components presented alone. Longest vowel duration	Static

that the Nucleus 22 and Ineraid devices are early-generation CIs that process speech

quite differently from current-day implants. The Nucleus 22 device incorporated

about 20 frequency channels, similar to contemporary devices, but used a feature-

encoding strategy (F0/F1/F2) that provided only a sparse representation of the stimulus. The Ineraid device was limited to four channels of analog information; thus, it provided limited spectral detail, and suffered from distortion due to direct current interaction in the cochlea (Wilson, 2004).

Overall, CI users in the Kirk et al. study demonstrated similar patterns of performance as NH listeners; however their average accuracy was roughly half that of the NH listeners. One key difference between groups was that the CI users were less able to make use of dynamic cues than the NH listeners. This suggests that dynamic cues were conveyed poorly by the speech processing strategies, or that other auditory deficits prevented the CI users from making use of these cues. Importantly, those CI users who made better use of dynamic cues in the vowel identification task also had higher CVC word recognition scores. This supports the idea that dynamic cues are critical to speech perception and that contemporary processing strategies that better encode dynamic spectral cues may support higher levels of speech perception in CI users.

Purpose of the present study. The ability to utilize both static and dynamic cues to vowel identify likely underlies NH listeners' ability to maintain high levels of speech recognition under difficult listening conditions, such as when speech occurs in background noise. Little is known about how well CI users with current generation devices can make use of static vs. dynamic cues to speech perception. A better understanding of their capabilities could lead to improved CI designs and therefore, improved speech recognition performance for future CI recipients. Accordingly, the purpose of the present study is to evaluate the ability of CI users

with current-generation implants to identify vowels on the basis of static and dynamic cues.

Methods

Participants

Participants were 8 cochlear implant (CI) users with current-generation devices and 12 young normal hearing (YNH) listeners. All participants were native, monolingual speakers of American English. Each YNH subject was recruited from the University of South Florida and was screened prior to testing to ensure that pure tone thresholds were no greater than 20 dB HL at octave frequencies from 500 Hz to 8000Hz. The CI users were recruited from the University of South Florida CI lab and CI centers of Tampa Bay and had at least one year of CI experience. All but one of the CI users were postlingually deafened; the remaining CI user was prelingually deafened but demonstrated good speech and language skills. Each participant completed testing in a single research session of three to four hours duration and received monetary compensation for their participation ($8 per hour for YNH subjects and $15 per hour for CI subjects). CI users who lived more than 30 minutes from the lab were also given a $20 travel stipend.

Stimuli and Procedures

Stimuli were /dVd/ syllables containing the seven target vowels /i, e, ɛ, ɪ, ɑ, æ, ʊ/. These syllables were naturally produced, recorded and modified to form 8 stimulus conditions as described in Table 4: **FULL** (naturally spoken syllables without modification), **Gap20**, **C80**, **C60**, **E40**, **E30**, **E40-NP3** and **E40-NM3**. The last

two conditions (E40-NP3 and E40-NM3) were assessed as part of a larger study and

will not be discussed here. Details of the other conditions are presented in Table 4.

Table 4. Summary of the 8 syllable conditions tested

Condition	Description	Cues included
FULL	Naturally-spoken syllables with no modifications	Duration, dynamic, static
Gap20	20 ms of vowel center attenuated to silence	Duration, dynamic, static
C80	80ms of Center component preserved. Initial and Final components attenuated to silence	Static
C60	60ms of Center component preserved. Initial and Final components attenuated to silence	Static
E40	Center component attenuated to silence. 40ms of each edge preserved	Dynamic
E30	Center component attenuated to silence. 30ms of each edge preserved	Dynamic
E40-NP3	Noise (3 dB SNR re: vowel amplitude) inserted in the silent gap of the E40 syllable	Dynamic
E40-NM3	Noise (– 3 dB SNR re: vowel amplitude) inserted in the silent gap of the E40 syllable	Dynamic

Three monolingual native speakers of American English produced the

stimuli. Stimuli were recorded in the phrase "Say a /dVd/ again," to ensure similar

phonemic transitions into and out of the targeted syllable. The use of three different

speakers served to prevent listener adaptation to specific characteristics of one

speaker. A total of 12 repetitions were recorded per speaker per stimulus, and three

tokens of each stimulus were chosen to be the FULL stimuli. The FULL stimuli were

roughly equated in loudness by matching the root-mean-square (RMS) intensity levels of the most intense 40 ms of the vowel.

