Master's Thesis

Study on Lipreading Performance in Speech Learning for Hearing Impaired People

Jie Hou

Supervisor Masashi Unoki

Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
(Information Science)

August, 2022

Abstract

Background: Visual information plays an important role in speech perception, especially for people with hearing loss. The perceptual compensation hypothesis proposes that auditory deprivation enhances visual processing of deaf people. However, whether hearing impaired people have superior lipreading ability than normal hearing people is still controversial.

Purpose: The main purpose of this study is to investigate the lipreading performance and visual attention strategy of hearing impaired people in speech learning. Our main hypothesis is that deaf people will have higher learning efficiencies in visual speech learning for their dependence on visual modality; and deaf people will spend more time focusing on the speakers' mouth due to their special visual language experience.

Method: Both prelingually deaf adults (n = 19) and normal hearing adults (n = 17) participated in this experiment. Chinese subjects with no Japanese experience were required to learn Japanese kana and words through pronunciation videos. The lipreading abilities of individuals and groups were assessed and compared by learning efficiencies in visual-only (VO) and audiovisual (AV) speech learning tasks. Meanwhile, eye movement data were recorded to explore differences in the visual distribution of speech perception.

Results: Lipreading accuracy of deaf people was not significantly higher than hearing people. Both prelingually deaf adults and normal hearing adults spent more attention on speakers' mouth in speech learning, but the proportion of hearing adults on the mouth was even higher than that of deaf adults.

Conclusion: Deaf people don't have superior lipreading ability and more reasonable selective attention in speech learning. Given that deaf people have higher lipreading performance in native speech perception, it is possible that the visual language experience facilitates lipreading ability. Therefore, it is of great importance to incorporate lipreading training into speech rehabilitation and language teaching for deaf people.

Keywords: deaf; hearing; lipreading; visual attention

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Speech perception is a cross-modal activity: in addition to auditory information, the visual information also plays an important role in speech processing of human being, because people's speech contains both acoustic cues and articulatory cues. Accordingly, it is found that congruent visual input facilitates people's speech perception (Lusk & Mitchel, 2016), while incongruent or ambiguous visual input misleads people's speech perception [McGurk effect, (McGurk & MacDonald, 1976)].

Usually, people can integrate visual and auditory sense naturally in speech perception (**multisensory integration**). But for hearing impaired people, degraded auditory input brings increased dependence on visual input. For example, many studies have shown that wearing masks have a large negative impact on deaf people's communication during the COVID-19 (Saunders et al., 2021; Poon & Jenstad, 2022).

There is widespread concern about how hearing impairment and visual dependence shape speech perception of deaf people. Previous studies believed that sensory deprivation in hearing modality alters visual modality [**cross-modal plasticity**, (Bavelier & Neville, 2002)]. A well-known viewpoint to explain the cross-modal plasticity is the **Perceptual Compensation Hypothesis** (Ronnberg, 1995). Based on this hypothesis, previous research have confirmed that hearing impaired people have enhanced peripheral vision and efficient attention distribution in periphery (Bavelier et al., 2006; Shalev et al., 2020).

Overall, previous research demonstrated that visual dependence can enhance vision of hearing impaired people on some specific measurements. However, it is still inconclusive whether visual dependence will lead to higher lipreading performance for hearing impaired people. Therefore, investigating visual speech performance of hearing impaired people can provide more meaningful evidence for or against the perceptual compensation hypothesis.

1.2 Research Purpose & Questions

Given the controversial lipreading ability of the deaf and the lack of research regarding lipreading performance in speech learning, this study will aim to evaluate the performance of visual speech perception for hearing impaired people.

Two important aspects of visual speech perception are considered:

- Lipreading performance (Outcome of the visual speech perception): to evaluate the learning efficiencies of different groups and different modalities.
- Visual attention allocation (Pathway of the visual speech perception): to compare the attention distribution between hearing-impaired people and normal hearing people.

By doing these, this study would like to answer the following two Research Questions:

Q1: Do hearing impaired people have enhanced lipreading ability compared to normal hearing people in speech learning?

Q2: What are the differences in visual attention strategies between hearing impaired people and normal hearing people in speech learning?

1.3 Originality and Significance

The originality of the current study is to assess lipreading ability of individuals and groups by introducing a new language. Since the hearing impaired people have richer experience, practice, and even systematic training in lipreading than normal hearing people, the speech learning with stimuli from the new language will bring hearing impaired people and normal hearing people to the same starting point, and eliminate interference effect from different visual language experiences.

This study will highlight the importance of visual information and lipreading in speech learning for hearing-impaired people, motivate clinical and personal lipreading training, and provide convincing evidence for or against the perceptual compensation hypothesis. Moreover, research on attention allocation will explain the cause of different learning efficiencies and reveal the mechanism of audiovisual integration for hearing-impaired people.

1.4 Outline of thesis

In Chapter 1, the research background of the study has been introduced; the research objectives and questions have been identified; and the originality and significance of the study have also been discussed.

In Chapter 2, the existing literature will be reviewed from two principal perspectives: empirical research on lipreading and eye movement research on visual speech perception.

In Chapter 3, the research philosophy, experimental design and experimental assumptions will be elaborated.

In Chapter 4, the details of the experiment will be presented as Participants, Stimuli and Apparatus, Procedure, and Data analysis.

In Chapter 5, both behavioral results and eye movement results will be presented.

In Chapter 6, the results will be interpreted and discussed from several important topics.

In Chapter 7, the main conclusions, important inferences, and future plan will be summarized.

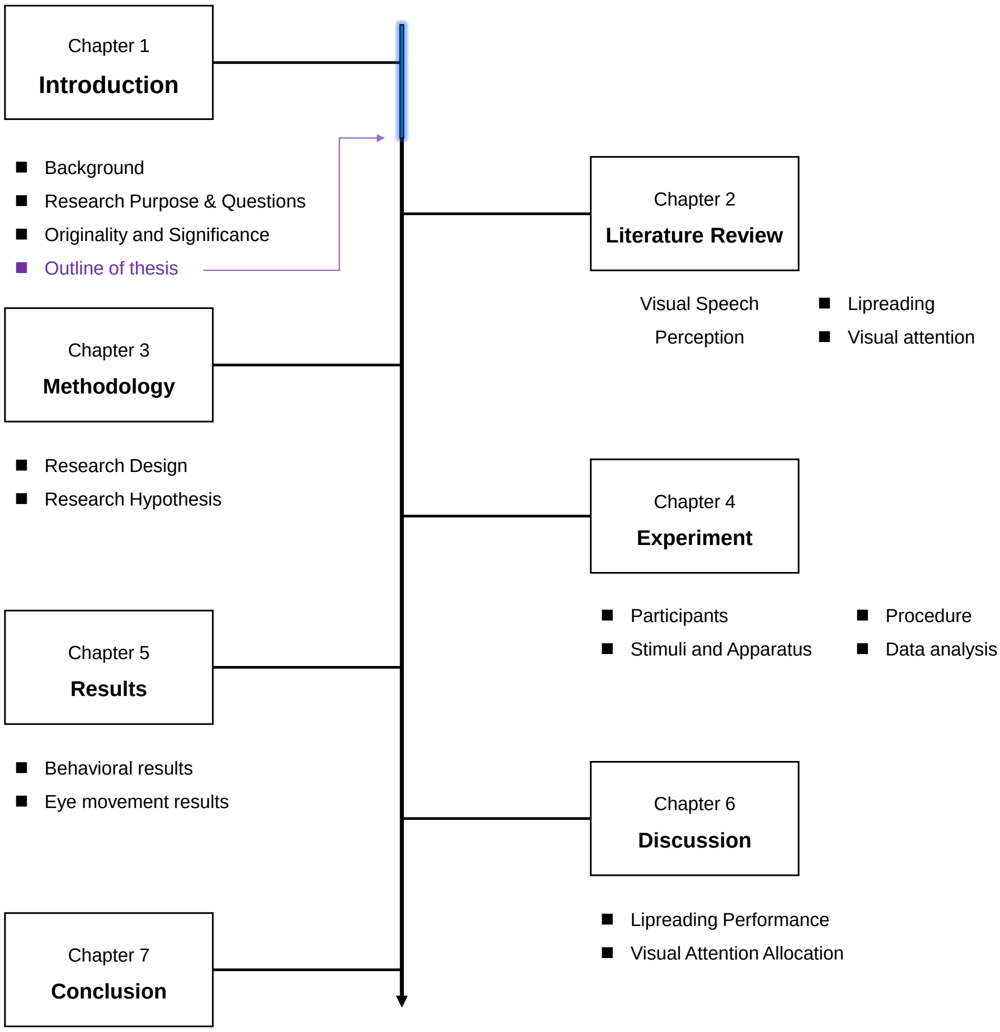

Figure 1.1: Outline of thesis

Chapter 2

Literature Review

2.1 Theoretical Debate on Deaf Vision

There is no doubt that hearing loss in deaf people alters their visual performance. In addition, deaf people differ from hearing people in two ways: First, impaired hearing input forces deaf people to increase their reliance on visual input for necessary linguistic and emotional information, which may make some changes in the speed and ability of visual processing. Second, most of the deaf people communicate using sign language, which may shape their visual attention between central and peripheral vision. However, whether the altering, changing or shaping is positive or negative is still debated [see Qin & Yan (2021) and Dye & Bavelier (2010) for review of different views and hypotheses].

