Audiovisual Sentence Recognition in
Bimodal and Bilateral Cochlear Implant Users

by

Shuai Wang

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2015 by the
Graduate Supervisory Committee:

Michael Dorman, Chair
Visar Berisha
Julie Liss

ARIZONA STATE UNIVERSITY

May 2015
The present study describes audiovisual sentence recognition in normal hearing listeners, bimodal cochlear implant (CI) listeners and bilateral CI listeners. This study explores a new set of sentences (the AzAV sentences) that were created to have equal auditory intelligibility and equal gain from visual information.

The aims of Experiment I were to (i) compare the lip reading difficulty of the AzAV sentences to that of other sentence materials, (ii) compare the speech-reading ability of CI listeners to that of normal-hearing listeners and (iii) assess the gain in speech understanding when listeners have both auditory and visual information from easy-to-lip-read and difficult-to-lip read sentences. In addition, the sentence lists were subjected to a multi-level text analysis to determine the factors that make sentences easy or difficult to speech read.

The results of Experiment I showed that (i) the AzAV sentences were relatively difficult to lip read, (ii) that CI listeners and normal-hearing listeners did not differ in lip reading ability and (iii) that sentences with low lip-reading intelligibility (10-15 % correct) provide about a 30 percentage point improvement in speech understanding when added to the acoustic stimulus, while sentences with high lip-reading intelligibility (30-60 % correct) provide about a 50 percentage point improvement in the same comparison.

The multi-level text analyses showed that the familiarity of phrases in the sentences was the primary driving factor that affects the lip reading difficulty.

The aim of Experiment II was to investigate the value, when visual information is present, of bimodal hearing and bilateral cochlear implants. The results of Experiment II showed that when visual information is present, low-frequency acoustic hearing can be of
value to speech understanding for patients fit with a single CI. However, when visual information was available no gain was seen from the provision of a second CI, i.e., bilateral CIs. As was the case in Experiment I, visual information provided about a 30 percentage point improvement in speech understanding.
ACKNOWLEDGMENTS

I still remember the first time I met my mentor, Prof. Dorman. I stopped by his office, knocked at the door, and said "hi" (with my strong accent).

He stopped typing, turned around gently and said, "Shuai? Come in, and have a seat."

Every imagination becomes real at that moment. All my previous thoughts on how he looks like, how the department looks like and how the life across the Pacific Ocean looks like became real at that moment. I had never been aboard before.

After five years, I still feel grateful to have Prof. Dorman as my mentor to explore the beauty of science. I learn both tangible and intangible knowledge from him. The latter part is so unique and especially important. I gain knowledge about how he thinks about scientific questions and solves problems in the world of science. That's something I can not get from reading papers or taking classes. I guess this is why mentorship is so important.

I would like to sincerely thank Prof. Dorman again for offering me enormous time, efforts and resources that I will benefit for the rest of my life. He always encourages me to try new ideas and thoughts with full support. I appreciate every single talk and discussion with him along these years. My dissertation is a summary of what I have learned during my five years of the PhD training.

I would like to thank Prof. Visar Berisha, who serves as my committee member. I admire him with his comprehensive knowledge and skills sets on engineering. He always knows the answer to solve the problems. I really enjoy working with him, both as a professor and as a friend.
I would like to thank Prof. Julie Liss, who always support her students with great personality. I can always feel her kindness and optimism whenever I talk with her. She is such a great lady.

I would like to thank my wife, Danni Wang (Yes, I am not only graduating soon but also got married!). Although sometimes she is even busier as a business doctoral student, she always supports me when I have problems. Before I met her I didn't know how miserable my life was. Now I couldn't imagine my life without her. She is not just my partner, but also my inspiration.

I would like to thank Ting Zhang, Louise Loiselle, Tony Spahr and Sarah Cook. I couldn't finish my dissertation without their help. As senior members in the lab, they make my life easier by generously sharing their experience with me.

I would like to thank Xuan Zhong and Myra Schatzki, who have been working with me in many great projects. Xuan's insight on applying knowledge into real world problems greatly inspires me. Myra is not only a great colleague but also a true friend of me.

I would also like to thank Shixiong Chen, Ningyuan Wang, Li Xu, Juan Huang, Yishan Jiao, Ming Tu, Yingchao Zhang for their help and advice on academics and personal life.

At last, I would like to thank my parents, Yang Yu and Futang Wang. As the only child in my family, I receive great patience and love from them. They unconditionally support me to study aboard. Thank you my parents, I miss you all so much now.

Shuai Wang, March 23rd, 2015
# TABLE OF CONTENTS

| LIST OF TABLES | vi |
| LIST OF FIGURES | vii |

## CHAPTER

1. EXPERIMENT I ................................................................. 1
   - Introduction .............................................................. 1
   - Methods ................................................................. 8
   - Results ................................................................. 11
   - Discussion ............................................................ 14

2. EXPERIMENT II ............................................................. 20
   - Introduction ............................................................ 20
   - Methods ................................................................. 24
   - Results ................................................................. 26
   - Discussion ............................................................ 31

3. CONCLUSION REMARKS .................................................... 42

REFERENCES ................................................................. 45
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Subject Demographic Information Such as Age at Testing, Year of Experience with a CI, Processor Manufacturer, and Etiology</td>
<td>9</td>
</tr>
<tr>
<td>2. Holm-Sidak’s Multiple Comparisons Test for 8 Bimodal Patients</td>
<td>36</td>
</tr>
<tr>
<td>3. Holm-Sidak’s Multiple Comparisons Test for 9 Bilateral Patients</td>
<td>39</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Survey of 121 CI Users on the Environments Which They Listened to............</td>
<td>2</td>
</tr>
<tr>
<td>2.</td>
<td>Lip Reading Scores of Kopra Sentences and AzAV Sentences on NH Subjects and Lip Reading Scores of AzAV Sentences in 19 CI Subjects ................</td>
<td>12</td>
</tr>
<tr>
<td>3.</td>
<td>Plot of Summary of Scores on A and A+V Kopra Hard List and Easy List on 10 CI Users and Plot of the A+V Score as a Function of A Score in Kopra Hard List and Kopra Easy List.................................................................</td>
<td>13</td>
</tr>
<tr>
<td>4.</td>
<td>Numeric Occurrence of Phrases from Google N-gram Project .......................</td>
<td>16</td>
</tr>
<tr>
<td>5.</td>
<td>Phoneme Counts in 4 Lists of Kopra Sentences........................................</td>
<td>18</td>
</tr>
<tr>
<td>6.</td>
<td>Unaided Audiograms of 9 Bimodal CI Subjects’ Non-implanted Ear in the Present Study.................................................................................................................................</td>
<td>19</td>
</tr>
<tr>
<td>7.</td>
<td>Plot of Scores of Six Test Conditions in 8 out of 9 Bimodal CI Users.........</td>
<td>27</td>
</tr>
<tr>
<td>8.</td>
<td>Plot of Scores of Six Test Conditions in 9 out of 10 Bilateral CI Users.......</td>
<td>29</td>
</tr>
<tr>
<td>9.</td>
<td>Plot of A+V score as a function of A-only score in bimodal and bilateral CI users.................................................................................................................................</td>
<td>30</td>
</tr>
<tr>
<td>10.</td>
<td>Summary Plot of Benefit from Adding HA When V is Absent and When V is Present .................................................................................................................................</td>
<td>37</td>
</tr>
<tr>
<td>11.</td>
<td>Individual Plot of Benefit from Adding HA When V is Absent and When V is Present .................................................................................................................................</td>
<td>38</td>
</tr>
<tr>
<td>12.</td>
<td>Summary Plot of Benefit from Adding a Second CI When V is Absent and When V is Present .................................................................................................................................</td>
<td>40</td>
</tr>
</tbody>
</table>
13. Individual Plot of Benefit from Adding a Second CI When $V$ is Absent and When $V$ is Present