The FULL syllables were then modified using signal-editing software (Adobe Audition) to create the remaining 7 stimulus conditions using a procedure similar to that described in Rogers and Lopez (2008). Gap20 was created by attenuating to silence a 20 ms segment of the vowel at the temporal center of the syllable. This condition was used to observe the effect of a brief interruption in the stimulus, without substantial loss of acoustic information. The two *center-only* conditions, C80 and C60, preserved 80 ms and 60 ms of the vowel center, respectively, and attenuated to silence the initial and final edges. These conditions contained only static vowel cues. The two *edges-only* conditions, E40 and E30, preserved 40 or 30 ms, respectively, of the initial formant transition (immediately following the vowel onset) and the final formant transition (immediately preceding the final consonant closure). These conditions contained dynamic cues only to vowel identity. The durations of acoustic information in the *edges-only* conditions were selected to match those of the *center-only* conditions. For example the total duration of acoustic information in the E40 condition (80ms) was the same as that in the C80 condition.

Stimuli were stored as digital files on a personal computer. During the experiment, they were played out from the Lynx 1 sound card, attenuated (Tucker Davis PA-5 attenuator in passive mode) and routed to a high-quality speaker (Spendor S3/5se) inside a double-wall sound room. Stimuli were presented at an average level of 70 dBA. Calibration was accomplished using a 1kHz calibration tone, and a Brüel & Kjær type 2250 sound level meter. Stimulus levels were roved

over a 6 dB range within each block to ensure that minimal variations in loudness across stimuli did not enhance vowel identity performance. Different randomizations of level roving were used in each stimulus block.

Vowel identification was assessed using a seven-alternative forced choice procedure in which the listener was presented with a stimulus and subsequently identified the stimulus heard by selecting one of seven corresponding words on a computer monitor. Stimuli were presented in sets, where each set included 1 test block of stimuli for each condition.   The FULL condition was always presented first and the order of the seven remaining conditions was randomized. Different randomizations were applied for each new data set and subject in order to decrease the chance that learning effects or fatigue would systematically influence outcomes. This also increased the likelihood that more difficult conditions would be interspersed with easier conditions. The first data set was dedicated to practice testing that allowed the listener to familiarize themselves with the task.   Each practice block contained 21 trials (3 speakers x 7 vowels x 1 token), and the listener was given correct-answer feedback after every response.   The remaining 3 data sets were used to obtain the actual test data. For these sets, no feedback was given following each response, but a percentage correct score was displayed at the end of each block. Each test block contained 42 trials (3 talkers x 7 vowels x 2 tokens).

Both stimulus presentation and tracking of subject responses were controlled by a custom script written for the EPrime version 1.1 software (Psychology Software Tools, Inc., 2002).

Results

   Individual data for YNH listeners.   Figure 1 shows the vowel identification

scores for each of the 12 YNH listeners across the six stimulus conditions. Overall,

the YNH subjects demonstrated a consistent pattern of performance. In general,

performance was near perfect for the Full and Gap20 conditions, decreased for the

center-only conditions (C80 and C60), and decreased further for the *edges-only*

conditions (E40 and E30). For some subjects, decreases in duration within a given