The **deficit view** states that deaf people have lower visual processing abilities than hearing people. ***Auditory scaffolding hypothesis*** (Conway et al., 2009) proposes that sound can provide time and sequence information to support individuals' processing of time series information, and provide 'scaffolding' for the development of related cognitive abilities. Impaired processing of time-series information in deaf individuals leads to deficits in visual function. ***Division of labor***

hypothesis (Mitchell & Quittner, 1996) proposes that the healthy development of attention requires the coordination of multiple sensory pathways. The absence of auditory information changes the visual attention pattern of deaf people: hearing people has selective attention and can flexibly allocate visual attention resources according to the experimental requirements, but the distribution of visual attention resources of deaf people is scattered.

Contrastively, the **enhancement view** believes that deaf people have better visual processing than hearing people. ***Reactivity enhancement hypothesis*** (Bottari et al., 2010) thinks that deaf people react faster for stimuli in parafoveal vision and peripheral vision. That is, the modal compensation is in the time dimension. ***Perceptual enhancement hypothesis*** (Buckley et al., 2010) thinks that hearing loss increases the sensitivity of deaf people to information in peripheral vision and increases the perceptible range of the peripheral vision. That is, the modal compensation is in the spatial dimension.

Numerous experimental studies have been conducted to support these theories in terms of visual sensory thresholds (Bavelier et al., 2006), visual attention and facial emotion recognition (Sidera et al., 2016). However, little attention has been paid to lipreading performance of deaf people.

2.2 Empirical Research on Lipreading

The term **lipreading** in current study specifically refers to 'the process and ability to acquire linguistic information from movement of the mouth and lips'. Some studies also call it **speechreading**. As the important component of the **Oralism** (communication through lipreading, residual hearing and spoken language; as opposed to **Manualism** which supports sign language communication), lipreading has a long history of more than two hundred years in deaf education, while the empirical research has only begun to attract attention in the last few decades.

The lipreading ability of deaf people is still controversial. Most studies concluded that deaf adults who have been deaf from an early age have equivalent or superior lipreading abilities compared to hearing adults (Worster et al., 2018). In a study of the lipreading performance of congenitally deaf adults with a silent video watching task, Pimperton et al. (2017) found that the word identification rates for CI participants were significantly higher than hearing partcipants. Moreover, the result showed a correlation between later implantation and superior speechreading skills, which seems to suggest that more visual dependence leads to higher lipreading skills for hearing impaired individuals. However, some studies have come to the opposite conclusion. It is also found that the young children and older adults have no difference in lipreading performance with their hearing counterparts (Kyle et al., 2013; Tye-Murray et al., 2007).

Meanwile, previous studies have also raised difficulties in testing lipreading ability since the individual difference is considerable for both deaf (Inceoglu, 2019) and hearing people (Sennott et al., 2011). In the study of Inceoglu (2019), the lipreading performance of deaf people showed high variability: accuracy rates ranged from 7% to 58% for vowel recognition, 0% to 44% for word recognition and 0% to 7% for sentence recognition. Moreover, a number of factors that may affect lipreading performance have been proposed, including age, gender, education, intelligence, visual memory and degree of hearing loss (Tye-Murray et al., 2016; Woodhouse et al., 2009).

Considering the indispensable place of lipreading in deaf communication, it is important to understand what makes a good lipreader. However, most of the studies on lipreading testing were conducted in the first language setting, which poses a potential problem: since most of the deaf people have rich experience, practice, and even systematic training in lipreading, one cannot determine the source of their good lipreading performance, is it from enhanced visual abilities or just extensive experience? Therefore, the current study considers evaluating lipreading of deaf and hearing people through learning efficiency in new language learning. The introducing of a new language can eliminate the effects of previous lipreading usage on deaf people to ensure more scientific results.

2.3 Eye Movement Research on Visual Speech Perception

People can unconsciously adjust their visual distribution pattern in speech perception to maximize access to information. For the speaker's face, previous research suggested that the eyes area corresponds to emotional and prosodic information, while the mouth area corresponds to phonetic and segmental information (Lansing & McConkie, 1999). This feature requires one to have a reasonable visual allocation in daily communication.

The visual attention varies at different stages of infant development. Infants younger than 8 months of age pay more attention to the talker's eyes than mouth (Smith et al., 2013). However, 8-month-old and 12-month-old infants care more about the talker's mouth than eyes (Pons et al., 2015). The transition from eyes to mouth has been reported starting at 6 months of age, reaching the peak at 10 months of age, and returning to eyes in adulthood (Lewkowicz & Hansen-Tift, 2012). The increased gaze on mouth may result from the huge demand for visual speech cues during the critical period of children's language development.

The development of visual attention in infants suggests that people's patterns of visual allocation may be task-determined. Many studies on adults also demonstrate the task-determined tendency: the mouth becomes important when the speech signal is degraded, or linguistic information is required.

It is worth noting that speech intelligibility has an impact on visual speech perception. Numerous studies have explored the role of noise in speech perception and found that people spent more time on talker's mouth than eyes as masking noise increase (Król, 2018). Particularly, a study by Vatikiotis-Bateson et al. (1998) showed that the percentage of people's looking at the mouths changed from 35% to 50% with noise level ranging from zero to medium, but proportion would no longer change with noise level after reaching about 60%.

Also, language familiarity affects people's visual assignments in speech perception. Barenholtz et al. (2016) investigated people's selective attention in a speech coding task and found that subjects were more concerned with the speaker's mouth when confronted with unfamiliar language than when confronted with familiar language. In contrast, similar results were not obtained in the listening task.

Until now, little attention has been paid to the visual distribution of the deaf in speech perception. Wang et al. (2020) investigated the visual distribution of deaf adults in native and non-native languages and found that deaf participants' fixations on mouth were consistently greater than 50 percent of speaker's face, which was significantly different from that of hearing participants (10%); and the deaf participants did not change their gaze patterns in native and non-native language speech listening. These findings validate the dependence of deaf people on visual modality and prove that it is hard for deaf people to flexibly adjust their visual attention. Additionally, language learning is an important intermediate state between language communication (familiar or native language) and language contact (unfamiliar or non-native language). However, the patterns adopted by deaf people in language learning remain unclear. Therefore, it is necessary to explore their characteristic of visual distribution in language learning tasks.

Chapter 3

Methodology

3.1 Research Design

- **Research Philosophy: Positivism**
- **Research Approach: Deductive Approach and Quantitative**
- **Research Strategy: Experimental Research**

The study is underpinned by positivism research philosophy to conduct a scientific psychological and linguistic research. In accordance with the positivism research philosophy, the research approach adopted for this study is deductive and quantitative. The purpose of this study is to evaluate lipreading ability of deaf people through visual speech learning efficiency, in order to provide evidence for the Perceptual Compensation Hypothesis. Therefore, the philosophical choice is reasonable and appropriate.

Experiment:

To answer the research questions, a empirical experiment is applied with 2 (Hearing Group) × 2 (Modality) × 2 (Content) × 7 (Session) mixed factorial design.

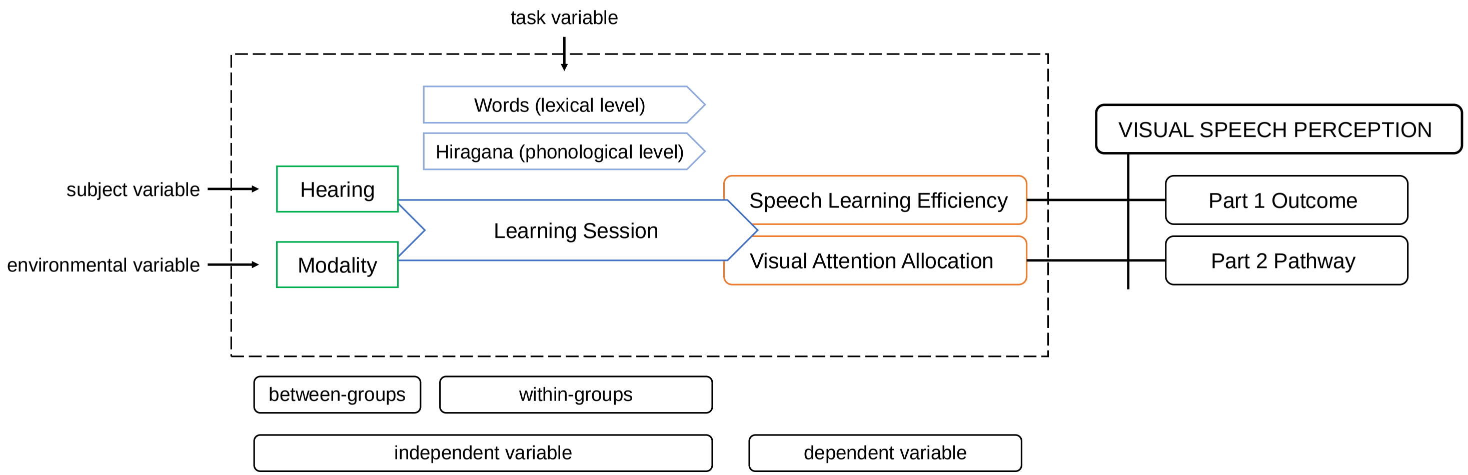

Figure 3.1: Framework of Experiment Design

The task of the subjects is to obtain the pronunciation of Japanese kana or words by repeatedly watching the silent videos in training block, and can restore the meaning it represents through the articulatory gestures in the silent videos in testing block. The subjects' lipreading abilities are quantified as the language learning efficiencies from silent videos. The framework of experiment design are shown in Figure 3.1 and each variable is described in detail below.

