

Comparison of Gated Audiovisual Speech Identification in Elderly Hearing Aid Users and Elderly Normal-Hearing Individuals: Effects of Adding Visual Cues to Auditory Speech Stimuli

Trends in Hearing
2016, Vol. 20: 1–15
© The Author(s) 2016
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2331216516653355
tia.sagepub.com


Shahram Moradi¹, Björn Lidestam², and Jerker Rönnerberg¹

Abstract

The present study compared elderly hearing aid (EHA) users ($n = 20$) with elderly normal-hearing (ENH) listeners ($n = 20$) in terms of isolation points (IPs, the shortest time required for correct identification of a speech stimulus) and accuracy of audiovisual gated speech stimuli (consonants, words, and final words in highly and less predictable sentences) presented in silence. In addition, we compared the IPs of audiovisual speech stimuli from the present study with auditory ones extracted from a previous study, to determine the impact of the addition of visual cues. Both participant groups achieved ceiling levels in terms of accuracy in the audiovisual identification of gated speech stimuli; however, the EHA group needed longer IPs for the audiovisual identification of consonants and words. The benefit of adding visual cues to auditory speech stimuli was more evident in the EHA group, as audiovisual presentation significantly shortened the IPs for consonants, words, and final words in less predictable sentences; in the ENH group, audiovisual presentation only shortened the IPs for consonants and words. In conclusion, although the audiovisual benefit was greater for EHA group, this group had inferior performance compared with the ENH group in terms of IPs when supportive semantic context was lacking. Consequently, EHA users needed the initial part of the audiovisual speech signal to be longer than did their counterparts with normal hearing to reach the same level of accuracy in the absence of a semantic context.

Keywords

audiovisual speech perception, EHA users, ENH listeners, gating paradigm

Date received: 18 December 2015; revised: 12 May 2016; accepted: 12 May 2016

Introduction

In daily face-to-face conversation, listeners benefit from combined auditory and visual speech signals that facilitate the identification of speech stimuli in comparison with auditory-only or visual-only presentation (Erber, 1969; Sumbly & Pollack, 1954). The audiovisual presentation of speech stimuli is particularly important for hearing-impaired individuals, who, even when using their hearing aids, have greater difficulties in perceiving auditory speech stimuli compared with normal-hearing listeners (Dimitrijevic, John, & Picton, 2004; Moradi, Lidestam, Hällgren, & Rönnerberg, 2014a). Walden, Grant, and Cord (2001) reported that the addition of visual cues to amplified auditory signals by hearing aids resulted in better identification of speech stimuli relative to unaided audiovisual or aided auditory-only

conditions. An important question that remains unexplored is whether hearing aid users have the same level of ability for audiovisual speech recognition as their age-matched normal-hearing counterparts.

A few studies have attempted to compare the audiovisual speech abilities of hearing-impaired and normal-hearing listeners; all were conducted under *unaided*

¹Linnaeus Centre HEAD, Department of Behavioral Sciences and Learning, Linköping University, Sweden

²Department of Behavioral Sciences and Learning, Linköping University, Linköping, Sweden

Corresponding author:

Shahram Moradi, Department of Behavioral Sciences and Learning, Linköping University, SE-581 83 Linköping, Sweden.
Email: shahram.moradi@liu.se



conditions, in which the auditory component of audiovisual stimuli was delivered to the ear(s) of listeners (Baskent & Bazo, 2011; Bernstein & Grant, 2009; Tye-Murray, Sommers, & Spehar, 2007a). Bernstein and Grant (2009) and Baskent and Bazo (2011) found that hearing-impaired listeners performed more poorly than normal-hearing listeners in both auditory-only and audiovisual conditions. In addition, Tye-Murray, Sommers, & Spehar (2007a) found that the benefit of the additional visual information was approximately the same in both normal hearing and hearing-impaired groups, once performance in the auditory-only condition was equated across the two groups. The auditory component of audiovisual speech signals is a key variable in audiovisual speech performance in hearing-impaired (Corthals, Vinck, De Vel, & Van Cauwenberg, 1997; Picou, Ricketts, & Hornsby, 2013) and normal-hearing (Baart, Vroomen, Shaw, & Bortfeld, 2014) listeners. As the clarity of the auditory component of the audiovisual speech signal is reduced, performance in audiovisual speech identification is decreased as well. Therefore, it seems that poorer auditory coding by hearing-impaired individuals (relative to normal-hearing listeners) results in inferior performance for these individuals in the audiovisual identification of speech stimuli presented at a constant signal-to-noise ratio (SNR) or sound pressure level (SPL; see Baskent & Bazo, 2011; Bernstein & Grant, 2009). However, by individually setting SPL or SNR across the groups, there would be no difference between hearing-impaired and normal-hearing groups in the audiovisual identification of speech stimuli (see Tye-Murray, Sommers, & Spehar, 2007a). This is supported by studies that found no differences between hearing-impaired and normal-hearing listeners in lip-reading ability (Lyxell & Rönnberg, 1989; Tye-Murray, Sommers, & Spehar, 2007a) and audiovisual integration ability (Tye-Murray, Sommers, & Spehar, 2007a).

The present study extends a previous study by Moradi et al. (2014a) by investigating the audiovisual rather than just the auditory modality. Specifically, this study aimed to compare elderly hearing aid (EHA) users and elderly normal-hearing (ENH) individuals in terms of isolation points (IPs, Grosjean, 1980; the shortest time from the onset of a speech stimulus required for correct identification of that speech stimulus) and accuracy (in identification) for different types of audiovisual speech stimuli (consonants, words, and final words in less predictable, LP, and highly predictable, HP, sentences) presented at the same SPL in silent conditions. Another aim was to investigate the extent to which adding visual cues would impact the IPs for different types of speech stimuli in EHA users and ENH individuals. To this end, we compared audiovisual IPs and accuracies of different speech stimuli from the present study with auditory IPs and

accuracies extracted from Moradi et al. (2014a). Moradi et al. (2014a) reported that EHA users needed longer IPs for the auditory identification of consonants, words, and final words in LP sentences than ENH individuals, although there was no difference between the two groups in terms of IPs for final word identification in HP sentences. With regard to accuracy, the EHA users had lower accuracy for the auditory identification of consonants and words than the ENH individuals, but no difference was observed between the two groups either in LP or HP sentences.

Since the addition of visual cues to auditory speech stimuli greatly helps the identification of speech stimuli in terms of both IP and accuracy (see Moradi, Lidestam, & Rönnberg, 2013), we assumed that the EHA users may reach similar performance as ENH individuals, in terms of both IPs and accuracy, in audiovisual identification of speech stimuli presented at the same SPL in silent conditions. In addition, we predicted that the audiovisual IPs of different types of speech stimuli will be shorter than auditory IPs (extracted from Moradi et al., 2014a) either in EHA and ENH groups.

Methods

Participants

We recruited two groups of participants in the present study: EHA users and ENH individuals.