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.</td>
<td>41</td>
</tr>
</tbody>
</table>
CHAPTER 1

EXPERIMENT I

Introduction

In the United States, the standard for assessing cochlear implant (CI) patient performance, both before and after CI surgery, is presentation of sentence material in an audio-only mode. Yet when Cook et al. (unpublished) surveyed cochlear implant users (n=121) on the environments in which they listened to speech, she found that the most common listening environment was one in which the CI patients could see the person they were talking to.

Some of the results from Cook et al. (unpublished) are shown in Figure 1. The question was “The most common sources of speech I hear come from ………?”. The subjects were allowed to choose one or more options from five categories: (1) radio, (2) television where I can see the talker, (3) television where I cannot see the talker, (4) another person who I can see and (5) telephone conversations. The percent responses for the five categories were as follows: (1) radio = 2%; (2) television where I can see the talker = 10%; (3) television where I cannot see the talker = 3%; (4) another person who I can see = 78%; and (5) telephone conversations = 12%

These data suggest that, if surgeons and audiologists are to provide a data based answer to the question, “How well will I do with my cochlear implant?” then testing in a combined audio-visual environment is necessary.

**Previous studies using AV test material with CI patients.** The benefits of adding visual information to auditory information in speech recognition for CI users have been observed in various reports.
Figure 1. Survey of 121 CI users on the environments that they listened to.
**Consonant level data.** Adding visual information helped consonant recognition in both pre-lingually deaf children CI recipients (Schorr et al., 2005; Tremblay et al., 2007; Tremblay et al., 2010) and post-lingually deaf adults (Rouger et al., 2008; Sterlnikov et al., 2009; Sheiffeld et al., 2014).

McGurk effect stimuli (McGurk, 1976) were also used in several experiments on CI users to examine the efficiency of audiovisual integration and visual benefit. A smaller than normal but observable McGurk effect was reported in both pre-lingually deaf children CI recipients (Schorr et al., 2005; Tremblay et al., 2007) and post-lingually deaf adult CI users (Dasei et al., 2008). Those experiments indicated that CI users were able to integrate both congruent and incongruent auditory and visual information to achieve better speech understanding, despite degraded information from CIs (Dasei et al., 2008; Tremblay et al., 2010; Rouger et al., 2008).

**Word level data.** Similarly, visual benefits were reported in word-level audiovisual speech recognition experiments with CI users. Rouger et al. (2007) reported an average 30% improvement in an audiovisual condition compared to an auditory-only condition in French word recognition in 91 CI users after implantation. Gray et al. (1995) reported about 36% visual gain on the Boothroyd word lists in 15 adult, Ineraid 4-channel CI users. Kaiser et al. (2003) reported about 35% visual gain on a lexically balanced word list in 20 post-lingually deaf adult CI users in quiet.

**Sentence level data.** Audio-visual version of Video Lists of Sentences (VLS) was used in the evaluation of performance of CI implantation in 37 post-lingually deaf CI users in Van Dijk et al. (1999). Twenty VLS sentences were presented in quiet in each of A, V, and A+V conditions. Van Dijk et al. (1999) reported the average speech reading
score was 28.7%, 33.4%, 39.9% in pre-operatively evaluation, evaluation on implantation after six months, evaluation on implantation after twelve months, respectively. Average A-only scores of VLS sentences were 0%, 20.2%, 24.7% in pre-operatively evaluation, evaluation on implantation after six months, evaluation on implantation after twelve months, respectively. Average A+V scores of VLS sentences were 28.7% (note this has be the same as V-only score), 70.4%, 76.9% in pre-operatively evaluation, evaluation on implantation after six months, evaluation on implantation after twelve months, respectively. The mean A+V gain at one year was 42.6 percentage points.

An audio-visual version of the City University of New York (CUNY) sentence recognition test (Boothroyd et al., 1988) was used in the Bergerson et al. (2003) study of 45 post-lingually deaf adult CI users. The CUNY sentences dataset was originally a speechreading sentence corpus containing topic-related sentences such as “food”, “family”, “work”, etc. (Boothroyd et al., 1988). The original version of the recordings was by a female talker from the Northeastern region of United States (Boothroyd et al., 1988). Boothroyd et al. (1988) reported average speech reading performance between 25% and 30% among 9 normal hearing adults. However, Altieri et al. (2011) reported a far different lip reading score for the CUNY sentences – for 84 normal hearing adults the score was only 12.4%. At all events, Bergerson et al. (2003) examined audio-alone, visual-alone and audio+visual conditions of the CUNY sentences in 45 CI users. The average speech reading performance of 45 CI users was around 15%; the average audio-only performance was near 50% and the average performance in the audio-visual condition was around 80%. Thus the mean A + V gain was approximately 30 percentage points.
Sentence level data from hearing-impaired listeners. Grant and Seitz (1998) used an audio-visual version of IEEE sentences in a speech recognition test for 30 out of 41 adult hearing-impaired listeners. The 41 candidates had mild to severe hearing loss with average (500Hz, 1000Hz and 2000Hz) pure-tone threshold at 39 dB HL. 30 subjects participated in A, V and A+V conditions using IEEE sentences at 0 dB SNR. Percent correct of 150 keywords was used as the measure of performance in each condition. Although Grant and Seitz (1998) did not report the average performance of subjects, visual inspection of Figure X from Grant and Seitz (1998) indicates that the average V-only score of subjects was near 10%, and the average benefit from adding vision was about 30%.

AV gain is related to speech-reading difficulty. Previous studies on audio-visual sentence perception on hearing impaired users have indicated that the difficulty of lip-reading is related to the amount of visual benefit when audio and visual information is combined. Bergeson et al. (2003) reported that adding vision provided about 30~35% improvements in performance for material that had approximately 15% intelligibility in a vision alone condition. Similarly, Grant and Seitz (1998) reported a lip reading score for IEEE sentences of about 10% correct and reported about a 30% point improvement in the A + V condition.

In contrast to the low speech reading intelligibility of CUNY and IEEE sentences, Van Dijk et al. (1999) reported a much higher speech reading performance -- near 40 percentage points -- on VLS sentences. Critically, the benefit from adding vision was also higher (above 40 percentage points) than in the Bergeson et al. (2003) and Grant and Seitz (1998) studies.
New sentence material for use in AV research – the AzAV sentence corpus. If, for example, changes in CI performance over time are to be assessed, then the difficulty of the test material must stay constant over time. This requirement motivated the development of the auditory-only AzBio sentence corpus for adults (Spahr et al., 2012) and for children (Spahr et al., 2014). The group that created these materials has recently created a new sentence corpus for use in A/V experiments. In the AzAV test corpus, the sentence lists have equal intelligibility in A-only conditions and equivalent gains from adding vision, i.e., in the A + V conditions (Cook et al., 2014).