There are three main independent variables in this study:

- Hearing (subject variable): to compare the difference between normal hearing participants (NH) and prelingual deaf participants (PD).
- Modality (environmental variable): to compare the performance between visual speech perception and audiovisual speech perception.
- Training Content (task variable): to compare the learning efficiencies between syllable learning and word learning.

The two dependent variables correspond to two research questions:

- Speech Learning Efficiency: the response accuracies are recorded in multiple testing as an indicator of lipreading performance.
- Visual Attention Allocation: eye tracking data are collected to explore the visual distribution in speech learning.

Effectiveness:

To our knowledge, all proposed lipreading testing methods employ silent videos of sound segments, words and sentences ***in native speech*** as testing materials. Compared with previous studies, this study selects videos from ***inexperienced***

non-native speech, and the lipreading abilities of individuals and groups are assessed and compared by learning efficiencies in visual-only (VO) and audiovisual (AV) speech learning tasks.

Our method has two advantages: First, the introducing of a new language can eliminate the effects of previous lipreading usage on deaf people to ensure more scientific results. Second, because lipreading is an acquired ability of human being rather than a talent, the focus of lipreading should be to examine people's ability to use visual information to get new messages and new knowledge. Recognition accuracy of lipreading only applies to specific stimuli and specific languages. Therefore, the ***speech learning efficiency (process)*** is a more reasonable indicator of lipreading testing than ***speech recognition rate (result)***: high learning efficiency corresponds to high information acquisition within a certain period of time, suggesting high lipreading ability.

Limitation:

The major limitation of this study is that there is no good resolution for the large background differences among deaf people. Many factors are thought to influence the speech ability and language development of deaf people: different ages of hearing loss, different degrees of hearing loss, different selections of Hearing Aids (HA) or Cochlear Implants (CI), and different coding strategies of hearing device. Although the impact from these factors is unavoidable, this study selects suitable subjects according to strict criteria. Furthermore, these factors were not the main sources of influence on lipreading. Therefore, we believe that the proposed method is still effective in measuring the lipreading levels.

3.2 Research Hypothesis

Based on the research questions and experimental design, the following two research hypotheses are proposed:

Hypothesis 1: If hearing impaired people have higher learning efficiency in visual speech learning than normal hearing people, it can be concluded that they have enhanced lipreading ability.

Hypothesis 2: Hearing impaired people will spend more gaze time on the speakers' mouth in speech learning considering their special visual language experience.

Chapter 4

Experiment

4.1 Participants

Thirty-six participants were recruited in this experiment: 17 normal hearing adults (8 females and 9 males, mean age ± SD: 20.8±2.3, age range: 19-26) and 19 prelingually deaf adults (11 females and 8 males, mean age ± SD: 22.4±1.7, age range: 19-26). All deaf participants use Chinese sign language as their first language and speak Standard Chinese as their second language and all normal hearing participants speak Standard Chinese as their first language.

The deafness-related characteristics of the deaf participants were shown in Table 4.1. All of them were congenital deaf (n = 4) or lost their hearing before 36 months of age (n = 15), with a symmetrical bilateral profound hearing loss (>90 dB). Besides, they started wearing hearing aids (HAs) or receiving cochlear implants (CIs) at an early age.

Overall, all participants had no language experience of Japanese and had normal or corrected-to-normal vision with glasses. None of them reported any other speech or hearing disorders. And all of them signed the informed consent and were paid for their participation.

Table 4.1: Demographic information and deafness-related characteristics of the deaf participants

ID	Sex	Age (yrs)	Modality	Stimulation (L + R)	Unaided PTA (dB)	Age of hearing loss (month)	Age at device using (month)	Frequency of language(%): Spoken:Sign
101	M	23	VO	HA + /	120	17	32	5:95
102	M	22	VO	HA + HA	94	6	24	95:5
103	M	25	VO	/ + CI	120	24	32	50:50
104	F	21	VO	HA + /	102.5	0	24	80:20
105	F	22	VO	/ + HA	105	28	60	25:75
106	F	21	VO	HA + HA	96	8	68	80:20
107	F	21	VO	HA + HA	115	21	62	20:80
108	F	22	VO	HA + CI	110	36	39	30:70
109	M	19	VO	HA + CI	113	27	38	100:0
201	F	20	AV	HA + HA	115	31	31	70:30
202	F	22	AV	HA + HA	95	8	24	100:0
203	F	22	AV	HA + /	102.5	10	27	65:35
204	F	24	AV	HA + HA	102.5	25	25	45:55
205	F	22	AV	HA + HA	95	30	40	80:20
206	F	22	AV	HA + HA	97.5	1	53	50:50
207	M	26	AV	HA + HA	110	24	60	90:10
208	M	25	AV	/ + CI	107.5	0	14	90:10
209	M	24	AV	HA + HA	100	0	24	50:50
210	M	23	AV	/ + CI	105	0	62	5:95

Participants were randomized into two modality groups: Visual Only (VO) and Audio-Visual (AV). Stimuli in the VO modality were silent videos without any sound. Particularly, deaf people in AV group need to using hearing aids or cochlear implants, while deaf people in VO group need NOT to using hearing aids or cochlear implants.

4.2 Stimuli and Apparatus

4.2.1 Materials

There were two sets of materials for training and testing: Japanese Hiragana and Japanese words. The two stimuli contents required different levels of speech processing or speech learning: participants were expected to establish a link between visual speech cues and the orthographic representation (Speech - Symbol mapping, phonological level) in Japanese Hiragana conditions; and participants were expected to establish a link between visual speech cues and the conceptual meaning (Speech - Meaning mapping, lexical level) in Japanese words conditions.

Japanese Kana

Fourty Japanese Hiragana (hereafter called kana) characters, corresponding to 40 Japanese monosyllables in consonant-vowel (CV) format, were selected from Japanese phonological system. They were 5×8 combinations: 5 columns (/a/, /i/ /u/, /e/ /o/) and 8 rows (/Ø/, /k/, /s/, /t/, /n/, /h/, /m/, /r/).

Japanese Word

Fourty Japanese words were chosen from BCCWJ (Maekawa et al., 2014). All of them were two-syllable (mora) in Japanese and can be represented by a single

Chinese character in Chinese. The selected words were all high-frequency nouns with concrete concepts. The psycholinguistic parameters of these nouns were as follows: Concreteness (CNC): 596 ± 20, Imageability (IMG): 597 ± 29, Familiarity (FAM): 575 ± 34, which from Abstract Conceptual Feature (ACF) (Crutch et al., 2013) and MRC Psycholinguistic Database (Wilson, 1988).

4.2.2 Recording

Stimuli (video and audio) were recorded by a Sony video camera in a soundproof booth in mp4 format (1920 × 1080 pixels resolution, 60 Hz frame rate and 48,000 Hz sampling rate for audio). The video had a white background color and the speakers' faces were in the center of the video frame. After recording, the video was split into 3000ms video clips, each video clip for one stimuli token.

Five native speakers of Japanese (2 females and 3 males) were recruited as stimuli speakers. High Variability Phonetic Training (HVPT) (Barriuso & Hayes-Harb, 2018) was adopted in the current experiment: 3 speakers (1 female and 2 males) only appeared in training and 2 speakers (1 female and 1 male) only appeared in testing.

4.2.3 Apparatus

Stimuli were presented on a 24-inch monitor with a resolution of 1920 × 1080 pixels and a refresh rate of 60 Hz. Sounds were played by an ISK HP-800 monitoring headphone. Participants' eye movement was recorded with EyeLink 1000 Plus (SR Research, 1000 Hz sampling rate) or Tobii Pro Nano (Tobii Technology, 60 Hz sampling rate) during the experiment. And nine-point calibration was conducted prior to each experiment.