EHA users. A total of 20 native Swedish speakers (13 men and 7 women) with a symmetrical bilateral mild-to-moderate hearing impairment took part in this study. The participants were experienced hearing aid users selected from an audiology clinic patient list at Linköping University Hospital, Sweden. Their ages ranged from 69 to 77 years ($M=73.1$ years) at the time of testing. They had been habitual hearing aid users for at least 1 year. On average, the participants reported having had hearing loss for 6.2 years ($SD=5.5$; range, 1 year and 1 month to 14 years and 7 months). In Moradi et al. (2014a), the average duration of hearing loss was 5.4 years ($SD=3.4$; range, 2 years to 13 years and 10 months). There was no significant difference in the duration of hearing loss between the EHA group in the present study and the EHA group in Moradi et al. (2014a), $t(30.64)=0.56$, $p=.58$. In addition, when comparing pure-tone average thresholds of the across seven frequencies (PTA7) for the EHA users in the present study and Moradi et al. (2014a), there were no significant differences neither in the PTA7 left ear, $t(42)=0.04$, $p=.97$, nor in the PTA right ear, $t(42)=0.80$, $p=.43$.

In the present study, EHA users wore various in-the-ear, behind-the-ear, and receiver-in-the-ear digital

Table 1. Brands and Models of Hearing Aids Used by EHA Users.

Hearing aid	BTE, ITE, CIC, RITE	Number of participants
Oticon, Hit Pro I3	BTE	3
Oticon, Vigo Pro I3	BTE	2
Oticon, Vigo Pro T	BTE	2
Oticon, EPOQ XW	RITE	1
Oticon, EPOQ XW	CIC	1
Oticon, Vigo Pro 312	BTE	1
Phonak, Versata Art VZ	ITC/HS	1
Phonak, AMBRA M H20	BHE	1
Phonak, Versata Art micro	BHE	1
Phonak, Exelia Art micro	BTE	1
Phonak, Exelia Art M	BTE	1
Phonak, Versata Art M	BTE	1
Phonak, Exelia Art	ITE	1
Beltone, True9 78DW	BTE	1
Beltone, True9 66DW	BHE	1
Resound, Live5 LV571-DVI	BTE	1

Note. EHA = elderly hearing aid; BTE = behind the ear; ITE = in the ear; CIC = completely in the canal; ITC/HS = in-the-canal/half-shell.

hearing aids. Table 1 shows the brands and models of hearing aids used by these participants. For 12 of the hearing aid users, the current hearing aids were their first. Eight of the hearing aid users had experiences of other hearing aids before the current hearing aids. A total of 19 of the participants had been using their current hearing aids for 1 to 3 years. One participant had been using their current hearing aid for 3 years and 6 months. The hearing aids had been fitted based on each listener's individual needs, by licensed audiologists who were independent of the present study. All of the hearing aids used non-linear processing and had been fitted according to manufacturers' instructions.

As in Moradi et al. (2014a), the EHA users wore their own hearing aids, and the amplification settings of their hearing aids were not changed throughout the testing in order to prevent a novelty effect that might impact on their performance in the speech tasks.

The study inclusion criteria were as follows: (a) age over 65 years, (b) Swedish as the native language, and (c) bilateral hearing impairment with an average threshold of > 35 dB for pure-tone frequencies of 500, 1,000, 1,500, and 2,000 Hz.

Elderly people with normal hearing. A total of 20 native Swedish speakers with age-appropriate normal hearing (9 women and 11 men) took part in the present study. Their ages ranged from 67 to 76 years ($M = 71.7$ years). These individuals were from the general population living within the hearing clinic catchment area.

They were recruited primarily via invitation letters sent to their addresses and via flyers.

The inclusion criteria for this group were the following: (a) age over 65 years, (b) Swedish as the native language, and (c) a mean threshold of < 20 dB for pure-tone frequencies of 500, 1,000, 1,500, and 2,000 Hz.

Pure-tone thresholds. The mean and standard deviation of audiometric thresholds for frequencies 125, 250, 500, 1,000, 2,000, 4,000, and 8,000 Hz in the right and left ears of the participants in the EHA and ENH groups are reported in Table 2.

Participant characteristics. Participants in both groups (ENH and EHA groups) reported themselves to be in good health. They did not suffer from tinnitus, middle-ear pathology, dementia, seizures, Parkinson's disease, or psychological disorders that might compromise their ability to perform the speech and cognitive tasks.

The participants in both groups completed the Mars Letter Contrast Sensitivity Test (Arditi, 2005) and a word comprehension test (Järpsten, 2002) to measure their visual acuity and vocabulary knowledge, respectively. To be included in this study, the participants' scores in the Mars Letter Contrast Sensitivity Test had to be within age-appropriate ranges (i.e., above 1.52 contrast sensitivity log), according to the test manual (Mars Perceptrix, n.d.), and the participants had to score over 30 in the word comprehension test.

Table 3 shows the means for age, years of formal education, Mars visual acuity test, word comprehension test scores, and pure-tone average thresholds across seven frequencies (or PTA7) for the right and left ear of the EHA and ENH groups. Except PTA7 for the right and left ears, there were no significant differences between two groups in the other variables.

Ethical Considerations

All participants were fully informed about the study and gave written consent for their participation. The Linköping regional ethical review board approved the study, including the informational materials and consent procedure.

Stimuli

Talker. A female native talker with a general Swedish dialect read all of the speech stimuli at a natural articulation rate in a quiet studio while looking straight into the camera. The talker maintained a neutral facial expression, avoided blinking, and closed her mouth before and after articulation. Each target speech stimulus was recorded several times, and the best of the video and audio items recorded were selected.

Table 2. Mean and Standard Deviations (in Parentheses) of Audiometric Thresholds for EHA Users and ENH Individuals.

	125 Hz (SD)	250 Hz (SD)	500 Hz (SD)	1000 Hz (SD)	2000 Hz (SD)	4000 Hz (SD)	8000 Hz (SD)
EHA group							
Right ear	25.75 (12.06)	23.50 (10.53)	26.50 (9.33)	34.75 (9.93)	51.50 (10.77)	65.75 (11.95)	75.00 (17.09)
Left ear	26.75 (11.95)	24.75 (9.93)	25.75 (9.50)	38.25 (13.31)	55.50 (9.85)	70.00 (12.46)	74.50 (17.39)
ENH group							
Right ear	6.50 (3.66)	8.00 (3.40)	10.75 (2.94)	14.25 (3.35)	18.75 (3.58)	25.25 (4.99)	38.50 (5.64)
Left ear	7.25 (3.43)	9.25 (1.83)	11.00 (3.08)	15.25 (3.02)	20.50 (3.94)	29.25 (5.45)	39.25 (4.38)

Note. EHA = elderly hearing aid; ENH = elderly normal-hearing.

Table 3. Means, Standard Deviations and Significance Levels for EHA and ENH Groups for the Age, Years of Formal Education, Word Comprehension Test, Mars Letter Contrast Sensitivity Test, and PTA7 for the Right and Left Ears.