The materials were inspired by, and copied from, an audio-visual test created by MacLeod and Summerfield (1990). The materials described by MacLeod and Summerfield (1990) were recorded in British English and had equal across-list auditory intelligibility (for 10 lists) and equal gain from visual information. To create the AzAV materials, the same sentences were rerecorded in American English (Cook et al., 2014). The AzAV corpus contains 10 lists of sentences. Each list contains 15 sentences. Each sentence has the same “Noun Phrase—Verb Phrase—Noun Phrase” structure. It is of some interest that list equivalence was maintained across very different speakers and dialects of English, i.e., a male speaker of British English and a female speaker of American English.

Experiment I - Assessing the speech reading difficulty of the AzAV test material and the gain in A+V intelligibility as a function of speech reading difficulty. In order to assess and scale the speech reading difficulty of sentences, Kopra et al. (1986) created sentences (the Kopra Sentences) that varied in lip reading difficulty. The Kopra Sentences consists of 12 lists of sentences that are ranked from the highest lip reading
index (0.995, equivalent to 99.5% correct) to lowest lip reading index (0.008, equivalent to 0.8%). Each list consists of 25 sentences that are used in “daily life” communication (Kopra et al. 1986). The sentences within each list were considered to have “equivalent” lip reading difficulty, and the average lip reading difficulty of sentences was considered significant “different” across the lists.

In the present study, materials from the Kopra sentences were recorded in AV formant. Lists 1, 4, 8, 12 spanned the continuum of easy-to-speech-read to difficult-to-speech-read material. Listeners were presented the Kopra sentence material as well as the AzAV sentence material. The outcomes are relevant to two issues: (1) the speech reading difficulty of the AzAV sentences and (2) the speech-reading ability of CI patients vs. normal-hearing listener. In addition, in this experiment the Kopra sentences were subjected to a multi-level text analysis to determine the factors that make sentences easy or difficult to speech read. In the last part of Experiment 1, the effect of easy-to-speech-read material vs. difficult-to-speech-read material on the gain in intelligibility when V was added to A was assessed.
Methods

**Subjects.** 16 young adult female normal hearing (NH) listeners volunteered in the present study. All subjects had normal or corrected-to-normal vision. 19 post-lingually deaf adult CI users (10 bilateral, 9 bimodal) also participated in the Experiment I (see demographic information in table 1). There are 9 males and 10 females. The average age at the time of study was 65 years old, ranging from 21 years to 87 years old. The manufacturers of CIs included Cochlear Corporation, MED EL Corporation and Advanced Bionics (see Table I for demographic information of subjects).

**Materials.** Lists 1 (easiest to lip read), 4, 8, and 12 (hardest to lip read) were selected from the Kopra Sentences. A young, female, English speaker who previously had recorded the AZAV sentences recorded video clips of these four lists. The recording conditions were the same as for the AzAV sentences. Two lists of the AzAV sentences, selected from the larger list of sentences, were also used for each listener.

**Procedure – Speech reading difficulty.** 16 NH listeners finished the lip reading tests of lists 1 and 12 of the Kopra Sentences and two lists of the AzAV sentences. 8 NH listeners finished speech-reading tests of Kopra lists 4 and 8. (This project grew over time and tests of lists 4 and 8 were added after the data from the other lists had been collected.) 19 adult CI subjects finished two lists.

Subjects sat in a sound treated booth with a loudspeaker and a 22”Dell video monitor at the distance of one foot. Subjects typed their answers with a keyboard. A trained experimenter later analyzed the transcripts. Average percent correct of words on 25 sentences was used as the score of that condition.

**Procedure – AV gain as function of speech reading difficulty.** 10 CI listeners
participated in this project. List 1 ("Easy List") and list 12 ("Hard List") from the Kopra Sentences were used in both A-only and in A + V test conditions.

To avoid possible practice effects, the A-only condition and A+V condition of the Kopra Easy List and Kopra Hard List were tested on separate days. Additionally, the list order and index order of sentences were both randomized. Multi-talker babble noise was used to drive subjects’ A-only performance on each condition to near 50% correct. The SNR was adjusted for each subject. The presentation level was minimally fixed at 60 dB SPL. The test environment was the same as for the speech-reading tests.

Table 1
*Subject demographic information such as age at testing, year of experience with a CI, processor manufacturer/type, and etiology*

<table>
<thead>
<tr>
<th>ID</th>
<th>Age</th>
<th>Type of CI</th>
<th>experience with CI</th>
<th>etiology</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIL1</td>
<td>72</td>
<td>Cochlear</td>
<td>8yrs</td>
<td>unknown</td>
</tr>
<tr>
<td>BIL2</td>
<td>53</td>
<td>Med EL</td>
<td>10yrs</td>
<td>hereditary</td>
</tr>
<tr>
<td>BIL3</td>
<td>53</td>
<td>MedEL</td>
<td>12yrs</td>
<td>unknown</td>
</tr>
<tr>
<td>BIL4</td>
<td>49</td>
<td>Cochlear</td>
<td>6yrs</td>
<td>unknown</td>
</tr>
<tr>
<td>BIL5</td>
<td>82</td>
<td>Medel</td>
<td>4yrs</td>
<td>hereditary</td>
</tr>
<tr>
<td>BIL6</td>
<td>62</td>
<td>Medel</td>
<td>4yrs</td>
<td>unknown</td>
</tr>
<tr>
<td>BIL7</td>
<td>69</td>
<td>Medel</td>
<td>3yrs</td>
<td>unknown</td>
</tr>
<tr>
<td>BIL8</td>
<td>75</td>
<td>Cochlear</td>
<td>11yrs</td>
<td>multiple factors</td>
</tr>
<tr>
<td>BIL9</td>
<td>87</td>
<td>Cochlear</td>
<td>12yrs</td>
<td>unknown</td>
</tr>
<tr>
<td>BIL10</td>
<td>21</td>
<td>Cochlear</td>
<td>2.5 yrs</td>
<td>unknown</td>
</tr>
<tr>
<td>BMD1</td>
<td>61</td>
<td>Cochlear</td>
<td>7yrs</td>
<td>unknown</td>
</tr>
<tr>
<td>Patient</td>
<td>Age</td>
<td>CI Brand</td>
<td>Years</td>
<td>Diagnosis</td>
</tr>
<tr>
<td>---------</td>
<td>-----</td>
<td>----------</td>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td>BMD2</td>
<td>79</td>
<td>Cochlear</td>
<td>5.5yrs</td>
<td>measles</td>
</tr>
<tr>
<td>BMD3</td>
<td>66</td>
<td>Medel</td>
<td>8yrs</td>
<td>unknown</td>
</tr>
<tr>
<td>BMD4</td>
<td>57</td>
<td>AB</td>
<td>6.5yrs</td>
<td>unknown</td>
</tr>
<tr>
<td>BMD5</td>
<td>69</td>
<td>AB</td>
<td>6yrs</td>
<td>unknown</td>
</tr>
<tr>
<td>BMD6</td>
<td>56</td>
<td>Cochlear</td>
<td>12yrs</td>
<td>unknown</td>
</tr>
<tr>
<td>BMD7</td>
<td>83</td>
<td>Medel</td>
<td>10yrs</td>
<td>Meniere's Disease</td>
</tr>
<tr>
<td>BMD8</td>
<td>70</td>
<td>AB</td>
<td>14yrs</td>
<td>hereditary</td>
</tr>
<tr>
<td>BMD9</td>
<td>73</td>
<td>Cochlear</td>
<td>5 yrs</td>
<td>unknown</td>
</tr>
</tbody>
</table>

**Note.** For year of experience with two CIs, it is calculated based on the fitting of the first CI for bilateral CI users. The year of experience with a CI for all subjects is rounded to the 0.5 year accuracy.
Results

**Speech-reading difficulty of the AzAV sentences.** The speech-reading scores of the Kopra Sentences and the AzAV sentences by NH subjects are shown in Figure 2. The performances of CI patients on the AzAV sentences are also shown in Figure 2.