4.3 Procedure

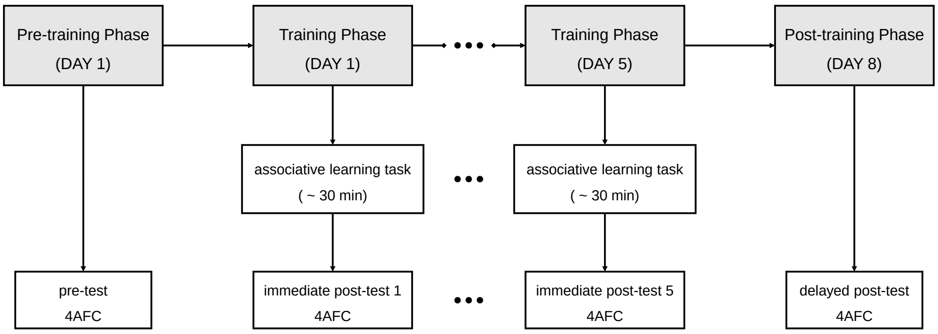

Figure 4.1: The flow of experimental protocol

It took 8 days for one experiment block, which included 5 days of training, one pre-test before training, five immediate post-tests after daily training and one delayed post-test 72 hours after finishing all training. The experiment took a total of 16 days for two blocks.

4.3.1 Training: Associative Learning Task

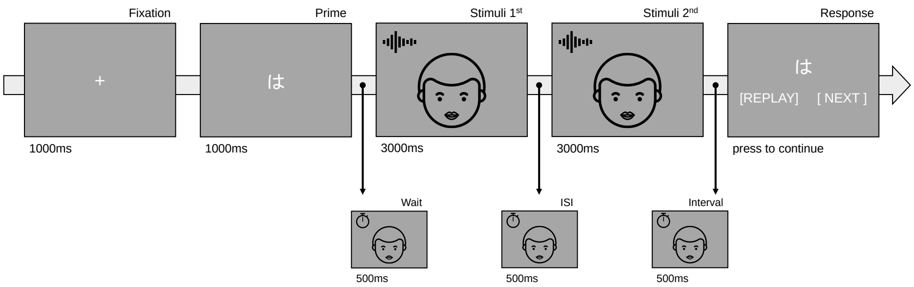

Figure 4.2: The sketch of one training trial

First, a cross was presented in the center of the screen followed by the target kana or word as priming. Then the video of the target kana or word was played, and the same video would be played twice with 500-ms interstimulus interval (ISI). After the video played, the kana or word would appear again on the monitor. Participants were instructed to ‘replay this item’ or ‘turn to the next item’ by pressing the corresponding key. A total of 80 trails (40 Stimuli × 2 Speaker) were presented to each participant in sequential order.

4.3.2 Testing: Four-alternative Forced Choice (4AFC)

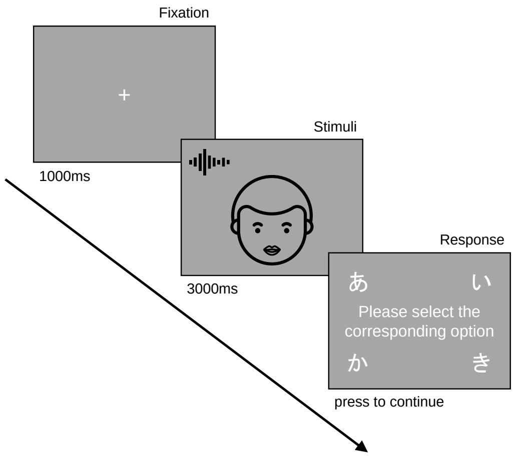

Figure 4.3: The sketch of one testing trial

In this task, a cross appeared in the center of the screen. Then videos were presented to participants followed by four options. Participants were asked to choose the one corresponding to the previous video by pressing the corresponding button. All 40 target stimuli were randomly presented to each participant.

4.4 Data analysis

4.4.1 Behavioral data

A generalized linear mixed-effects model was fitted to the whole testing data, with response (1 = 'correct', 0 = 'incorrect') as dependent variable, **hearing** (categorical predictor, sum coding: -1 = 'HN', 1 = 'PD'), **modality** (categorical predictor, sum coding: -1 = 'AV', 1 = 'VO'), **content** (categorical predictor, sum coding: -1 = 'kana', 1 = 'word') and **session** (numerical predictor, 0 = 'pre-test', 1 to 5 = 'immediate post-test', 8 = 'delayed post-test') as independent variable, and the model also included random intercepts and random slopes for *participant* (36 levels), *stimulus* (80 levels) and *speaker* (2 levels).

4.4.2 Eye tracking data

There were several steps of preprocessing for eye tracking data before statistical analysis:

- Re-sampling: all data from different eye tracking devices were downsampled to 60Hz.
- Data cleaning: the threshold of track-loss frames was set to 25%: trials whose trackloss proportion is greater than 25% will be excluded. A total of 9576 trails of training (28.40% of 33723 trails) and 4029 trails of testing (19.99% of 20160 trails) were removed in this step.
- Period of interest (POI) setting: the time window was set to the period when stimulus appears on the screen: 6500 ms for testing and 3000 ms for testing.
- Area of interest (AOI) setting: three AOIs were defined in this experiment (face, mouth, and eyes) for each speaker, then calculated the proportion of

the dwell time on eyes, mouth and face to the overall dwell time, respectively. Then all proportions were transformed into rationalized arcsine unit (RAU) for statistical purpose.

A linear mixed-effects models was fitted to the eyetracking data, with adjusted fixation proportion as dependent variable (RAU ranging from -23.0 to 123.0 corresponding to Proportion from 0% to 100%), **hearing** (categorical predictor, sum coding: -1 = 'HN', 1 = 'PD'), **modality** (categorical predictor, sum coding: -1 = 'AV', 1 = 'VO'), **area** (categorical predictor, sum coding: -1 = 'eyes', 1 = 'mouth'), **task** (categorical predictor, sum coding: -1 = 'Learning', 1 = 'Testing') and **session** (numerical predictor, 0 = 'pre-test', 1 to 5 = 'immediate post-test', 8 = 'delayed post-test') as independent variable, and the model also included random intercepts and random slopes for *participant* (36 levels), *stimulus* (80 levels) and *speaker* (5 levels).

4.4.3 Statistical package

Statistical analysis was performed in R 4.2.0 (R Core Team, 2022). Generalized linear mixed model was fitted using glmer() function and linear mixed-effects model was fitted using lmer() function in lme4 package (Bates et al., 2015) and lmertest package (Kuznetsova et al., 2017). Akaike Information Criterion (AIC) was used in model comparison to ensure all models mentioned above were the optimal models. The model with mixed-effects were be chosen if random factor showed improvement to the model. Post Hoc comparisons or pairwise comparisons were conducted using emmeans package (Lenth, 2022), with degrees-of-freedom derived using Kenward-Roger method and p-values adjusted using Tukey method. Eye tracking data were processed using some functions from eyetrackingR package (Dink & Ferguson, 2015).

Chapter 5

Results

5.1 Learning Efficiency in Visual Speech Perception

The learning curves for different groups and different contents were displayed in Figure 5.1. One could see that: first, the response accuracy increased with training sessions, all four groups made great progress after training, and the training effects were well retained in the delayed post-test; second, AV participants had higher response accuracy than VO participants, and the accuracy of deaf group was slightly higher than their hearing counterparts in only lipreading conditions (VO); third, particularly for NH + AV group, they showed extreme higher learning efficiency than other groups.

Results of generalized linear mixed-effects model showed significant main effect of Session (β = 0.585, SE = 0.016, z = 37.141, p <0.0001) on participants' responses, reflecting participants' improved performance with training. However, the main effects of Hearing (β = 0.016, SE = 0.068, z = 0.238, p = 0.812), Modality (β = -0.034, SE = 0.068, z = -0.501, p = 0.616) and Content (β = 0.059, SE = 0.071, z = 0.832, p = 0.405) did not reach significance, indicating that the better hearing conditions and multiple modalities did not enhance learning efficiency,