	EHA M (SD)	ENH M (SD)	Inferential statistics EHA vs. ENH ($df=38$)
Age	73.05 (2.84)	71.65 (2.54)	$t=1.64, p=.108$
Years of formal education	12.65 (2.41)	13.50 (2.57)	$t=-1.08, p=.287$
Word comprehension test	32.60 (0.883)	33.15 (0.875)	$t=-1.98, p=.055$
Mars letter contrast sensitivity test: binocular	1.674 (0.030)	1.668 (0.032)	$t=0.61, p=.543$
PTA7 right	43.25 (5.85)	17.43 (2.55)	$t=18.09, p<.001, d=6.15$
PTA left	45.07 (5.95)	18.82 (2.58)	$t=18.10, p<.001, d=6.16$

Note. EHA = elderly hearing aid; ENH = elderly normal-hearing.

Video recording. Visual speech stimuli were recorded with a RED ONE digital camera (RED Digital Cinema Camera Company, CA, USA) at a rate of 120 frames per second (each frame = 8.33 ms, see Figure 1), in $2,048 \times 1,536$ pixels. Note that at this frame rate, the camera cannot record sound; therefore, the auditory component of the audiovisual speech signal had to be recorded separately. The video recording was segmented into separate target speech items using Final Cut Pro software, version 7.0.3 (Apple Inc., CA, USA). In the next step, the video files were cropped so that the number of pixels to be processed was reduced to 600×670 pixels, and then saved as non-compressed “.mov” files. The reducing of pixels of the recorded stimuli had two aims. First, it lowered the processing demands for playback, ensuring that presentation could be executed without synchronization errors according to Psychophysics Toolbox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). Second, it matched the pixels of “.mov” files with the settings of the screen used for presentation (i.e., no loss in spatial resolution). Each video file on the testing computer monitor showed the hair, face, and top part of the talker’s shoulders against a dark gray background. The video files were inspected for anything that may distract the participants. The start and end frames of each video file showed a still face.

Audio recording. The auditory speech stimuli were recorded with a directional electret condenser stereo microphone at 16 bits and a sampling rate of 48 kHz. The recorded auditory stimuli were segmented into separate auditory target speech stimuli using Sound Studio 4 software (Felt Tip Inc., NY, USA). The onset and offset of each auditory speech stimulus were set carefully according to inspection of the speech waveform (using Sound Studio 4) and auditory feedback by the first two authors. Each auditory speech stimulus was then saved as a “.wav” file. The root mean square value was calculated for each speech stimulus, and the stimuli were then rescaled to equate levels across the speech stimuli. The audio speech stimuli were inspected for clicks, noise, and phonemic distinctiveness.

Measures

A detailed description of the gated speech tasks employed in the present study is available in Moradi et al. (2013, 2014a). We provide a brief description of the gated tasks below. Note that the gated speech tasks in the present study used exactly the same speech stimuli employed by Moradi et al. (2014a) in auditory identification of different types of speech stimuli. In the present study, we presented the same speech stimuli audiovisually.

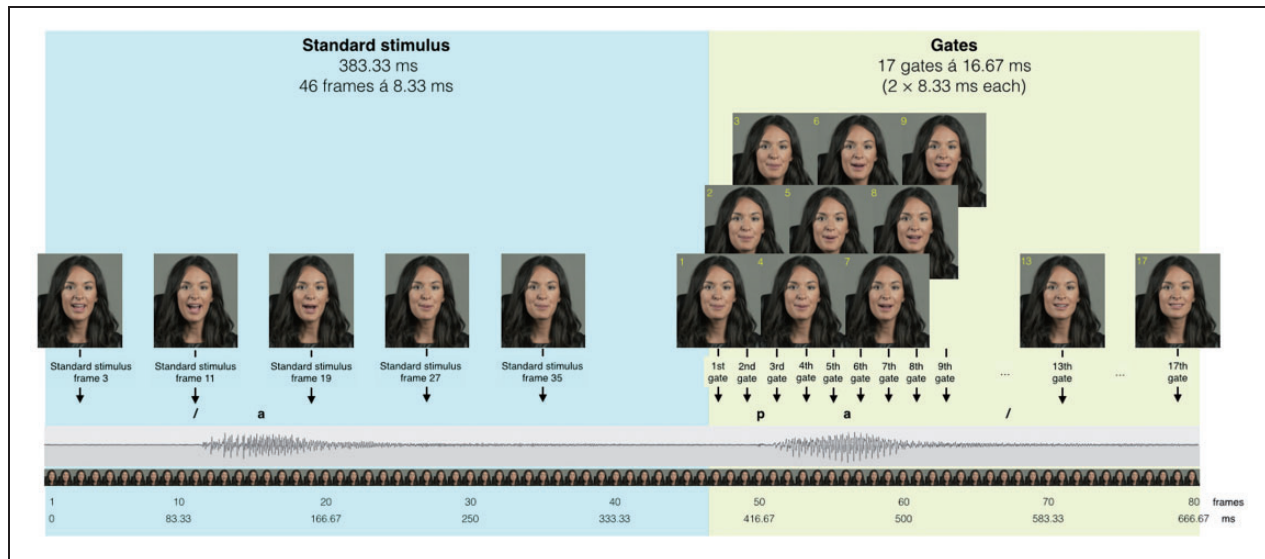


Figure 1. An illustration of gating for audiovisual identification of consonants.

Consonants. A total of 18 Swedish consonants, structured in a vowel-consonant-vowel syllable format (/aba, ada, afa, aga, aja, aha, aka, ala, ama, ana, aᅇa, apa, ara, aᅇa, asa, aᅇa, ata, and ava/) were employed in the present study. The phonemic context /aCa/ was used to minimize coarticulation effects. The gate size for consonants was set at 16.67 ms. Gating started after the first vowel, /a/, immediately at the start of the consonant onset. Thus, the first gate included the vowel /a/ plus the initial 16.67 ms of the consonant, the second gate added a further 16.67 ms of the consonant (total of 33.33 ms), and so on. The consonant-gating task took 10–15 minutes per participant to complete. Figure 1 shows an example of audiovisual gating presentation for consonant identification.

Words. We employed 23 Swedish monosyllabic words in a consonant-vowel-consonant format (CVC, all nouns). These words were selected from 46 Swedish monosyllabic words used in the study by Moradi et al. (2013). Each word used in the present study had a small-to-average number of neighbors (i.e., three to six alternative words with the same pronunciation of the first two phonemes). The gate size for words was set at 33.3 ms, as used by our previous studies. The explanation for this gate size was based on our pilot studies showed that the identification of words with the gate size of 16.67 ms started from the first phoneme in CVC format lead to exhaustion and loss of motivation. Hence, a double gate size (33.3 ms) started from the onset of second phoneme has been used to avoid fatigue in participants. The word-gating task took around 15 to 20 minutes to complete.

Final words in sentences. There were two sentence types in this study; the types differed according to how predictable the last word in each sentence was. The sentences ended with either an HP word, for example, “Lisa gick till biblioteket för att låna en *book*” (“Lisa went to the library to borrow a *book*”), or an LP word, for example, “I förorten finns en fantastisk *dal*” (“In the suburb there is a fantastic *valley*”). The final (target) word in each sentence was always a monosyllabic noun. The gate size for identification of final words in sentences was set at 16.67 ms. In total, there were 22 sentences (11 HP sentences and 11 LP sentences). The sentence-gating task took around 10 to 15 minutes to complete.