type of syllable (i.e., C80 vs. C60 or E40 vs. E30) resulted in decreased performance;
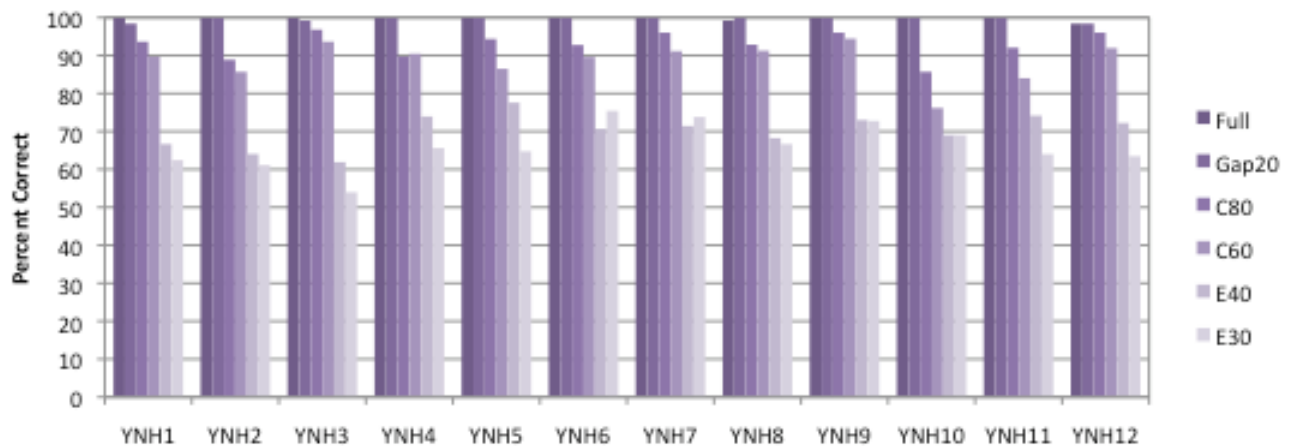
in others, duration had little or no effect.

Figure 1: Individualized data for 12 the YNH listeners

   Mean data for YNH listeners. Mean values and standard deviations for the

YNH data are shown in Figure 2 (left panel).   As observed for the individual data,

there is a slight reduction in performance (approximately 10%) from the

Full/Gap20 conditions to the *center-only* conditions, and an additional reduction in

performance (approximately 22%) from the *center-only* to the *edges-only*

conditions.   A one-way analysis of variance (ANOVA) showed a main effect of

condition (F=252.1, p<0.001). Post-hoc testing indicated that all comparisons were

statistically significant (Holm-Sidak, p < 0.001) except for the Full vs. Gap20 and
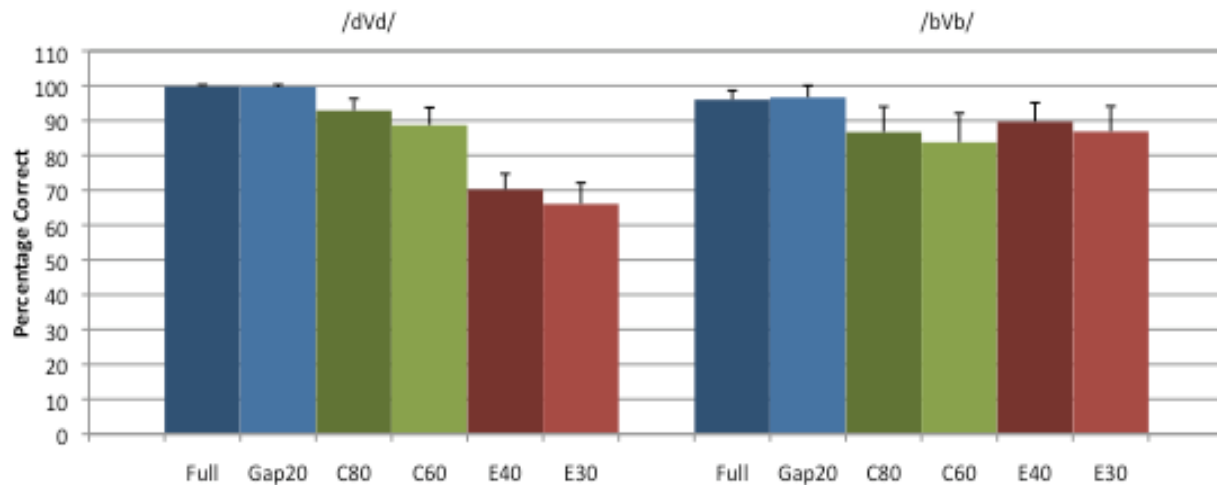
E40 vs. E30 conditions.