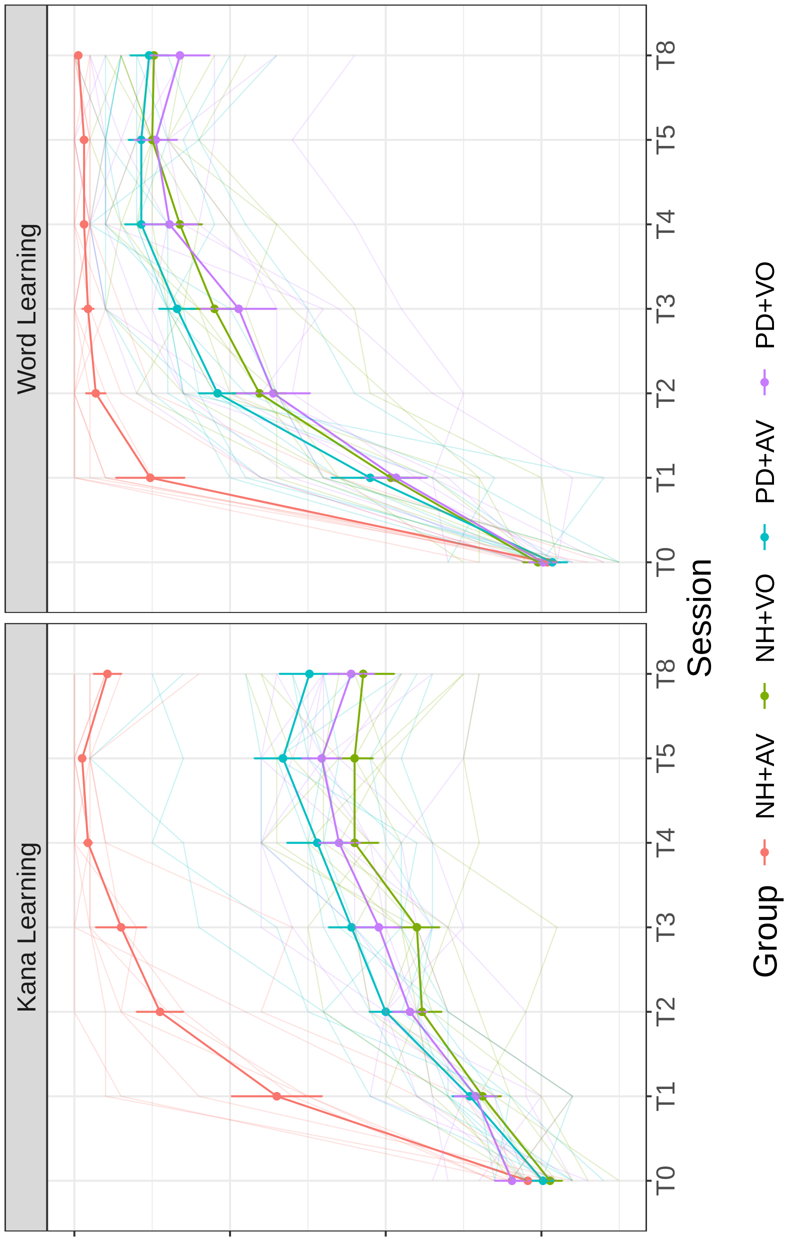

Figure 5.1: The response accuracies for different groups over session (bold lines represent learning curves of different groups and different modalities; thin lines represent learning curves of different individuals)

Table 5.1: Summary of generalized linear mixed-effects model

	Estimate	**Std. Error**	**z**	**p**
(Intercept)	−0.561	0.102	−5.5	**<.0001**
Hearing	0.016	0.068	0.2	0.812
Modality	−0.034	0.068	−0.5	0.616
Content	0.059	0.071	0.8	0.405
Session	0.585	0.016	37.1	**<.0001**
Hearing:Modality	0.035	0.068	0.5	0.606
Hearing:Content	0.034	0.029	1.2	0.236
Modality:Content	0.052	0.029	1.8	0.075
Hearing:Session	−0.248	0.016	−15.9	**<.0001**
Modality:Session	−0.275	0.016	−17.7	**<.0001**
Content:Session	0.219	0.016	14.1	**<.0001**
Hearing:Modality:Content	−0.077	0.029	−2.6	**<.01**
Hearing:Modality:Session	0.235	0.016	15.1	**<.0001**
Hearing:Content:Session	−0.069	0.015	−4.5	**<.0001**
Modality:Content:Session	−0.071	0.015	−4.6	**<.0001**
Hearing:Modality:Content:Session	0.059	0.015	3.9	**<.001**

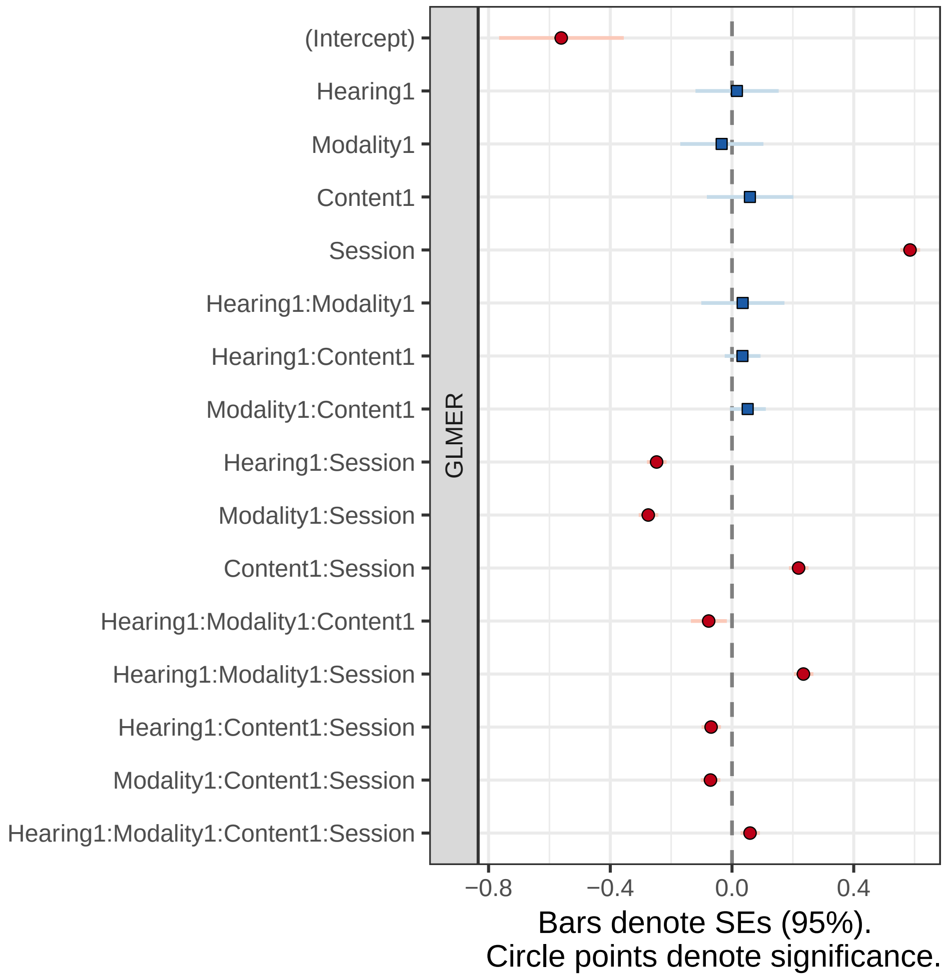

Figure 5.2: Estimates of generalized linear mixed-effects model

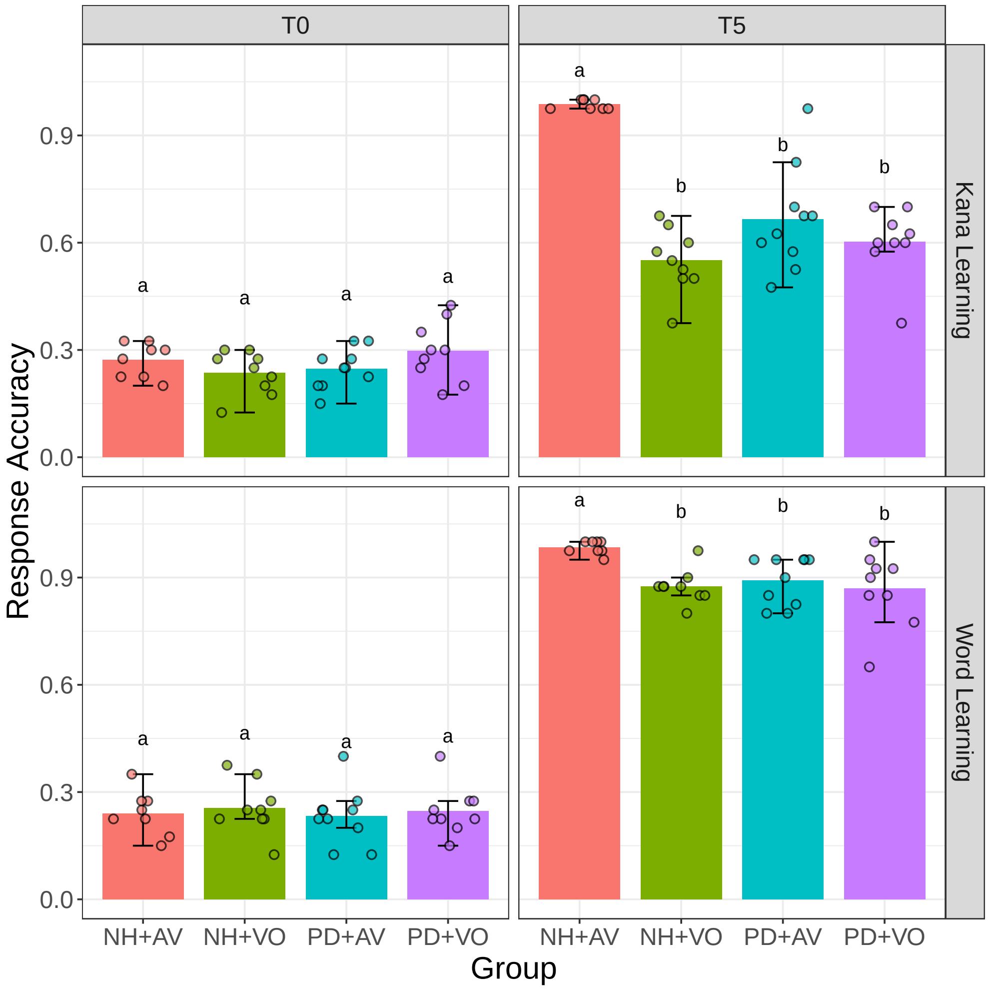

Figure 5.3: The simple effect of Hearing: Modality on Session: Content

and this tendency remained unchanged between kana learning and word learning.