Procedure

An iMac (OS X 10.8.5) running MATLAB (R2013b) and Psychophysics Toolbox (version 3.0.11) were used to synchronize the audio and video speech stimuli and to present the audiovisual gated stimuli. Details about the synchronization of audio and video stimuli, and about the MATLAB script used to gate the speech stimuli, are available in Lidestam (2014). The iMac was equipped with a fast solid-state hard drive (Pegasus J2), and a fast interface to ensure adequate speed for video rendering and playback. The iMac was configured for dual-screen presentation. The visual stimuli were displayed on a 21” CRT monitor (DELL UltraScan P1110, 120-Hz refresh rate, 800 × 600 pixel resolution) inside the sound booth and viewed from a distance of 70 cm. The audio stimuli were delivered via the iMac, which was routed to the input of two loudspeakers (Genelec 8030A) located to the right and left of the CRT monitor.

The experimenter used the iMac outside the sound booth to present the gated stimuli, monitor the participants' progress, and record the participants' responses. A microphone (in the sound chamber, routed into the audiometry device) delivered the participants' verbal responses to the experimenter through a headphone connected to the audiometry device. The average overall SPL for the audiovisual gated speech stimuli was 65 dB SPL (as in Moradi et al., 2014a) for both EHA and ENH groups. This was measured in the vicinity of the participant's head with a Larson-Davis System 824 (UT, USA) sound level meter in free field.

The testing procedure was similar to that described by Moradi et al. (2014a); however, the current study additionally included the Mars Letter Contrast Sensitivity Test, which was utilized to assess participants' visual contrast sensitivity. Participants were tested individually in a sound booth. Initially, pure-tone hearing thresholds (125–8000 Hz) were obtained (using an Interacoustics AC40 audiometer) and then the visual contrast sensitivity scores were acquired (using the Mars Letter Contrast Sensitivity Test).

The participants underwent a practice session to become familiarized with the gated presentation of stimuli, which involved completing some trial runs. The practice session comprised three gated consonants (/v k ŋ/) and two gated words (/tum [inch]/ and /bil [car]/). Feedback was provided during the training session but not during the experiment. After the practice, the gating paradigm started.

All participants began with the consonant identification task, followed by the words task, and ending with the final words in sentences task. There were short rest periods to prevent fatigue. The order of item presentation within each gated task (i.e., consonants, words, and final words in sentences) varied among the participants. Participants gave their responses orally and the experimenter wrote these down.

The presentation of gates continued until the target item was correctly recognized on six consecutive presentations; this meant that random guessing was avoided. If the target item was not correctly recognized, presentation continued until the end of the stimulus. When a target was not correctly identified, its entire duration plus one gate size was calculated as the IP for that item (this scoring method corresponds to our previous studies and to other studies that have employed the gating paradigm; Elliott, Hammer, & Evan, 1987; Hardison, 2005; Lidestam, Moradi, Petterson, & Ricklefs, 2014; Metsala, 1997; Moradi et al., 2013, 2014a; Moradi, Lidestam, Saremi, & Rönnerberg, 2014; Walley, Michela, & Wood, 1995).

The word comprehension test (a measure of vocabulary knowledge) was administered in a second session with the other cognitive and speech-in-noise tests. In

the present study, we only report the results for the gated speech stimuli.

Results

Group Comparison of Gated Audiovisual Speech Task Results

The mean IPs for the gated audiovisual speech tasks are reported in Table 4. A 2 (Hearing loss: EHA, ENH) \times 4 (Gated task: consonants, words, final words in HP and LP sentences) mixed analysis of variance (ANOVA) with repeated measures on the second factor was conducted to examine the effect of hearing loss on the IPs for the identification of different types of audiovisual speech stimuli. The results showed a main effect of aided hearing loss, $F(1, 38) = 12.67$, $p < .001$, $\eta_p^2 = 0.25$, and a main effect of tasks, $F(1.66, 63.21) = 3085.97$, $p < .001$, $\eta_p^2 = 0.99$. The interaction between aided hearing loss and tasks was also significant, $F(1.66, 63.21) = 8.41$, $p < .001$, $\eta_p^2 = 0.18$. Four planned comparisons showed that the EHA users needed longer IPs than the ENH individuals for the identification of consonants, $t(38) = 2.42$, $p = .020$, and the identification of words, $t(38) = 3.47$, $p < .001$. However, there were no significant differences between the two groups for the identification of final words in LP sentences, $t(38) = 1.79$, $p = .081$, and final words in HP sentences, $t(38) = -0.40$, $p = .689$.

Table 5 shows the mean accuracy for the identification of stimuli in the different audiovisual gated speech tasks in EHA users and ENH individuals. A 2 (Hearing loss: EHA, ENH) \times 4 (Gated task: consonants, words, final words in HP and LP sentences) mixed ANOVA with repeated measures on the second factor was conducted to examine the effect of aided hearing loss on the accuracy for the identification of different types of audiovisual speech stimuli. The results showed that the main effect of aided hearing loss was not significant, $F(1, 38) = 0.73$, $p = .398$. However, the main effect of gated tasks was significant, $F(3, 114) = 19.49$, $p < .001$, $\eta_p^2 = 0.34$. The interaction between aided hearing loss and gated tasks was not significant, $F(3, 114) = 0.56$, $p = .644$.

Thus, the results showed that the EHA users needed longer IPs for the identification of speech stimuli when a supportive semantic context was lacking. In terms of accuracy, the EHA users and ENH individuals demonstrated a similar level of performance for the identification of different types of audiovisual speech stimuli.

Comparison of Gated Audiovisual Versus Auditory Speech Task Results

In the next step, we compared the IPs and accuracies for different types of audiovisual speech tasks in the present study with those for different auditory speech tasks

Table 4. Mean IPs, SD (in Parentheses), and Significance Levels for the Identification of Different Types of Speech Stimuli in EHA Users and ENH Individuals Presented Audiovisually and Auditorily (Moradi et al. 2014a).

Types of gated tasks	Descriptive statistics				Inferential statistics			
	Audiovisual		Auditory		Audiovisual vs. auditory		EHA users vs. ENH individuals	
	EHA users (a)	ENH individuals (b)	EHA users (c)	ENH individuals (d)	EHA users (df=42)	ENH individuals (df=42)	Audiovisual (df=38)	Auditory (df=46)
Consonants	112.85 (21.21)	97.98 (17.46)	145.28 (27.02)	117.46 (18.02)	t=4.36, p<.001, d=1.35	t=3.62, p<.001, d=1.10	t=2.42, p=.021, d=0.77	t=3.99, p<.001, d=1.24
Words	449.10 (41.77)	406.92 (34.77)	560.34 (34.20)	502.01 (31.32)	t=9.72, p<.001, d=2.89	t=9.54, p<.001, d=2.86	t=3.47, p<.001, d=1.10	t=6.11, p<.001, d=1.78
Final words in LP	128.31 (11.98)	121.11 (13.34)	140.40 (23.59)	122.22 (19.73)	t=2.08, p=.044, d=0.66	t=0.21, p=.826	t=1.79, p=.081	t=2.90, p=.006, d=0.84
Final words in HP	20.03 (4.53)	20.59 (4.18)	20.20 (3.46)	20.25 (2.84)	t=0.14, p=.892	t=-0.32, p=.753	t=-0.40, p=.689	t=-0.59, p=.953

Note. EHA = elderly hearing aid; ENH = elderly normal-hearing; LP = less predictable; HP = highly predictable; IP = isolation points.