Figure 2: Comparison of means of /dVd/ vs /bVb/ studies

Comparison to /bVb/data. Figure 2 (right panel) shows comparable mean

data for /bVb/ syllables reported by Donaldson et al. (2010). Mean performance

levels for the Full, Gap20 and *center-only* conditions is similar to performance

shown for the /dVd/ syllables (left panel). However, performance for the *edges-only*

conditions is substantially higher for the /bVb/ stimuli than for the present /dVd/

stimuli.

Comparison of equal-duration conditions. Recall that we hypothesized

that *center-only* and *edges-only* conditions with equal total durations of acoustic

information would produce similar levels of performance. This was true for both

YNH and older NH subjects for the /bVb/ data reported by Donaldson et al.

However, it clearly was not the case for the /dVd/ data obtained here.   That is, comparison of vowel identification scores for the *center-only* and *edges-only* conditions with equal total durations (i.e., C80 vs. E40; C60 vs. E30) showed relatively poorer performance for the *edges-only* conditions.

Individual data for CI users. Figure 3 displays the individual data for the 8 CI users. Two overall patterns of performance are evident in these data: First, the four subjects whose data are plotted in the left half of the graph show patterns that resemble the patterns of performance of the NH listeners.   That is, they demonstrate the highest percent-correct scores for the Full and Gap20 conditions, somewhat lower scores for the C80 and C60 conditions, and still lower scores for the E40 and E30 conditions. In contrast, the four subjects whose data are shown in the right half of Figure 3 demonstrate a slightly different pattern of performance. For these subjects percent-correct scores are highest for the Full and Gap20 conditions, but demonstrate a substantial reduction for the *center-only* conditions, and then maintain this general level of performance for the *edges-only* conditions.
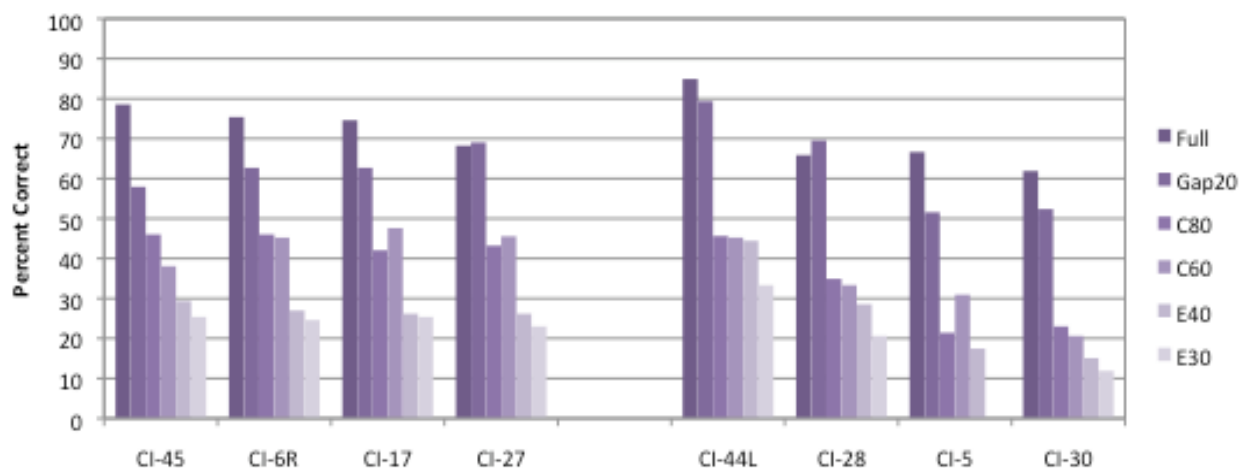