The two-way interactions were only significant when Session was included (Hearing * Session: β = -0.248, SE = 0.016, z = -15.916, p <0.0001; Modality * Session: β = -0.275, SE = 0.016, z = -17.667, p <0.0001; Content * Session: β = 0.219, SE = 0.016, z = 14.067, p <0.0001).

Nested model comparison showed that the random factors significantly improved the modal fit. Among the three random factors, the *stimulus* explained 33.21% of the variance and 57.62% of the deviance; the *participant* explained 13.47% of the variance and 36.71% of the deviance; and the *speaker* explained 0.31% of the variance and 5.5% of the deviance.

In line with what we observed in the learning curves, simple effect results (see Figure 5.3) suggest that: at the beginning, the four groups were at the same level. But after training, the NH + AV group showed significantly higher accuracy than the other groups, while accuracies between all the other three groups did not show significance.

5.2 Visual Allocation in Visual Speech Perception

Figure 5.4 and 5.5 showed the mean fixation proportions on the mouth AOI and the eyes AOI. One can see that both normal hearing participants and prelingually deaf participants had greater overall fixation on mouth AOI than eyes AOI and greater overall fixation in audiovisual modality than in visual-only modality. The proportions of fixation on eyes were 12.7%, 10.8%, 17.7% and 12.4% and on mouth were 51.6%, 58.8%, 44.6% and 54.3%, respectively for normal hearing in audio-visual, normal hearing in visual only, prelingually deafness in audio-visual, prelingually deafness in visual only.

Results of linear mixed-effects model showed significant main effects of Area

(β = 24.940, SE = 0.174, t = 143.416, p <0.0001), Modality (β = 1.970, SE = 0.761, t = 2.588, p <0.05) and Session (β = 0.212, SE = 0.050, t = 4.267, p <0.0001) on fixation proportion. However, the main effects of Hearing (β = 0.184, SE = 0.761, t = 0.242, p = 0.81) and Task (β = 1.935, SE = 2.472, t = 0.783, p = 0.491) did not reach significance. From these results, it can be concluded that people spend more time on speakers' mouths than on speakers' eyes in speech learning; different modalities have impact on people's attention distribution, while different hearing conditions have no impact on people's attention distribution.

Since the interactions were significant among Hearing, Modality, Task and Area (β = -0.280, SE = 0.049, t = 5.741, p = <0.0001), a battery of simple effect analysis was also applied and the results were displayed in Figure 5.4 and Figure 5.5. From the above results, two points could be summarized as follows:

First, both normal hearing participants and prelingually deaf participants spend more time on mouth in visual-only modality than in audiovisual modality, which indicated that the absence of auditory information increased subjects' attention to the speakers' mouth. Moreover, all differences between different modalities reached significance except NH+VO in learning. Conversely, both normal hearing participants and prelingually deaf participants spend less time on eyes in visual-only modality than in audiovisual modality. The exception is NH+VO in learning, in which the ratio of VO was slightly higher than that of AV.

Second, normal hearing participants spend more time on mouth than prelingually deaf participants in both visual-only modality and audiovisual modality, which indicated that normal hearing people pay more attention to visual speech information in speech learning. In contrast, prelingually deaf participants spend more time on eyes than normal hearing participants.

Table 5.2: Summary of linear mixed-effects model

	Estimate	**Std. Error**	**t**	**p**
(Intercept)	26.862	2.583	10.398	**<.001**
Hearing	0.184	0.761	0.242	0.81
Modality	1.970	0.761	2.588	**<.05**
Session	0.212	0.050	4.267	**<.0001**
Area	24.940	0.174	143.416	**<.0001**
Task	1.935	2.472	0.783	0.491
Hearing:Modality	0.889	0.761	1.168	0.251
Hearing:Session	−0.111	0.049	−2.253	**<.05**
Modality:Session	−0.225	0.049	−4.583	**<.0001**
Hearing:Area	−2.712	0.174	−15.595	**<.0001**
Modality:Area	3.579	0.174	20.580	**<.0001**
Session:Area	−0.516	0.049	−10.592	**<.0001**
Hearing:Task	0.048	0.175	0.275	0.783
Modality:Task	−1.185	0.175	−6.778	**<.0001**
Session:Task	0.032	0.050	0.651	0.515
Area:Task	−1.089	0.174	−6.261	**<.0001**
Hearing:Modality:Session	−0.145	0.049	−2.950	**<.01**
Hearing:Modality:Area	−0.241	0.174	−1.388	0.165
Hearing:Session:Area	0.021	0.049	0.423	0.672
Modality:Session:Area	−0.180	0.049	−3.701	**<.001**
Hearing:Modality:Task	−0.518	0.175	−2.965	**<.01**
Hearing:Session:Task	0.082	0.049	1.671	0.095
Modality:Session:Task	0.203	0.049	4.137	**<.0001**
Hearing:Area:Task	0.637	0.174	3.661	**<.001**
Modality:Area:Task	1.227	0.174	7.053	**<.0001**
Session:Area:Task	0.876	0.049	17.965	**<.0001**
Hearing:Modality:Session:Area	0.393	0.049	8.062	**<.0001**
Hearing:Modality:Session:Task	0.297	0.049	6.047	**<.0001**
Hearing:Modality:Area:Task	−0.029	0.174	−0.166	0.868
Hearing:Session:Area:Task	−0.033	0.049	−0.669	0.504
Modality:Session:Area:Task	0.256	0.049	5.258	**<.0001**
Hearing:Modality:Session:Area:Task	−0.280	0.049	−5.741	**<.0001**