Table 5. Descriptive Statistics for the Accuracy of Consonants, Words, and Final Words in HP and LP Sentences in the EHA Users and the ENH Individuals Presented Audiovisually (present study) and Auditory (Moradi et al. 2014a).

Types of Gated Tasks	Audiovisual		Auditory	
	EHA M (SD)	ENH M (SD)	EHA M (SD)	ENH M (SD)
Consonants	93.33 (8.94)	95.28 (6.57)	80.32 (11.70)	94.68 (6.45)
Words	98.48 (3.24)	99.14 (1.77)	84.76 (8.69)	98.73 (2.39)
Final words in LP	100.00 (0.00)	100.00 (0.00)	96.60 (4.15)	98.62 (3.18)
Final words in HP	100.00 (0.00)	100.00 (0.00)	100.00 (0.00)	100.00 (0.00)

Note. EHA = elderly hearing aid; ENH = elderly normal-hearing; LP = less predictable; HP = highly predictable.

observed in our previous study (Moradi et al., 2014a). This comparison (Table 4) enabled us to investigate the extent to which the addition of visual cues on the auditory speech stimuli affected the IPs and accuracy with different types of speech stimuli. A 2 (Modality: audiovisual, auditory) \times 2 (Aided hearing loss: EHA, ENH) \times 4 (Gated task: consonants, words, final words in HP and LP sentences) mixed ANOVA with repeated measures on the third factor was computed to examine the effects of presentation modality and aided hearing loss on the mean IPs for different types of gated task. The results showed a main effect of modality, $F(1, 84) = 128.62, p < .001, \eta_p^2 = 0.61$, a main effect of aided hearing loss, $F(1, 84) = 49.30, p < .001, \eta_p^2 = 0.37$, and a main effect of gated tasks, $F(1.89, 158.69) = 8278.40, p < .001, \eta_p^2 = 0.99$. The interaction between presentation modality and aided hearing loss was not significant, $F(1, 84) = 2.88, p = .093$. However, there were significant interactions between presentation modality and gated tasks, $F(1.89, 158.69) = 115.09, p < .001, \eta_p^2 = 0.58$, and aided hearing loss and gated tasks, $F(1.89, 158.69) = 23.47, p < .001, \eta_p^2 = 0.22$. The three-way interaction between presentation modality, aided hearing loss, and gated tasks was not significant, $F(1.89, 158.69) = 0.59, p = .548$. When comparing the IPs of audiovisual relative to auditory presentation among EHA users, the advantage of audiovisual presentation was observed for the identification of consonants, words, and final words in LP sentences. In the ENH group, the advantage of audiovisual presentation was observed only for the identification of consonants and words.

Consonants. Table 6 reports the mean IPs for the identification of different types of speech stimuli presented audiovisually and aurally in the EHA and ENH groups. A 2 (Modality: audiovisual, auditory) \times 2 (Aided hearing loss: EHA, ENH) \times 18 (Consonants) mixed ANOVA with repeated measures on the third factor was computed to examine the effects of modality and aided hearing loss on the mean IPs for Swedish consonants. The results showed a main effect of modality,

$F(1, 84) = 31.99, p < .001, \eta_p^2 = 0.28$, a main effect of aided hearing loss, $F(1, 84) = 21.63, p < .001, \eta_p^2 = 0.21$, and a main effect of consonants, $F(5.555, 466.613) = 188.82, p < .001, \eta_p^2 = 0.69$. The interaction between modality and aided hearing loss was not significant, $F(1, 84) = 1.99$. The interaction between aided hearing loss and consonants was not significant, $F(5.555, 466.613) = 2.07, p = .061$. However, the interaction between modality and consonants was significant, $F(5.555, 466.613) = 4.79, p < .001, \eta_p^2 = 0.05$. The three-way interaction between modality, aided hearing loss, and consonants was not significant, $F(5.555, 466.613) = 1.57, p = .158$. When comparing the audiovisual IPs of consonants relative to auditory ones (see Table 6), the audiovisual presentation significantly shortened the IPs for 11 consonants (/b d f g h j l m s ʃ v/) in the EHA users. In the ENH group, audiovisual presentation significantly shortened the IPs for 7 consonants (/b l p r s t v/). When comparing the IPs of consonants between the EHA and ENH groups in audiovisual and auditory modalities, the EHA group needed longer IPs than the ENH group for /l n t/ in audiovisual modality and longer IP for /f/ in auditory modality.

Words. A 2 (Modality: audiovisual, auditory) \times 2 (Aided hearing loss: EHA, ENH) ANOVA was conducted to examine the effects of modality and aided hearing loss on the mean IPs for Swedish monosyllabic words (Table 4). The results showed a main effect of modality, $F(1, 84) = 184.77, p < .001, \eta_p^2 = 0.69$, and a main effect of aided hearing loss, $F(1, 84) = 43.84, p < .001, \eta_p^2 = 0.34$. However, the interaction between modality and aided hearing loss was not significant, $F(5.555, 466.613) = 1.13, p = .290$. When comparing the audiovisual IPs of words relative to auditory ones, audiovisual presentation significantly shortened the IPs for both EHA users, $t(42) = 9.72, p < .001$ and ENH group, $t(42) = 9.54, p < .001$.

Final words in sentences. A 2 (Modality: audiovisual, auditory) \times 2 (Aided hearing loss: EHA, ENH) \times 2 (Sentence predictability: high, low) mixed ANOVA with repeated

Table 6. Descriptive and Inferential Statistics for IPs of Consonants for EHA Users and ENH Individuals Presented Audiovisually and Auditorily (Moradi et al. 2014a).