The mean lip reading score of list 1 was 55.27% (SEM=3.308%). The mean lip reading score of list 4 was 46.25% (SEM= 4.002%). Mean lip reading score of list 8 was 30.25% (SEM=5.041%). Mean lip reading score of list 12 was 6.708% (SEM =1.720%).

The mean lip reading score of the AzAV sentences by NH listeners was 7.361% (SEM=2.049%). Mean lip reading score of AzAV sentences by CI patients was 10.78% (SEM=2.332%).

For the normal hearing listeners, a one-way analysis of variance (ANOVA) revealed a main effects of ‘list”(F (3,44)=53.03, P < 0.0001) among 4 lists. Sidak’s multiple comparisons test revealed that all pairs of mean of lists comparisons were significant different except comparison on mean of list 1 and mean of list 4 (55.27% vs 46.25%, t=1.837). This may due to the recording set-up of the present version Kopra sentences started much more difficult than original Kopra sentences (55.27% as mean of list 1 in ASU version vs 95% as mean of list 1 in Kopra’s original report). The paired-t test between list 12 and AzAV in speech reading performance of NH did not differ (t=0.4332, df=15). Thus, the AzAV lists should be considered as difficult to speech read. Additional unpaired-t test showed that NH performance and CI listeners’ performance on the AzAV lists did not differ (P=0.2871). Thus, for this material, CI listeners are not better speech readers than normal hearing listeners.
Figure 2. Lip reading scores of 4 lists Kopra sentences and AzAV sentences on NH subjects and lip reading scores of AzAV sentences in 19 CI subjects.
**Gain from easy- and difficult-to-speech-read material.** The A and A+V scores for the Kopra Easy List and Kopra Hard List are plotted in Figure 3 for the CI listeners. The mean A score of Kopra Hard List was 30.78% (SEM=3.393%). The mean A+V score of the Kopra Hard List was 63.313% (SEM=4.774%). The mean A score of the Kopra Easy List was 40.589% (SEM=2.737%). The mean A+V score of the Kopra Easy List was 86.423% (SEM=3.652%). Within-subject, two-way ANOVA revealed that condition (A or A+V), lip reading difficulty (Easy or Hard) as well as interaction between condition and lip reading difficulty all had significant effects on the scores (interaction: F(1,18)=11.88, P=0.0029; condition: F(1,18)=64.40, P<0.0001; Lip reading difficulty: F(1,18)=72.81, P<0.0001). The significance of the interaction between condition and lip reading difficulty reflected the outcome that the easier lip reading sentences provide larger magnitude of visual benefit. Easy List provided about 15% more visual benefit than Hard list, compared to corresponding A-only condition.

Figure 3(a). Summary of scores on A and A+V Kopra Hard list and Easy list on 10 CI users; Figure 3(b). The A+V score as a function of A score in Kopra Hard list and Kopra Easy list.
Discussion

The results for the lip reading difficulty of Kopra sentences validated Kopra’s original claim that the lists of sentences were ordered from easy to hard based on lip reading. However, we did see an effect of speakers or recording conditions, or perhaps in experience of the subjects, between the Kopra original version in 1986 and the present replication version in 2014. The average score of list 1 in Kopra et al. (1986) was 95%, which is nearly twice the score obtained in the present version.

Results on the speech reading performance of the Kopra sentences and the AzAV sentences by both NH and CI listeners indicated that the two groups of listeners did not differ on speech reading performance. Although a gender difference was seen in a previous report (female adult CI users were better lip readers than male adult CI users) (Strenlinkov et al., 2009), the unpaired t-test between 10 female CI users and 9 male adult CI users did not show a difference (13.52% vs 7.74%, P=0.2248). However, difference was in the same direction as in Strenlinkov et al. (2009).

Multi-level text analysis on 4 lists of Kopra sentences. In order to investigate factors that affected lip-reading difficulty, analysis of linguistic layers from a low level, statistical phoneme count to high level, semantic information were conducted. The four lists of Kopra sentences were used as materials.

First, each list of Kopra sentences has same number of sentences. So the number of sentences would not be the predictor that affected the lip reading difficultly. Secondly, Kopra indicated that sentences that were harder to lip read tended to have fewer different words. However, it was unlikely to be true since the difference among type-to-token ratio of lists are not close to the order of lip reading intelligibility. The list that had highest
type-to-token ratio was list 9 while list 3 and list 10 has same type-to-token ratio (Kopra et al., 1985). Finally, phonemes were counted across lists (see Figure 4). Text mining processes that extracted words from the text of sentences were done through the open-source Natural Language Tool Kit (NLTK) (Bird et al., 2009). The Carnegie Mellon University Pronunciation Dictionary (CMU Dictionary)\(^1\), which had over 125,000 words, was then used to decompose each word into a sequence of phonemes.

The results of the phoneme analyses are shown in Figure 4. No count of phonemes including bilabial phonemes could correlate with the speech read intelligibility trend in Figure 1, except phoneme “W”. However, the difference of count of “W” between list 1 and list 12 is just less than 10. It could not be able to explain the over 40% difference in speech read intelligibility between list 1 and list 12, especially there are 25 sentences per list.

The difference in the familiarity of sentences was suggested by Kopra et al. (1986) as a possible predictor of lip reading difficulty. In the present study, ideas derived from Natural Language Processing (NLP) were used to quantify the familiarity of sentences.

Phrased-based spoken language recognition of human speech has grown in popularity, especially in the approach to build an automatic speech recognition engine (Li et al., 2013). For example, when humans listen to a sentence such as “I love bacon ___” the follow up “strips” would be easier to recognize because the phrase “bacon strips” is much more familiar than other phrases such as “bacon cubes”, “bacon cheese” or even

\(^1\) http://www.speech.cs.cmu.edu/cgi-bin/cmudict
“bacon pizza”. In other words, the occurrence of “bacon strips” in daily life is much higher compared to other grammatically correct combinations.