Figure 3: Individualized data for 8 CI users

Comparison of CI to YNH. Figure 4 compares mean data for the CI users

with that shown previously for the YNH listeners.   Overall, the CI users performed

more poorly than the YNH listeners, as expected.   For example, scores for the Full

condition were 99.8 % for the YNH group and only 73.5 % for the CI group.

However, the average CI data reflect the same pattern of results across conditions as

the YNH data. That is, the average CI data are more similar in pattern to the CI

subjects shown on the left side of Figure 3 rather than the right side.
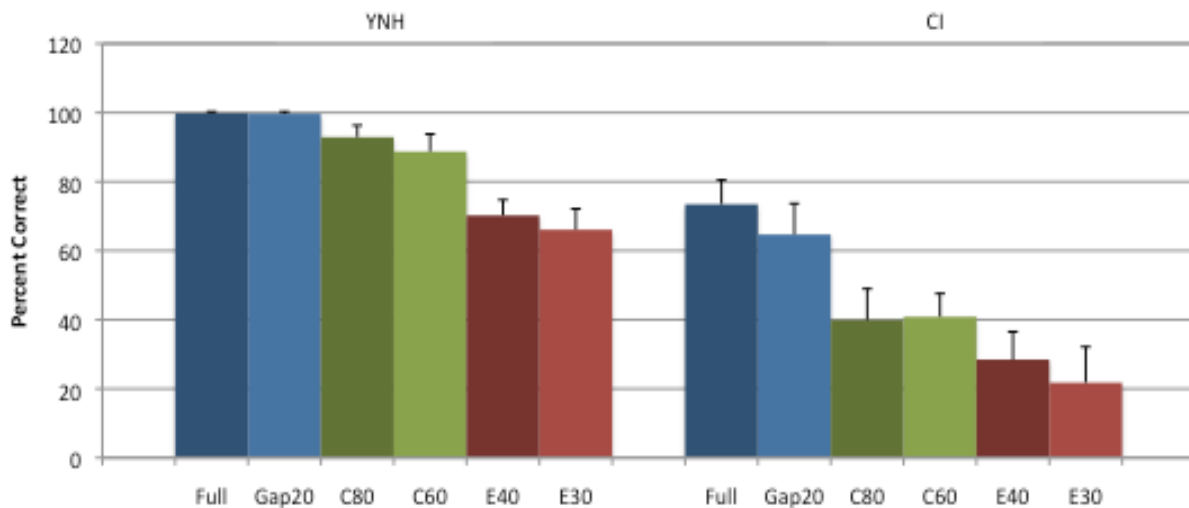

Figure 4: Comparison of Means CI vs YNH

A one-way ANOVA performed on the CI data confirmed a significant main

effect of condition (F=155.1, p<0.001). Post-hoc testing indicated that the Full,

Gap20, *center-only* and *edges-only* conditions yielded significantly different results

(Holm-Sidak, p < 0.001), but that scores were not significantly different within the

center-only (C80 vs. C60) or *edges-only* (E40 vs. E30) conditions.