Consonants	Modality				<i>p</i>			
	Audiovisual		Auditory		Audiovisual vs. auditory		EHA users vs. ENH individuals	
	EHA users (a)	ENH individuals (b)	EHA users (c)	ENH individuals (d)	EHA users (a – c)	ENH individuals (b – d)	Audiovisual (a – b)	Auditory (c – d)
b	104.19 (32.84)	81.68 (25.88)	154.20 (47.21)	132.67 (37.59)	.001	.001	.021	.088
d	119.19 (37.58)	110.02 (46.02)	154.20 (28.77)	134.75 (28.63)	.001	.035	.494	.022
f	100.85 (29.36)	85.85 (32.57)	151.42 (63.89)	102.80 (23.40)	.002	.05	.134	.002
g	124.19 (34.41)	121.69 (52.19)	169.48 (46.55)	154.20 (41.78)	.001	.027	.859	.238
h	91.69 (20.59)	85.02 (16.13)	122.94 (41.37)	99.33 (21.70)	.0026	.019	.262	.018
j	85.85 (12.42)	75.02 (17.53)	119.47 (55.11)	86.82 (28.23)	.001	.111	.031	.014
k	60.01 (13.68)	55.01 (12.21)	72.24 (18.83)	59.73 (19.61)	.017	.355	.231	.029
l	105.02 (18.81)	79.18 (18.64)	136.14 (42.76)	104.19 (24.70)	.003	.001	.001	.003
m	105.02 (23.01)	99.19 (27.83)	143.08 (63.89)	109.74 (46.35)	.001	.377	.475	.016
n	141.70 (41.37)	100.85 (24.47)	163.23 (71.40)	126.41 (44.49)	.22	.027	.001	.038
ŋ	195.87 (46.80)	171.70 (50.19)	210.46 (35.04)	173.65 (50.35)	.258	.899	.124	.005
p	60.85 (23.12)	44.18 (12.42)	80.57 (24.90)	70.15 (12.99)	.010	.001	.008	.078
r	105.02 (23.64)	90.85 (19.85)	131.97 (43.67)	118.77 (27.51)	.013	.001	.047	.261
t	312.56 (104.86)	299.23 (103.11)	330.62 (99.37)	239.63 (108.24)	.564	.070	.688	.004
s	47.51 (13.55)	45.84 (10.65)	99.33 (57.22)	78.49 (20.55)	.001	.001	.667	.104
ʃ	119.19 (25.53)	100.02 (32.90)	156.28 (39.88)	126.41 (32.20)	.001	.010	.047	.007
t	55.84 (12.42)	41.68 (10.12)	69.46 (21.24)	56.26 (17.60)	.012	.002	.001	.024
v	96.69 (22.04)	76.68 (27.26)	150.03 (48.66)	140.31 (44.76)	.001	.001	.015	.475

Note. Significant differences according to Bonferroni adjustment ($p < .00278$) are in bold. EHA = elderly hearing aid; ENH = elderly normal-hearing; LP = less predictable; HP = highly predictable; IP = isolation points.

measures on the third factor was computed to examine the effects of modality and aided hearing loss on the mean IPs for final words in sentences (Table 4). The results showed that the main effect of modality was not significant, $F(1, 84) = 2.51$, $p = .117$. However, the main effect of aided hearing loss, $F(1, 84) = 9.07$, $p = .003$, $\eta_p^2 = 0.10$, and the main effect of sentence predictability, $F(1, 84) = 3141.99$, $p < .001$, $\eta_p^2 = 0.97$, were significant. The interactions between modality and aided hearing loss, $F(1, 84) = 1.95$, $p = .166$, and modality and sentence predictability, $F(1, 84) = 3.03$, $p = .086$, were not significant. However, the interaction between aided hearing loss and sentence predictability was significant, $F(1, 84) = 11.42$, $p < .001$, $\eta_p^2 = 0.12$. The three-way interaction between modality, aided hearing loss, and sentence predictability was not significant, $F(1, 84) = 1.86$, $p = .176$. When comparing the IPs for audiovisual versus auditory presentation, the audiovisual presentation significantly shortened the IPs for final words in LP sentences in the EHA group but not in the ENH group.

There was no effect of audiovisual presentation on IPs for final words in HP sentences either in the EHA or the ENH group.

Discussion

The goals of the current study were (a) to compare the IPs and accuracies of different types of audiovisual speech stimuli (consonants, words, and final words in LP and HP sentences) between EHA users and ENH individuals and (b) to compare audiovisual IPs for different types of speech stimuli from the present study with auditory IPs for those speech stimuli extracted from Moradi et al. (2014a).

Main Findings

The results reveal that the EHA group needed longer IPs than the ENH group for the audiovisual identification of speech stimuli in the absence of a prior semantic context.

In terms of accuracy, the two groups reached ceiling, and there was no difference between the two groups in the audiovisual identification of different types of speech stimuli. The addition of visual cues to auditory speech stimuli (when comparing audiovisual IPs with auditory IPs) shortened the IPs for consonants, words, and final words in LP sentences in the EHA group. In the ENH group, the addition of visual cues only shortened IPs for consonants and words.

Consonants. In the present study, the EHA users needed longer IPs than the ENH individuals for the identification of Swedish consonants (113 vs. 98 ms), while there was no difference in terms of accuracy between the two groups. The correspondence between the visual and auditory components of consonants is not one-to-one as some consonants look the same during visual articulation, such as /b p m/, /v f/, /k g/ /r l/, and /d t s/. While visual cues provide information about the place of articulation, auditory cues provide information about the manner of articulation. Visual cues are almost always available earlier than auditory cues during the audiovisual articulation of speech stimuli (Chandrasekaran, Trubanova, Stillitano, Caplier, & Ghazanfar, 2009; Smeele, 1994). According to the predictive coding hypothesis (Friston & Kiebel, 2009; see also the on-line prediction hypothesis, van Wassenhove, Grant, & Poeppel, 2005), initial visual articulation activates some phonological representations (predictions or residual errors) in the brain regarding the identity of a given audiovisual phoneme that is matched with earlier visual cues. These predictions are constantly updated as more visual and auditory inputs are received; this decreases the number of predictive phonological representations (and/or residual errors) until a phonological representation is left that matches with the incoming visual and auditory cues.

As mentioned earlier, the clarity of the audio component of the audiovisual speech signal is crucial to the audiovisual identification of speech stimuli (Baart et al., 2014; Corthals et al., 1997). As the EHA users had inferior performance compared with the ENH group in the auditory coding of consonants (see Moradi et al. 2014a), we assume that the hearing-impaired individuals, even with their own hearing aids, suffered from poor auditory coding also during the audiovisual presentation of consonants. As a consequence, they had larger residual errors than the ENH group that required extended gated presentation of consonants (as indicated by delayed IPs) to view a coherent audiovisual speech signal for recognition. For instance, the EHA users are likely to have needed more gated presentations than the ENH individuals to discriminate between /t k/ or /l r/ (see Table 6 for comparison of audiovisual IPs in EHA and ENH groups). In addition,

we suggest that the initial visual presentation of some consonants, such as /t/, likely activated more phonological candidates in the EHA users than the ENH individuals, which necessitated more gated presentations for correct identification. However, there was no difference between the two groups in terms of accuracy for the audiovisual identification of consonants. This finding suggests that although EHA users needed longer IPs for consonants, they were eventually able to correctly recognize consonants, at the same level as their age-matched counterparts with normal hearing.

When comparing audiovisual to auditory presentation, the results indicate that audiovisual presentation speeds up the identification of consonants relative to auditory-only presentation, regardless of whether an individual has hearing loss. However, the addition of visual cues to the auditory speech signal (representing a complementary effect) benefited the EHA group more than the ENH group. As shown in Table 6, audiovisual presentation (compared with auditory-only presentation) significantly shortened the IPs for seven voiced (/b, d, g, j, l, m, v/) and four fricative (/f, h, s, ʃ/) consonant types in the EHA group, while in the ENH group audiovisual presentation shortened the IPs for five voiced (/b, l, p, r, v/), one fricative (/s/), and one plosive (/t/) consonant type. There was less benefit from the combination of video and audio (representing a redundancy effect) for 7 consonants in EHA group and 11 consonants in the ENH group in the silent condition. This finding is in line with the notion that the benefits of audiovisual presentation over auditory presentation are greatest under degraded listening conditions, such as noise (see Moradi et al., 2013) or hearing loss (see Sheffield, Schuchman, & Bernstein, 2015), when access to critical auditory cues for the identification of consonants is impoverished by background noise or a reduction in auditory acuity due to hearing loss. The addition of visual cues to a degraded auditory signal is a major source of disambiguation, as it provides complementary cues about the place of articulation (Summerfield, 1987) and indicates where and when to expect the onset and offset of a given consonant (see Best, Ozmeral, & Shinn-Cunningham, 2007).