Figure 4. Phoneme counts in 4 lists of Kopra sentences. Top: count of bilabial phonemes in 4 lists of Kopra sentences. Bottom: count of total phonemes in 4 lists of Kopra sentences.
To estimate the familiarity of sentences, the numeric of occurrence of phrases in all scanned books published in the United States were retrieved through Google Books Project (aka., Google N-grams) (Michel et al. 2011)\(^2\). To automate and standardize the process across all lists of sentences, each sentence was decomposed into a sequence of 3-Grams (three-word phrases). For example, “Do you want a cookie?” decomposed into “Do you want”, “you want a”, “want a cookie”. The occurrence of each 3-gram phrase in the 3-Gram American English database with smoothing factor 3 (Michel et al., 2011) from year 1986 to year 2008 was retrieved thought a web crawler and the NLTK platform. For example, the familiarity of “Do you want” was estimated by the percentage of numeric occurrence of “Do you want” in the summary count of all 3-grams in the over 5 million books in the Google N-grams Project (Michel et al., 2011). Since Google N-grams did have a threshold (occurs over 40 times in scanned books) to determine whether a 3-Gram occurrence was too low to store, a 0% was placed if no occurrence data was stored in Google N-grams.

**Outcomes for analyses of “familiarity”.** Several metrics of the occurrence of phrases were plotted in Figure 5. The mean occurrence, median occurrence, maxim occurrence, minimum occurrence of phrases in each sentence was first retrieved. In order to compare the occurrences across lists, the average of those metrics on 25 sentences was used. The year trend was also plotted in the Figure 4. Each chunk of lists contained data from 1986 to 2008. The results for each year are shown as a vertical striation.

---

\(^2\) For a sample task that fetches Google N-Gram data, please look at an open source GitHub project, https://github.com/econpy/google-ngrams
Automated Google 3-gram Occurrence Based on the Mean of Each Sentence

Automated Google 3-gram Occurrence Based on the Median of Each Sentence

Automated Google 3-gram Occurrence Based on the Max of Each Sentence

Automated Google 3-gram data mining results based on min

Figure 5. Numeric Occurrence of phrases from Google N-gram Project.
Overall, the results support the proposition that the major factor that affects the speech reading score is the familiarity of the word sequences in a sentence. All four metrics on the occurrence of phrases reflected the overall trend in the lip reading difficulty score in Figure 1. It is also interesting that year had an effect on the occurrence, especially in phrases with high occurrence such as in list 1.

Auditory information is predominantly important for speech perception with no doubt. Previous results on audiovisual consonant recognition in both NH and CI listeners (Desai et al., 2008; Sheffield et al., 2014) indicated that visual information provided place of articulation information primarily. However, simple consonant place information from visual information is not able to explain why 60% intelligibility can be achieved with only visual information – as in List 1 of the Kopra Sentences. This observation helps us distinguish the roles of audiovisual information integration at an early stage of sentence recognition vs. at a later stage. In other words, Audiovisual integration models that work in either consonant or word level recognition (Braida 1991; Massaro 1987; Blamey et al., 1989), or models that indicates CI users were better integrators than NH at an early stage (Rouger et al., 2007), may not apply in a discussion of sentence recognition.
CHAPTER 2

EXPERIMENT II

Introduction

Over decades of effort, scientists have made significant progress on improving the performance of cochlear implants (Dorman and Wilson, 2004). Back to the late 1980s, when cochlear implantation was first approved in the US, only patients with bilateral severe-to-profound deafness fit the candidacy requirements for a CI (Wilson and Dorman, 2012). These indications have now loosened. A large body of evidence now suggests that individuals with a CI fitted in one ear benefit significantly when low-frequency residual hearing in the implanted ear, or in the contralateral ear, receives amplification (Ching et al., 2007; Gifford et al., 2010). Bilateral implants also offer benefits to performance relative to a single CI (Dorman et al., 2011).

Speech perception is not entirely auditory. Bimodal fitting (a CI in one ear and low-frequency hearing in the other ear), bilateral fitting, and hearing preservation surgery (low-frequency hearing preserved in the operated ear) are major approaches to maximize information input based on the “auditory-centered” view of speech understanding (Ching et al., 2007; Zhang et al., 2010; Dorman & Gifford, 2010). However, it is a fact that face-to-face communication plays a vital role in daily life for both normal hearing listeners and hearing impaired patients such as CI users (Cook et al., 2014). To reveal the role of visual cues in the service of speech perception, Campbell (1988) reviewed a large body of studies that focused on the basis of audiovisual speech perception in neurological and psychophysical views. A more extensive view, by some neurologists, claimed that multisensory speech was primary in the service of speech perception, rather than the view
that visual information was just supportive/add-on information to the information derived from the auditory modality (Rosenblum et al., 2007). Rosenblum pointed out that audiovisual speech perception was “automatic” and “ubiquitous” that behavioral and neuropsychological evidence supported that view that multi-modality processing even occurs at an early stage of speech perception. Silent lip-reading (without information input from an auditory stimulus) activates sites in both primary auditory cortex and secondary auditory cortex in both NH and deaf patients (Calvert et al., 1997; Pekkola et al., 2005; MacSweeney et al., 2002; Woll 2012). Corresponding to evidence showing that visual information, such as silent lip reading, activated auditory cortex sites, Von Kriegstein (2012) found that acoustic input activated standalone cortex areas that were related to visual information processing after a short time period of audiovisual information practice. This prior face-to-face short training also benefitted auditory-speech recognition behaviorally. In summary, speech perception is multimodality dominated.

**Benefits from visual information for CI listeners.** CI users show the ability to integrate auditory and visual cues to get benefits for speech perception, despite receiving degraded signals from the CI. Van Dijk (1999) reported near 50% improvement from a CI-only condition to a CI-plus-visual-cues condition when performance on sentence recognition was assessed. In a longitudinal study, Rouger et al. (2007) reported about 20% improvement on a French word recognition test when a CI-only condition was compared to a CI-plus-visual condition. Dasei et al. (2008) reported about a 30% improvement in percent correct phonemes when comparing auditory-plus-visual to either auditory-only or visual-only conditions. Pre-lingually deaf children who were implanted for over 3 years got average of 30% benefit from adding visual cues to acoustic cues,
while early implanted children got a larger 40% benefit from adding visual cues to acoustic cues (Bergeson et al., 2005).

Additionally, studies on CI users using “McGurk” (McGurk, 1976) stimuli investigated audio-visual information integration on phonemes. Pre-lingually implanted children before age 10 did not show the “McGurk effect” (Tremblay et al., 2007) while proficient adult CI users were able to show the ability to integrate both congruent and incongruent audio and visual stimuli on phoneme perception (Dasei et al., 2008; Tremblay et al., 2010; Rouger et al., 2008).

**Factors underlying A-V benefit.** Previous reports in NH listeners and CI recipients indicated that the sources of benefits from adding visual information varied. For example, adding visual information helped decrease the speech detection threshold (Strelnikov et al., 2009). It also led to better speech understanding for listeners when speaker had a foreign accent (Strelnikov et al., 2009). Zue (1985) pointed out that only an average of 2.4 words would fit a given feature sequence in a generic mental lexicon. Van Tassel et al. (1987) suggested that using only visual cues and envelope cue provided 95% percent correct of phoneme perception. Results from Zue (1985) and Van Tassel et al. (1987) may explain why CI users did reasonably well in audiovisual speech perception. Numerous reports also have indicated that positive statistical relationships exist between visual cues such as eyebrow movements and lip movements and acoustic properties (Hadar et al., 1983; Munhall et al., 2004; Scarborough et al., 2009; Grant & Seitz, 2000; Chandrasekaran et al., 2009). Besides behavioral and statistical evidence, visual cues such as lip movements alter and can enhance cortical neuronal responses (Ross et al., 2007; Davis et al., 2008; McGettigan et al., 2012). For example, seeing the face of the
selective talker led cortical neurons in auditory cortex to better phase-locking of the acoustic envelope in a cocktail-party like environment (Golumbic et al., 2013).