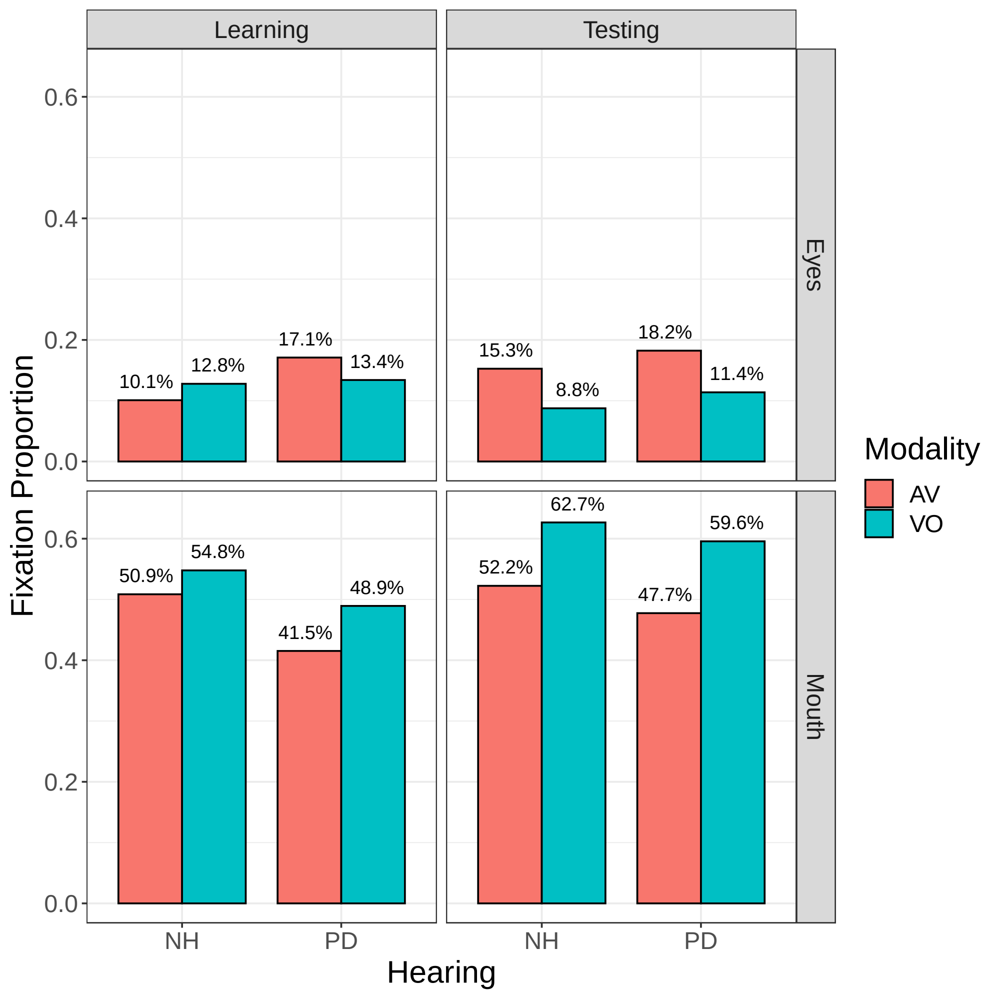

Figure 5.4: Fixation proportions for different areas: between modalities

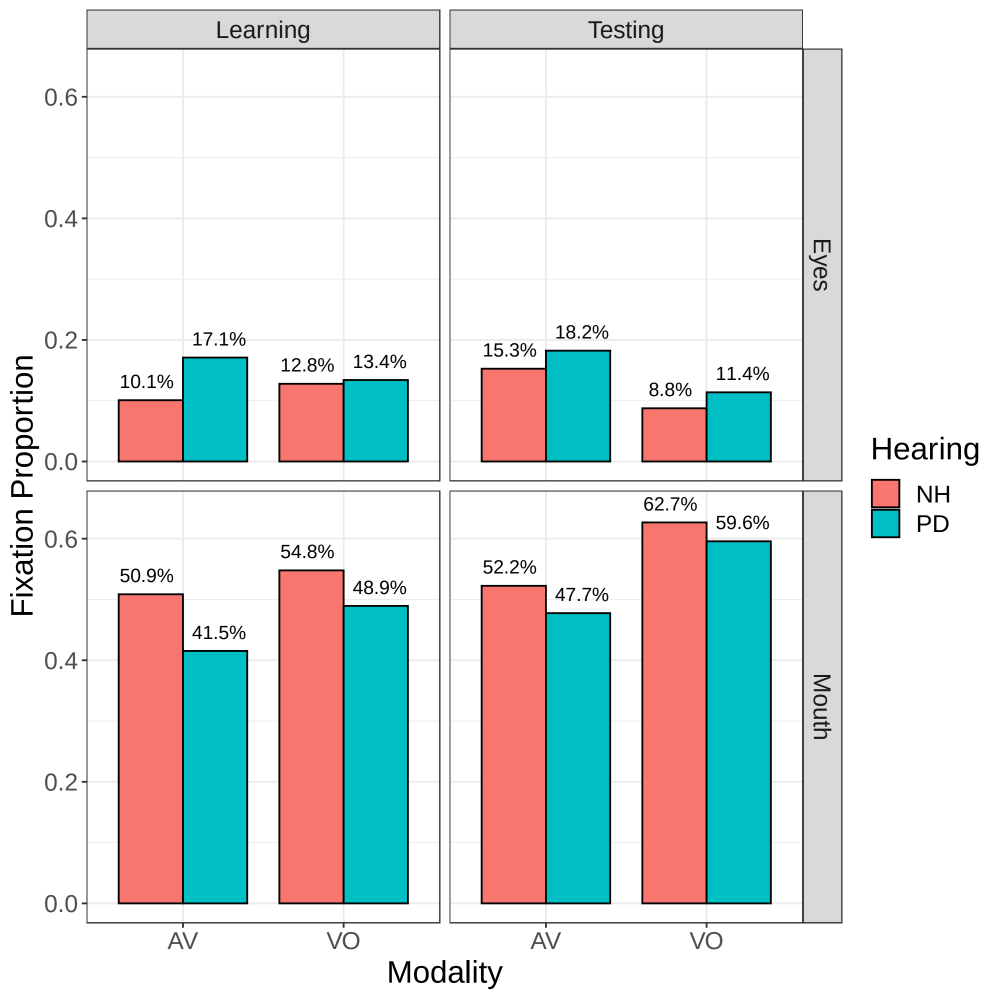

Figure 5.5: Fixation proportions for different areas: between groups

Chapter 6

Discussion

6.1 Lipreading Performance

The most surprising finding of this study is that there is no significant difference between deaf and hearing adults in lipreading learning efficiency. This finding is consistent with previous research showing similar lipreading performance for deaf and hearing children (Kyle et al., 2013), but contradicts the result of speechreading advantages of deaf adults (Pimperton et al., 2017). The following two points may support our inference and explain the opposite conclusion on lipreading ability:

First, the testing languages are different between this study and Pimperton et al. (2017): the target words from Pimperton et al. (2017) are selected from the subjects' native language, while materials of our study are non-native syllables and words. Deaf adults have extensive visual language experience (from daily conversation) and systematic lipreading training (from speech rehabilitation therapy) compared to hearing adults. These give deaf people an unbalanced advantage in lipreading assessment. Therefore, if the lipreading experience and training are language-specific, the lipreading advantage of the deaf will not be available in the face of non-native speech.

Second, both deaf and hearing people improve their lipreading accuracy through training. In contrast to some researchers regarding lipreading as an inborn talent to explain individual variation in lipreading performance, our results show that lipreading can be acquired for both deaf and hearing people. Therefore, deaf people will not have better lipreading than hearing people without additional training.

In summary, this finding provides some evidence against the perceptual compensation hypothesis. Hearing deprivation does not enhance people's visual perception. Of course, deaf people have no defect in lipreading ability compared to hearing people. Our research claims that training and experience can greatly improve lipreading performance of deaf people to facilitate their communication.

In addition, this study confirms the indispensable role of visual input in speech perception of deaf people. Our resluts show that the visual modality can contribute 71.3% of speech information to hearing people (AV: 98.6%, VO: 71.3%), but can contribute up to 94.5% to deaf people (AV: 77.9%, VO: 73.6%). Although hearing aids and cochlear implants provide comparable auditory signal input, deaf people still have great visual dependence in speech perception.

6.2 Visual Attention Allocation

The current study found that all participants increased their gaze time on mouth in VO conditions than in AV conditions (+ 10.5% for hearing and + 11.9% for deaf). This result is in line with previous studies that degrade of speech signal increases people's gaze on mouth (Vatikiotis-Bateson et al., 1998). In our experiments, the VO modality is equivalent to the environment where noise masks speech completely, which creates huge demand for speech cues from the mouth and its surroundings. Similar to previous findings, this study also showed a ceiling for the attention to the mouth: the proportion of the highest group was about

60%. The ceiling may come from the most important limitation of lipreading: not all phonemes can be perceived through visible articulators. Some phonemes are easily confused for their high visual similarity, for example, /b/, /p/ and /m/ are visually indistinguishable; some phonemes have variable articulation positions due to coarticulation, such as velar consonants. Fernandez-Lopez et al. (2017) found that humans can decode 44% of word and sentence units and 52.20% of phoneme unit in visual speech recognition; in contrast, the recognition rates of automatic recognition machine are 20% and 51.25% respectively. Therefore, it is possible that the upper limit of lipreading usability determines the upper boundary of visual fixation to the mouth.

Furthermore, this study suggests that deaf and hearing people adopt similar visual allocation strategies in speech learning: both groups look more at speakers' mouth than at speakers' eyes. Correspondingly, a study from Wang et al. (2020) proved that deaf people mainly look at talkers' eyes and hearing people mainly look at talkers' mouth in audiovisual speech perception. Compared to speech perception (native language), hearing people have a large attention shift from eyes to mouth in speech learning (non-native language), while deaf people persistently maintain interest in mouth area. These results imply flexible selective attention for hearing people. However, deaf people just pay more attention to the mouth regardless of different tasks or different language familiarities. The possible reason is that hearing deprivation of deaf people altered their organization of visual modality. The long term hearing loss makes deaf people heavily dependent on speech information from the mouth and the visual dependence has not changed with hearing improvement.

Finally, hearing people even pay more attention to speakers' mouth than deaf people in both visual speech recognition (VO) and audiovisual speech recognition (AV). Despite the great discrepancy in visual dependence and extensive difference in speech experience, deaf people don not have more efficient visual allocation strategy during the speech perception. Our results confirm the view of Bavelier et al. (2006) and Shalev et al. (2020): there is no distribute attention advantage in central vision for deaf people.

6.3 Summary

As stated before, our research shows that deaf people rely on lipreading but are not good at lipreading, possibly because ***lipreading is language-specific and cannot be applied to other languages***; deaf people are dependent on visual modality but cannot make full use of visual information, possibly because ***hearing loss changes visual allocation and attention of deaf people***.

Chapter 7

Conclusion

7.1 Summary

Before summarizing the current study, we would like to give the answers to research questions:

To Q1: Do hearing impaired people have enhanced lipreading ability compared to normal hearing people in speech learning?

No, there is no significant difference in lipreading between deaf and hearing people. Although deaf people are not superior lipreaders, lipreading information is indispensable for speech perception of deaf people.

To Q2: What are the differences in visual attention strategies between hearing impaired people and normal hearing people in speech learning?