Overall, our findings corroborate those of prior studies by showing that the audiovisual compared with auditory-only presentation of consonants improves performance in people with hearing loss in both aided and unaided conditions (Grant, Walden, & Seitz, 1998; Tye-Murray, Sommers, & Spehar, 2007a; Walden et al., 2001; Walden, Prosek, & Worthington, 1975) and in people with normal hearing (Sommers, Tye-Murray, & Spehar, 2005). Further, the greatest benefit of the audiovisual over auditory presentation of consonants in the EHA group was at the accuracy level, since accuracy improved to the same level as the ENH group.

Words. The results of the present study show that EHA users needed longer IPs relative to ENH individuals for the identification of Swedish monosyllabic words (449 vs. 407 ms), while the participants in both groups achieved ceiling levels in terms of accuracy. Word recognition occurs when the incoming speech signal maps with a lexical representation in the mental lexicon (Lively, Pisoni, & Goldinger, 1994). According to the cohort model of word recognition (Marslen-Wilson, 1993; Marslen-Wilson & Welsh, 1978), the initial presentation of a given speech signal activates particular lexical candidates in the mental lexicon. As more of the speech signal is acquired, the number of activated lexical candidates is decreased, until one lexical candidate remains that matches with the incoming speech signal. The number of activated lexical candidates is greatly dependent on lexical frequency and phonological neighborhood density (Dufor & Frauendelder, 2010; Luce & Pisoni, 1998), and modality presentation (i.e., auditory, visual, or audiovisual; see Tye-Murray, Sommers, & Spehar, 2007b). In addition, the presentation of words under degraded listening conditions (background noise or hearing loss) results in longer IPs for the identification of stimuli presented in either auditory or audiovisual modalities (Moradi et al., 2013; Moradi, Lidestam, Hällgren, et al., 2014, Moradi, Lidestam, Saremi, et al., 2014). This is most likely due to difficulty in moving from one lexical candidate to the target lexical item (see Singer, Bronstein, & Miles, 1981).

As noted earlier, the words in our study had average-to-high frequencies, with a small-to-average number of neighbors (three to six alternative words with the same pronunciation of the first two phonemes). The longer IPs in the EHA group relative to the ENH group may be due to poor auditory coding of words during processing of the incoming audiovisual speech signal, which activates a greater number of similar phonological-lexical candidates, or leads to a persistent focus on a non-target lexical item during the gated presentation of words in the EHA group. As a consequence, the EHA group required more of the incoming audiovisual lexical signal (as indicated by IPs) to correctly map the audiovisual speech signal onto the target lexical item in the mental lexicon. The increase in the length of the incoming audiovisual lexical signal required by the EHA group (as indicated by IPs) eventually enabled the group to correctly map the incoming signals onto their corresponding lexical representation in the mental lexicon, which resulted in the same level of accuracy as the ENH group.

When comparing audiovisual to auditory presentation, the results of our study suggest that audiovisual presentation significantly speeds up the identification of consonants compared with auditory-only presentation. In fact, the addition of visual cues to a poor auditory

lexical signal may facilitate the lexical access by amplifying bottom-up processing (viewing the initial articulation of the lexical signal to discriminate stimuli, e.g., /bar/ and /far/) and by reducing the number of phonological-lexical candidates as a result of the overlap of words presented visually and aurally as opposed to aurally only (see Tye-Murray, Sommers, & Spehar, 2007b). As a consequence, the accurate mapping of lexical signals with corresponding lexical representations in the mental lexicon is less difficult in an audiovisual relative to an auditory-only modality, and this resulted in shortened IPs in the audiovisual relative to the auditory modality in both the EHA and ENH groups. This finding is in agreement with prior studies showing that the addition of visual cues to auditory lexical signals expedites lexical access in correctly identifying words (see de la Vaux & Massaro, 2004; Moradi et al., 2013).

Final words in sentences. The results of the present study revealed no difference between the EHA group and the ENH group in the identification of final words in sentences, in either LP or HP sentences, both in terms of IPs and accuracy.

Prior semantic context facilitates the identification of target words embedded in congruent sentences compared with the presentation of words alone, particularly under degraded listening conditions (Boothroyd & Nittrouer, 1988; Grant & Seitz, 2000; Salasoo & Pisoni, 1985). Prior semantic context activates only lexical candidate(s) in the mental lexicon that are congruent with the meaning of a given sentence, which facilitates the identification of final words in sentences. The facilitative effect of semantic context greatly depends on the degree of predictability provided by the prior semantic context (see Bradlow & Alexander, 2007; Molis et al., 2015; Moradi, Lidestam, Hällgren, et al., 2014, Moradi, Lidestam, Saremi, et al., 2014). A highly predictable sentence may activate only one lexical candidate (i.e., “a pigeon is a kind of *bird*”), whereas a sentence with less predictability will activate a set of lexical candidates that are compatible with the meaning of the sentence (i.e., “bird” in “she pointed at the *xxxx*”). In young normal-hearing listeners, the addition of visual cues to semantic context resulted in faster and more accurate identification of speech stimuli than auditory-alone presentation of sentences, particularly under degraded listening conditions (Moradi et al., 2013; Van Engen, Phelps, Smiljanic, & Chandrasekaran, 2014).

Moradi et al. (2014a) reported that EHA users needed longer IPs than ENH individuals for the auditory identification of target words in LP sentences, but there was no difference between the two groups in terms of accuracy for LP sentences. The results of the present study indicate that the EHA group additionally benefited

from the combination of prior context and visual cues, helping the individuals in this group to disambiguate the target words in the LP sentences, resulting in the same level of performance between the EHA and ENH groups both in terms of IPs and accuracy. The explanation for the non-significant differences in final words is that prior semantic context restricts the number of activated lexical candidates in the mental lexicon and visual cues by discriminating the initial phonemes of target words in sentences (e.g., “bet” vs. “pet”), and by reducing the number of phonological neighbors as a result of the overlap of auditory and visual speech cues (see Tye-Murray, Sommers, & Spehar, 2007b), making the identification of target words at the end of LP sentences less difficult for the EHA group. Jesse and Janse (2012) reported that the benefit obtained from adding visual cues to meaningful sentences in a phoneme-monitoring task was more evident in older listeners with hearing loss than in younger adults with normal hearing.

The effect of prior semantic context is stronger for final words in HP sentences than for final words in LP sentences. Moradi, Lidestam, Hällgren, et al. (2014) and Moradi, Lidestam, Saremi, et al. (2014) showed that listeners are able to correctly guess the identity of final words in HP sentences between the first and second gates for speech stimuli presented in an auditory modality. Visual information has little or no effect on the identification of final words in HP sentences compared with LP sentences because of the strength of the semantic context effects in HP sentences. This explains why the EHA and ENH groups performed similarly, in terms of both IP and accuracy, when identifying final words in HP sentences.