The rationale for Experiment II. Earlier it was noted that bimodal CI fittings (CI in one ear and a hearing aid in the other ear) and bilateral CI fittings provided a level of speech understanding higher than that provided by a single CI. In Experiment II we ask whether this improvement also occurs when visual information is available to the listeners. It is possible that when the information from vision is added to that from a CI, the information from a low-frequency acoustic hearing or a second CI becomes redundant and does not lead to an improvement in performance. Thus, we ask if (i) low-frequency acoustic hearing adds to the intelligibility received from a single CI when visual cues are available (i.e., is CI+V+HA > CI+V) and (ii) whether a second CI benefits a single CI when visual cues are available (i.e., is CI+CI+V>CI+V)?
Methods

Subjects. Nine bimodal CI users (3 males, 6 females) and 10 bilateral CI users (6 males, 4 females) participated in the study. The audiograms for the acoustically stimulated ear of the bimodal patients are shown in Figure 6. The mean thresholds at 125, 250, 500 and 750 Hz were 47.2 dB HL, 51.7 dB HL, 71.1 dB HL, 80 dB HL respectively. The mean age of the bimodal patients was 68.2 years. The mean age of the bilateral patients was 62.3 years. All participants had at least two years’ experience with their CI. These and other demographic data are summarized in Table 1 (BMD refers to bimodal subjects; BIL refers to bilateral subjects).

Materials. The AzAV sentences, described in Experiment 1, were used as target sentences.

Procedure. The CI patients sat in front of a 21-inch video monitor in a sound booth with a single loudspeaker at 1-foot distance. The presentation level of the sentences was nominally set at 60 dB SPL. Multi-talker babble noise was used to drive each subject’s performance to near 40% percent correct in the A-only condition. 15 out of 19 subjects were tested at +3 dB SNR. 3 (BMD1, BMD3, BIL3) subjects were tested at +7 dB SNR and 1 subject (BIL9) was tested at +10 dB SNR.

For bimodal CI subjects, HA only, CI only, CI+V, HA+CI, HA+CI+V sessions were tested in random order. For bilateral CI subjects, both CIs were tested to determine the poorer and better sides. The better CI+V and bilateral (Bi) CI+V conditions were then tested in random order. For both bimodal and bilateral CI users, an additional V-only (i.e., speech reading without sound) condition was tested. Two lists of the AzAV
sentences were presented in each test condition. A short practice session preceded each test condition.

Figure 6. Unaided audiograms of 9 bimodal CI subjects’ non-implanted ear in the present study. For No response input, a 125 dB HL was used instead in the audiogram. Error bars represent to 10-90 percent range.
Results

**Bimodal subjects.** The bimodal CI subjects’ sentence recognition scores are shown in Figure 2. One subject’s score (BMD3) was dropped since his CI was programmed to receive no information from 500 Hz or below, which may alter the integration of bimodal hearing since the HA provides information primarily from that region (Zhang et al., 2010). The mean score in the V-only condition was 9.8 % (Standard Error of Mean (SEM)= 3.68%). The mean score in the HA-only condition was 4.3% (SEM =1.82%). The mean score in the CI-only condition was 38.3% (SEM = 6.23%). The mean score in the CI+HA condition was 42.4 % (SEM = 5.88 %). The mean score in the CI+V condition was 60.1% (SEM = 7.91 %). The mean score in the CI+V+HA condition was 75.0 % (SEM = 5.63 %).

A repeated measure ANOVA showed a significant effect of conditions (F (5,35)=53.7, p<0.0001). Holms-Sidak post-hoc pair-wise comparisons were performed on all possible pairs of conditions. All outcomes are shown in Table 2. Critically, the mean score in the HA+CI condition (42 percent correct) was not significantly different from the mean score in the CI-only condition (38 percent correct). Thus the low-frequency information from the hearing aid did not improve intelligibility when that information was added to the CI-alone.

Consider now performance in the CI+V condition (60 percent correct). When low frequency information from the HA was added to this condition, then performance did improve (CI+V+HA = 75 percent correct). Thus, low frequency information from the HA was of value when visual information was also available.
Figure 7. Plot of scores of six test conditions in 8 out of 9 bimodal CI users. The mean and SEM of each condition was plotted.
**Bilateral subjects.** The bilateral CI subjects’ sentence recognition scores are shown in Figure 8. One subject (BIL3), who was tested at +7 dB SNR, showed a possible ceiling effect (score of poorer CI = 80 %). For that reason the following description of results is based on the results from 9 subjects. The mean score in the V-only condition was 12.7 % (SEM= 3.75%). The mean score in the poorer CI condition was 29.9 % (SEM = 3.77 %). The mean score in the better CI condition was 37.9 % (SEM = 6.60 %). The mean score in the Bi CI condition was 46.9 % (SEM = 6.13%). The mean score in the better CI+V condition was 69.3 % (SEM = 5.83%). The mean score in the BiCI+V condition was 70.24 % (SEM = 5.73%).

A repeated measures ANOVA showed a main effect of conditions (F(5,40)=74.3, p<0.0001). Holm-Sidak post-hoc pairwise comparisons were performed in all possible pairs of conditions. All outcomes are shown in Table 3. The BiCI condition (47 correct) was significantly higher than the poor CI-only condition (30 % correct) but not higher than the better CI-only condition (38 percent correct). Thus no evidence of summation (BiCI > better CI) was obtained. A similar outcome was obtained when V was present, i.e., BiCI+V (70 percent correct) was not higher than the better CI + V condition (69 percent correct).

**Average benefit of adding vision.** In order to investigate the average benefit from adding vision, pairs of conditions such as CI+V vs CI only, HA+CI+V vs HA+CI, Better CI+V vs Better CI only, Bi CI+V vs Bi CI were combined. Linear regression fitting of the A+V score as a function of the A score was also plotted (Figure 9). The average A-only score across all selected subjects was 42 percent correct. The average
A+V score across all subjects was 69 percent correct. Linear regression goodness of fit (P<0.0001) showed that subjects received about 30% improvement from adding vision regardless of the starting point of the A-only score (Figure 9).

Figure 8. Plot of scores of six test conditions in 9 out of 10 bilateral CI users. The mean and SEM of each condition was plotted.
Figure 9. Plot of A+V score as a function of A-only score in bimodal and bilateral CI users.
Discussion

Gain from adding V to A. Previous studies using audiovisual sentence recognition tests with CI users used various test materials and test environments, e.g., Van Dijk et al., 1999; Bergerson et al., 2003). Van Dijk et al. (1999) reported an average 43 points improvement from A to A+V using the VLS sentences (Video Lists of Sentences) in quiet. In this study the average V-only score was 32%. Bergerson et al. (2003) reported a slightly lower level of benefit (about 30%) in a group of CI users using CUNY sentences. The average score in the V-only condition in the Bergerson et al. (2003) study was around 15% (estimated from Figure 1 in Bergerson et al), which was similar to the lip reading score in the present study. And, in the present study, the average gain from A to A+V was 27% -- very similar to that found by Bergerson et al. (2003).