Both deaf and hearing people have increased mouth attention in visual speech perception than in audiovisual speech perception and spend more fixation time on the mouth than on the eyes. Hearing people have more flexible selective attention and spend more time on the mouth than deaf people. Deaf people are more dependent on visual information but are not good users of visual information.

This study investigated the lipreading performance and visual attention strategy of hearing impaired people in speech learning. Overall, deaf people have no better lipreading performance than hearing people. Both deaf and hearing people pay more attention to speakers' mouths than speakers' eyes when learning a new language. These results do not support the compensatory hypothesis that vision will increase naturally with hearing loss. Alternatively, it highlights the significant effect of training on lipreading performance. Therefore, clinical rehabilitation and deaf education should pay attention to lipreading training. Moreover, the selective attention deficit of deaf people requests attention directing with special gestures in face-to-face communication, and prominent signs or impressive colors in content design.

7.2 Contribution

The main contribution of the current study lies in the development of a new method for lipreading testing: compared with speech recognition accuracy, we use visual speech learning efficiency, which can exclude the influence of visual speech experience, as an indicator of lipreading ability. Correspondingly, this study found that deaf people did not have superior lipreading skills than hearing people. The results highlight the role of lipreading training in speech perception of deaf people and inspire the deaf community to engage in lipreading education.

7.3 Remaining Works

A remaining and possible work on the current topic is to compare lipreading performance in native speech and second language acquisition (SLA) of deaf people. Exploring correlations between native and non-native lipreading abilities will provide more powerful evidence for lipreading abilities of deaf people.

References

Barenholtz, E., Mavica, L., & Lewkowicz, D. J. (2016). Language familiarity modulates relative attention to the eyes and mouth of a talker. *Cognition*, *147*, 100–105.
10

Barriuso, T. A., & Hayes-Harb, R. (2018). High variability phonetic training as a bridge from research to practice. *CATESOL Journal*, *30*(1), 177–194.
19

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01
23

Bavelier, D., Dye, M. W., & Hauser, P. C. (2006). Do deaf individuals see better? *Trends in cognitive sciences*, *10*(11), 512–518.
2, 7, 37

Bavelier, D., & Neville, H. J. (2002). Cross-modal plasticity: where and how? *Nature Reviews Neuroscience*, *3*(6), 443–452.
2

Bottari, D., Nava, E., Ley, P., & Pavani, F. (2010). Enhanced reactivity to visual

stimuli in deaf individuals. *Restorative neurology and neuroscience*, *28*(2), 167–179.
7

Buckley, D., Codina, C., Bhardwaj, P., & Pascalis, O. (2010). Action video game players and deaf observers have larger goldmann visual fields. *Vision research*, *50*(5), 548–556.
7

Conway, C. M., Pisoni, D. B., & Kronenberger, W. G. (2009). The importance of sound for cognitive sequencing abilities: The auditory scaffolding hypothesis. *Current directions in psychological science*, *18*(5), 275–279.
6

Crutch, S. J., Troche, J., Reilly, J., & Ridgway, G. R. (2013). Abstract conceptual feature ratings: the role of emotion, magnitude, and other cognitive domains in the organization of abstract conceptual knowledge. *Frontiers in human neuroscience*, *7*, 186.
19

Dink, J. W., & Ferguson, B. (2015). *eyetrackingr: An r library for eye-tracking data analysis.* Retrieved from http://www.eyetrackingr.com
23

Dye, M. W., & Bavelier, D. (2010). Attentional enhancements and deficits in deaf populations: An integrative review. *Restorative Neurology and Neuroscience*, *28*(2), 181–192.
6

Fernandez-Lopez, A., Martinez, O., & Sukno, F. M. (2017). Towards estimating the upper bound of visual-speech recognition: The visual lip-reading feasibil-

ity database. In *2017 12th ieee international conference on automatic face & gesture recognition (fg 2017)* (pp. 208–215).
36

Inceoglu, S. (2019). Individual differences in l2 speech perception: The role of phonological memory and lipreading ability. *The Modern Language Journal*, *103*(4), 782–799.
8

Król, M. E. (2018). Auditory noise increases the allocation of attention to the mouth, and the eyes pay the price: An eye-tracking study. *Plos one*, *13*(3), e0194491.
10

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi: 10.18637/jss.v082.i13
23

Kyle, F. E., Campbell, R., Mohammed, T., Coleman, M., & MacSweeney, M. (2013). Speechreading development in deaf and hearing children: introducing the test of child speechreading. *Journal of Speech, Language, and Hearing Research*.
8, 34

Lansing, C. R., & McConkie, G. W. (1999). Attention to facial regions in segmental and prosodic visual speech perception tasks. *Journal of Speech, Language, and Hearing Research*, *42*(3), 526–539.
9

Lenth, R. V. (2022). emmeans: Estimated marginal means, aka least-squares means [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=emmeans (R package version 1.7.3)
23

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, *109*(5), 1431–1436.
9

Lusk, L. G., & Mitchel, A. D. (2016). Differential gaze patterns on eyes and mouth during audiovisual speech segmentation. *Frontiers in psychology*, 52.
1

Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., ... Den, Y. (2014). Balanced corpus of contemporary written japanese. *Language resources and evaluation*, *48*(2), 345–371.
18

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.
1

Mitchell, T. V., & Quittner, A. L. (1996). Multimethod study of attention and behavior problems in hearing-impaired children. *Journal of Clinical Child Psychology*, *25*(1), 83–96.
7

Pimperton, H., Ralph-Lewis, A., & MacSweeney, M. (2017). Speechreading in deaf adults with cochlear implants: Evidence for perceptual compensation. *Frontiers in psychology*, *8*, 106.

8, 34

Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. *Psychological science*, *26*(4), 490–498.

9

Poon, B. T., & Jenstad, L. M. (2022). Communication with face masks during the covid-19 pandemic for adults with hearing loss. *Cognitive Research: Principles and Implications*, *7*(1), 1–18.

1

Qin, Z., & Yan, G. L. (2021). The theoretical debate on the effect of hearing impairment on visual performance of deaf individuals: Deficit or compensation? *Journal of Psychological Science*, *44*(5), 1266-1272.

6

R Core Team. (2022). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

23

Ronnberg, J. (1995). Perceptual compensation in the deaf and blind: Myth or reality. *Compensating for psychological deficits and declines: Managing losses and promoting gains*, 251–274.

2

Saunders, G. H., Jackson, I. R., & Visram, A. S. (2021). Impacts of face coverings on communication: an indirect impact of covid-19. *International Journal of Audiology*, *60*(7), 495–506.

1

Sennott, B., Straehley, I., Mork, D., Coppalle, R., & Loebach, J. L. (2011). Lipreading ability in normal hearing listeners. *Unpublished poster presentation*.
8

Shalev, T., Schwartz, S., Miller, P., & Hadad, B.-S. (2020). Do deaf individuals have better visual skills in the periphery? evidence from processing facial attributes. *Visual Cognition*, *28*(3), 205–217.
2, 37

Sidera, F., Amadó, A., & Martínez, L. (2016, 12). Influences on Facial Emotion Recognition in Deaf Children. *The Journal of Deaf Studies and Deaf Education*, *22*(2), 164-177. Retrieved from https://doi.org/10.1093/deafed/enw072 doi: 10.1093/deafed/enw072
7

Smith, N. A., Gibilisco, C. R., Meisinger, R. E., & Hankey, M. (2013). Asymmetry in infants' selective attention to facial features during visual processing of infant-directed speech. *Frontiers in Psychology*, *4*, 601.
9

Tye-Murray, N., Sommers, M. S., & Spehar, B. (2007). Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing. *Ear and hearing*, *28*(5), 656–668.
8

Tye-Murray, N., Spehar, B., Myerson, J., Hale, S., & Sommers, M. (2016). Lipreading and audiovisual speech recognition across the adult lifespan: Implications for audiovisual integration. *Psychology and aging*, *31*(4), 380.
8

Vatikiotis-Bateson, E., Eigsti, I.-M., Yano, S., & Munhall, K. G. (1998). Eye movement of perceivers during audiovisualspeech perception. *Perception & psychophysics*, *60*(6), 926–940.
10, 35

Wang, J., Zhu, Y., Chen, Y., Mamat, A., Yu, M., Zhang, J., & Dang, J. (2020). An eye-tracking study on audiovisual speech perception strategies adopted by normal-hearing and deaf adults under different language familiarities. *Journal of Speech, Language, and Hearing Research*, *63*(7), 2245–2254.
10, 36

Wilson, M. (1988). Mrc psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior research methods, instruments, & computers*, *20*(1), 6–10.
19

Woodhouse, L., Hickson, L., & Dodd, B. (2009). Review of visual speech perception by hearing and hearing-impaired people: clinical implications. *International Journal of Language & Communication Disorders*, *44*(3), 253–270.
8

Worster, E., Pimperton, H., Ralph-Lewis, A., Monroy, L., Hulme, C., & MacSweeney, M. (2018). Eye movements during visual speech perception in deaf and hearing children. *Language Learning*, *68*, 159–179.
8