Discussion

For the /dVd/ stimuli tested here, YNH listeners showed near-perfect performance on the FULL and Gap20 conditions, as expected from the earlier study of /bVb/ stimuli by Donaldson et al. (2010). However, unlike that study, their performance differed for the *center-only* and *edges-only* conditions having similar overall durations. Specifically, for the /dVd/ stimuli, performance for the *edges-only* condition was substantially poorer than performance for the *center-only* conditions. This may reflect differences in the articulators involved in closure for the /b/ vs. /d/ phonemes. Because /b/ closure depends on the lips, rather than the tongue, the tongue can move sooner during the /cV/ transition. As a result, information about the vowel target may occur during the nominal transition in the *edges-only* conditions. This is not possible for /d/ closures, because the tongue must make contact with the alveolar ridge. As a result, the *edges-only* conditions for the /bVb/ syllables may have contained portions of the vowel target itself, that was not present in corresponding conditions for the /dVd/ stimuli.

As expected, the CI users in our study showed poorer and more variable performance than the YNH listeners.   Where the YNH maintained relatively high levels of performance for the partial syllables (approximately 9% decrease for the *center-only* conditions and 31% decrease for the *edges-only* conditions) the CI users demonstrated substantially greater decrements for both of the partial-syllable conditions (30% and 45%, respectively).   This was particularly true of the four CI users whose data were shown in the right half of Figure 3. This suggests that the CI

users have a greater need to have all three types of cues (duration, static spectral
and dynamic spectral) present in the stimulus.

Another factor that may contribute to some of the results we observed is the
vowel-inherent spectral change (VISC) cue. Although vowel centers are relatively
static, slow changes in formant frequency occur during the center portion of the
vowel, allowing NH listeners to distinguish pairs of vowels whose with relatively
"close" F1 and F2 frequencies.   Use of this VISC information may account for the
relatively high performance of NH listeners on the *center-only* stimuli. It may also
explain why some CI users (those on the left side of Figure 3) were able to maintain
higher levels of performance for the *center-only* conditions than other CI users
(those on the right side of Figure 3).

Summary and conclusions

- YNH listeners can make use of either static cues or dynamic cues to identify
the vowels in /CVC/ syllables; however, vowel identification on the basis of dynamic
cues alone (*edges-only* stimuli) depends strongly on consonant context, and is
poorer for /dVd/ syllables than for /bVb/ syllables.

- CI users demonstrate poorer overall vowel-identification than YNH listeners
for /dVd/ syllalbes, and show larger decrements in performance than YNH listeners
when only static or dynamic cues are present in the syllables.

- CI users differ considerably in their ability to identify vowels in /dVd/
syllables on the basis of static cues alone (*center-only* stimuli). This may reflect

individual differences in the ability to make use of cues related to vowel-inherent

spectral change.

References:

Donaldson, G.S., Talmage, E.K., Rogers, C.L. (2010). Vowel identification by younger
      and older listeners: Relative effectiveness of vowel edges and vowel centers.
      The Journal of the *Acoustical Society of America,* 128 (3), 105-110.

Jenkins, J.J., Strange, W., Edman, T.R. (1983). Identification of vowels in "vowelless"
      syllables. *Perception & Psychophysics, 34(5),* 441-450.

Kirk, K I., Tye-Murray, N., Hurtig, R. R. (1992). The use of static and dynamic vowel
      cues by multichannel cochlear implant users. *The Journal of the Acoustical*
      Society of America, 91(6), 3487-3498.

Miller, G.A., Nicely, P.E. (1955). An analysis of perceptual confusions among some
      English consonants. *The Journal of the Acoustical Society of America*, 27, 338-
      352.

Nelson, D.A., Donaldson, G.S., Kreft, H.A. (2008). Forward-masked spatial tuning
      curves in cochlear implant users. *The Journal of the Acoustical Society of*
      America, 123, 1522-1543.

Strange, W., Jenkins, J.J., Johnson, T. L (1983). Dynamic Specification of coarticulated
      vowels. *The Journal of the Acoustical Society of America,* 74(3), 695-704.

Wilson, B. S., Dorman, M. F. (2008). Cochlear implants: Current designs and future
      possibilities. *Journal of Rehabilitation Research & Development,* 45(5), 695
      -730.