This degree of visual benefit (27 percentage points) was close to that found in the Grant and Seitz (1998) study on hearing impaired listeners when the material was an audiovisual version of the IEEE sentences. Grant and Seitz (1998) reported the mean benefit from adding vision was approximately 30% (estimated from Figure 1 in Grant and Seitz) when the mean V-only score was less than 10%. Sommres et al. (2005) study on young and older normal-hearing adults also reported a similar level of benefit from adding vision (30%) when the V-only score was reported to be near 10 % when using the Iowa Audiovisual sentences (estimated from Figure 1’s 3rd sub figure in Sommres et al.). The level of V-only only score in the present study was 11.4 %, which is close to Grant and Setiz (1998), Bergerson et al. (2003) and Sommres et al. (2005) study.

In summary, multiple studies, including this one, find that when the V-only score is very low (10-15 % correct), i.e., when the material is difficult to lip read (referenced to
the Kopra Hard Lists of Experiment 1) then adding V to A produces about a 30 percentage point improvement in performance. When the material is easier to lip read, then the benefit of adding V to A is larger – as found in Van Dijk et al. (1999). These reports support the observation from Experiment I that the lip reading difficulty of sentence materials has a close correlation with the amount of benefit when visual information is added to auditory information. This is not a surprising outcome.

In summary, both Experiment I’s results on the Kopra Easy sentences (where the V-only score was 55%) and Van Dijk et al. (1999) study using the VLS sentences (where the average V-only score was 32% among three tests) should be considered relatively easier to lip reading than those sentences described earlier. These authors also reported a larger benefit from adding vision - Experiment I reported 46% benefit adding vision to the Kopra Easy sentences; Van Dijk et al. (1999) reported average 42.7% benefit from adding vision to the VLS sentences.

**Bimodal CIs.** As indicated in the introduction, it is well established that low-frequency acoustic information from a HA can improve the performance of a CI by up to 30-40 percent points (Zhang et al., 2010; Dorman and Gifford 2010). At issue in this experiment was whether HA information is still of value when the CI listener has visual information also available.

In order to see the value of the HA in the CI+V+HA condition, the CI+V score needed to be far off of the ceiling. The results of previous experiments suggested that the gain from V was about 30 percent points. To keep the CI+V score far from the ceiling, the CI-alone score need to be driven down by adding a significant amount of noise (i.e., by using a SNR of +3 dB, which for CI patients is a very difficult SNR).
Hearing aids do not provide a high level of information in high noise environments. Thus the A-alone score at +3 dB SNR was less than 10 percent correct. In this circumstance, the HA did not benefit performance when added to the CI. This is a different outcome from previous studies and is, most likely, related to the starting level of performance in the CI-alone condition. In previous experiments when vision was not a factor, performance with the CI could start higher – and that was accomplished by adding less noise than that used in the present experiment (+10 dB SNR is a common value in previous experiments).

The novel and unexpected outcome of the present experiment was that the HA did improve performance when the reference condition was CI+V. All eight subjects showed an increase in performance, which indicated that this effect was consistent among subjects (Figure 5(a)). The similarity in outcome with previous experiments is that when the baseline performance is relatively high, either because a better SNR is used in the CI-alone condition, or V is added to the CI-only condition (with a poor SNR), then the low-frequency information from a HA benefits performance.

Sources of benefit from a HA include low frequency information, especially F0 cues. F0 cues can help better segment speech streams (Dorman and Gifford 2010; Zhang et al., 2010; Spitzer et al., 2009). Statistical analysis of the relationship between visual information and speech envelope information indicated that visual information helped track and segment speech streams since they provided unique marks (such as lips opening and closing) (Chandrasekaran et al., 2009). By using test materials that short phrases (4 to 5 words) were balanced of strong and weak syllables to detect segmentation strategies in speech recognition (Spitzer et al., 2009), Tillery et al. (2013) showed that visual
information helped reduce lexical boundary errors in normal hearing listeners. Thus HA may help better segment speech streams when V is present, which increases the chance of recognizing words in a sentence.

Another account for this outcome can be based on the amount of information available to the listener in the CI-only condition, when the performance level was less than 40% correct, and in the CI+V condition, when the performance level was 60 percent correct, presumably due to the added information about place of articulation provided by vision. In other words, low frequency information is not that helpful when intelligibility is low due the high level-noise (CI only score < 40%, see Figure 2). The starting point reaches 60% when V is present. The HA provides complementary low frequency information while V provides large portion of place information (Sheffield et al., 2014).

Experiment I indicated that top-down information such as familiarity was the driving factor for both lip reading difficulty. So information from HA became useable when V is present since the available context level was much higher compared to condition that V is absent.

**Bilateral CIs.** Similar to the question raised with bimodal CI users, the key question here is whether a second CI still provides benefit when V is present.

The data indicated that summation, i.e., BiCI > best CI, was not seen in in the condition without vision or in the condition with vision.

A single loudspeaker speaker listening environment was clearly not optimal to show the value in speech perception of having a second CI for CI users. The small, non-significant, benefit (9%) from having a second CI was consistent with previous findings (see Schafer et al., 2011 for a meta-analysis on bilateral CI users). Other test
environments that allow both squelch and summation, such as simulated real listening environment called R Space™ (Revitt, 2007; also see Loiselle 2013 for a comprehensive study on complex listening environments on bilateral CI users) would, in retrospect, have been a better choice of test environment.
Table 2

*Holm-Sidak’s multiple comparisons test for 8 bimodal patients*

<table>
<thead>
<tr>
<th>Holm-Sidak’s multiple comparisons test</th>
<th>Mean Diff.</th>
<th>Significant?</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>HA only vs. HA+CI</td>
<td>-38.06</td>
<td>Yes</td>
<td>****</td>
</tr>
<tr>
<td>HA only vs. CI only</td>
<td>-34.03</td>
<td>Yes</td>
<td>****</td>
</tr>
<tr>
<td>HA only vs. CI+V</td>
<td>-55.84</td>
<td>Yes</td>
<td>****</td>
</tr>
<tr>
<td>HA only vs. HA+CI+V</td>
<td>-70.70</td>
<td>Yes</td>
<td>****</td>
</tr>
<tr>
<td>HA only vs. V only</td>
<td>-5.514</td>
<td>No</td>
<td>ns</td>
</tr>
<tr>
<td>HA+CI vs. CI only</td>
<td>4.030</td>
<td>No</td>
<td>ns</td>
</tr>
<tr>
<td>HA+CI vs. CI+V</td>
<td>-17.78</td>
<td>Yes</td>
<td>**</td>
</tr>
<tr>
<td>HA+CI vs. HA+CI+V</td>
<td>-32.64</td>
<td>Yes</td>
<td>****</td>
</tr>
<tr>
<td>HA+CI vs. V only</td>
<td>32.54</td>
<td>Yes</td>
<td>****</td>
</tr>
<tr>
<td>CI only vs. CI+V</td>
<td>-21.81</td>
<td>Yes</td>
<td>**</td>
</tr>
<tr>
<td>CI only vs. HA+CI+V</td>
<td>-36.67</td>
<td>Yes</td>
<td>****</td>
</tr>
<tr>
<td>CI only vs. V only</td>
<td>28.51</td>
<td>Yes</td>
<td>****</td>
</tr>
<tr>
<td>CI+V vs. HA+CI+V</td>
<td>-14.86</td>
<td>Yes</td>
<td>*</td>
</tr>
<tr>
<td>CI+V vs. V only</td>
<td>50.32</td>
<td>Yes</td>
<td>****</td>
</tr>
<tr>
<td>HA+CI+V vs. V only</td>
<td>65.18</td>
<td>Yes</td>
<td>****</td>
</tr>
</tbody>
</table>
Figure 10. Benefit from adding HA when V is absent and when V is present. The x-axis marks the corresponding HA only score (%).
Figure 11: Benefit from adding HA when V is absent and when V is present. Each bimodal subject’s CI only, CI+HA, CI+V, CI+HA+V was plotted in a subfigure.
Table 3

*Holm-Sidak’s multiple comparisons test for 9 bilateral patients*

<table>
<thead>
<tr>
<th>Holm-Sidak's multiple comparisons test</th>
<th>Mean Diff.</th>
<th>Significant?</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>poor vs. BiCl</td>
<td>-16.97</td>
<td>Yes</td>
<td>***</td>
</tr>
<tr>
<td>poor vs. better</td>
<td>-7.990</td>
<td>No</td>
<td>ns</td>
</tr>
<tr>
<td>poor vs. Better+V</td>
<td>-39.32</td>
<td>Yes</td>
<td>****</td>
</tr>
<tr>
<td>poor vs. BiCl+V</td>
<td>-40.30</td>
<td>Yes</td>
<td>****</td>
</tr>
<tr>
<td>poor vs. V only</td>
<td>17.26</td>
<td>Yes</td>
<td>***</td>
</tr>
<tr>
<td>BiCl vs. better</td>
<td>8.982</td>
<td>No</td>
<td>ns</td>
</tr>
<tr>
<td>BiCl vs. Better+V</td>
<td>-22.35</td>
<td>Yes</td>
<td>****</td>
</tr>
<tr>
<td>BiCl vs. BiCl+V</td>
<td>-23.33</td>
<td>Yes</td>
<td>****</td>
</tr>
<tr>
<td>BiCl vs. V only</td>
<td>34.23</td>
<td>Yes</td>
<td>****</td>
</tr>
<tr>
<td>better vs. Better+V</td>
<td>-31.33</td>
<td>Yes</td>
<td>****</td>
</tr>
<tr>
<td>better vs. BiCl+V</td>
<td>-32.31</td>
<td>Yes</td>
<td>****</td>
</tr>
<tr>
<td>better vs. V only</td>
<td>25.25</td>
<td>Yes</td>
<td>****</td>
</tr>
<tr>
<td>Better+V vs. BiCl+V</td>
<td>-0.9850</td>
<td>No</td>
<td>ns</td>
</tr>
<tr>
<td>Better+V vs. V only</td>
<td>56.58</td>
<td>Yes</td>
<td>****</td>
</tr>
<tr>
<td>BiCl+V vs. V only</td>
<td>57.56</td>
<td>Yes</td>
<td>****</td>
</tr>
</tbody>
</table>
Figure 12. Benefit from adding a CI (Poor CI) when V is absent and when V is present.

The x-axis marks the corresponding Poor CI only score (%).
Figure 13. Summary plot of benefit from adding 2\textsuperscript{nd} CI when V is absent and when V is present for each bilateral subject. Each bimodal subject's better only, BiCl, better+V, BiCl+V was plotted in each corresponding sub figure.
CHAPTER 3

CONCLUSION REMARKS

Limitations and beyond: offering additional hearing for a single CI in various environments

As described earlier in the introduction, A HA or a second CI is common to add to a single CI in order provide additional speech information for CI users. These additional hearing could provide up to 40 percent points benefit in speech understanding (Zhang et al., 2010) or additional cues to reduce localization errors (Dorman et al., 2011).

Though these effects were not seen in the present study due to the artificial high-level noise or the single speaker setting. It should be noted that additional hearing such as HA was still useable even when visual information could provide 30% of benefit. The current study was rather a simulation of speech recognition in some environments than reflections in comprehensive and various listening conditions that CI users will meet in the real world. Conclusions such as a second CI was not useable should be referenced carefully with context in the present study, where a single speaker set-up is used. Other experiment set-ups such as roving targets and cocktail-party like complex listening environment in the view of audiovisual speech are highly encouraged to test and examine.

Bounded magnitude of visual benefit: from classic behavior experiments to massive data mining

The present work describes general bounds for visual benefit in audiovisual sentence recognition in CI users but not limited to the group of CI users. Very low lip reading intelligibility sentences provide about 30% benefit when visual information is
added to auditory-only condition. Sentences that are very easy to lip reading can provide near 50% of benefit when visual information is added.

Both the lower bound and upper bound can be examined at some degree. Now consider a brainstorm experiment that normal sentences are reconstructed with a random order of consonant and vowels. It makes sense to predict that audiovisual sentence recognition will be closely to be considered as a stream of audiovisual consonant and vowel recognition test. Previous studies (Desai et al., 2008; Sheffield et al., 2014; Sommres et al., 2005) reported about 30 to 35% improvement when V is added to A in consonant recognition test. It helps us understand why some very low lip reading intelligibility sentences can still provide at least near 30% visual benefit.

Kopra et al. (1986) reported a near perfect lip reading intelligibility (over 90% in list 1), which is not even close to the replication in the present study (55%) since no assumption about conditions of materials were made as the same between two experiments. It is logically and notionally correct to think that the upper bound of visual benefit is over 90% if Kopra’s original claim can be held, at least.

The present work does not start from the scratch. Instead, it starts with solid background from previous researchers’ work such as Summerfield (1990) and Kopra et al. (1986). Standing on the “giant’s shoulder”, it is fortunate to look further nowadays. It is also a fortunate that Kopra et al. (1986)’s original assumption on the effect of familiarity can be tested in the present study. Kopra’s assumption is not likely to validate without the boom of cutting-edge technology such as massive data mining on the web since 21st century.
Classic behavior experiments require both a fine control and an accurate experiment design in order to work with limited amount of data, often with the help from statistics. As pointed out earlier, it is optimal that practices and applications of massive data mining and other technique such as deep neural network simulations advance not only the knowledge in the field of speech and language, but also in the whole body of knowledge in human beings in a much larger scale in the near future.

A famous researcher and pioneer in the field of Cochlear Implants, Prof. Philip Loizou, used a title for a review paper, which I like very much. The title also inspires me a lot since it is a concrete understanding from a renowned researcher in the field of Cochlear Implants. The title is “mimicking the human ear” (Loizou 1998).

Current CI provided only degraded speech signals to CI users due to the limitation of signal processing strategies. It is optimal to think that Dr. Loizou’s prediction on “mimicking the human ear” will become real in the near future, which definitely will benefit people who suffer from the loss of hearing. It is even not too bold to imagine that the natural bounds of human beings such as 20Hz - 20,000Hz frequency response can be extended.
REFERENCES