The present study findings (with the exception of EHA users' results for the LP sentences task) indicated no beneficial effects for elderly people of adding visual cues to semantic context (as supported by EHA users' results for the HP sentences task, and the ENH group's results for both the LP and HP sentence tasks). This finding is not in agreement with prior studies on young normal-hearing persons, where it was reported that the presentation of both semantic context and visual cues improved the intelligibility of target words in meaningful sentences (Moradi et al., 2013; Van Engen et al., 2014). One explanation might be that older adults generally have a greater reliance on the semantic context than younger adults (see Rogers, Jacoby, & Sommers, 2012) and seemingly the benefit from congruent semantic context is greater in elderly people (see Pichora-Fuller, 2008; Rogers et al., 2012; Sheldon, Pichora-Fuller, & Schneider, 2008). Similarly, Sommers and Danielson (1999) reported that although older adults had greater difficulty than younger adults in identifying low-frequency words with similar phonological neighbors, the effect was eliminated when these words were embedded

in a congruent semantic context. In fact, because of experiences accumulated over time, elderly people are more skilled than younger adults to benefit from semantic context, since they need to compensate for their sensory and cognitive decline in identification of target speech signal (see Aydelott, Leech, & Crinion, 2010; Frisina & Frisina, 1997; Pichora-Fuller, Schneider, & Daneman, 1995). We argue that because of the greater benefit from semantic context in elderly people (compared with young normal-hearing listeners), lexical candidates that are not matched to prior sentential context will quickly be dropped, and no further aid can be attained from visual cues. However, the additive effect of visual cues and semantic context was observable in LP sentences for the EHA group only and not for the ENH group. Thus, it can be argued that the additive effect of visual cues and semantic context was evident under degraded listening conditions (i.e., noise or hearing loss) in the current study, whereby visual cues in combination with semantic context facilitated the identification of target words at the end of sentences.

The interplay between semantic context and visual cues in the identification of embedded words in sentences needs further research. We suggest that the interactive effects of visual cues and semantic context greatly depend on the sentence level of predictability, the population of listeners being assessed (e.g., young vs. elderly people), and the listening conditions (e.g., clear vs. degraded). For instance, the predictability of sentences is a key factor, as when predictability is highest (e.g., final words in HP sentences), there would be less or even no benefit from the addition of visual cues to speech stimuli. However, when the sentence predictability level is decreased (e.g., final words in LP sentences), visual cues can be extremely beneficial, and, when combined with semantic context, they can facilitate target word identification in sentences. Furthermore, the addition of visual cues to semantic context is more evident under degraded listening conditions, particularly for elderly people (see Pichora-Fuller, 2008); the reduced clarity of semantic context (by noise or hearing loss) can highlight the contribution of visual cues in the disambiguation of a target signal.

Sensitivity of the Measures

Psycholinguistic research has demonstrated that the latency measures such as response time are more sensitive than accuracy because measurement for each item is continuous whereas accuracy is discrete (i.e., correct or not). For instance, response times were generally much shorter with use of hearing aids, whereas accuracy was not affected nearly as much (Gatehouse & Gordon, 1990). Adverse listening conditions (e.g., background noise) affected intelligibility of speech tasks in Houben,

van Doorn-Bierman, and Dreschler (2013) and in Huckvale and Leak (2009). Phonemes could be better categorized based on response times than on accuracy (Pisoni & Tash, 1974). Similarly, IP (by measuring the shortest time required for identification of a speech stimulus from the onset of a speech signal) is another latent measure that provides a great range of responses even in optimum listening conditions, unlike performance accuracy that can reach ceiling levels (e.g., Moradi et al., 2013). The results of the present study demonstrated the sensitivity of IPs over accuracy in revealing differences between the EHA and ENH groups in the identification of speech stimuli. Although there was no difference between the two groups in terms of accuracy, as both groups performed at ceiling, EHA users needed longer audiovisual IPs for consonants and words. That is, the IP reflects that EHA users need a longer amount of signal than ENH individuals to map the sensory signal onto corresponding phonological and lexical representations. This can reflect the established sensory disadvantage at the phonological and lexical levels in aided hearing-impaired listeners than their counterparts with normal hearing (Ahlstrom, Horwitz, & Dubno, 2014; Dimitrijevic et al., 2004; Moradi, Lidestam, Hällgren, et al., 2014), even in audiovisual modality.

Limitations and Future Considerations

One limitation of the present study is that we compared ENH individuals with EHA users who wore their own hearing aids, with no changes in the settings of their hearing aids. It is probable that some signal processing (e.g., noise reduction algorithms) might have affected the performance of EHA users, particularly IPs when supportive semantic context was lacking. We suggest that future studies compare audiovisual performance under simple linear amplification conditions and when some signal processing is active during the experiment. This may elucidate the extent to which advanced signal processing positively or negatively influences IPs at phonemic and lexical levels.

The between-subject comparison of IPs in audiovisual and auditory modalities seems to be a second limitation of the present study, as individual differences across participants (between-group comparisons) for stimuli presented in auditory and audiovisual modalities may influence IPs to some extent. A within-subject experimental design may provide more robust interpretations by controlling for individual differences. Nevertheless, within-group comparison of audiovisual and auditory speech stimuli may have its own drawback, as for instance, early exposure to multisensory stimuli subsequently boost unisensory processing of stimuli (for a review, see Shams, Wozny, Kim, & Seitz, 2011). In speech perception, evidence supporting this notion comes from our previous studies on young

normal-hearing listeners (Lidestam et al., 2014; Moradi et al., 2013) showing that prior exposure to audiovisual speech stimuli subsequently facilitated the auditory performance of participants, whereas prior exposure to auditory speech stimuli did not. We hypothesize that if the present study had been a within-subject design and the modality of presentation had been randomized across participants (e.g., half of the participants started with gated auditory task and the other half with gated audiovisual task), those who had been tested first in the audiovisual modality subsequently would have had shorter IPs and improved accuracy in the auditory modality. This improvement in auditory IPs and accuracies (caused by perceptual doping) may create a Type II error by generating non-significant differences in comparing IPs of a given speech task between the audiovisual and auditory modalities (unless the sample size had been increased). We suggest that future studies should consider these limitations caused by between-subject and within-subject experimental designs when comparing audiovisual and auditory speech stimuli.

Conclusions

The addition of visual cues to an amplified speech signal in the EHA group resulted in the same level of performance in terms of accuracy as the ENH group. However, in terms of IPs, the EHA users had inferior performance than their age-matched counterparts with normal hearing when a supportive semantic context was lacking. In addition, audiovisual presentation greatly speeded up the identification of speech stimuli relative to auditory-only presentation in the absence of a semantic context, in both the EHA and ENH groups. Nevertheless, the effect of audiovisual presentation was more evident in the EHA group as the accompanying visual cues (see Moradi et al. 2014a) helped the EHA users to disambiguate the speech signal.

Acknowledgments

The authors thank Carl-Fredrik Neikter, Amin Saremi, and Niklas Rönnerberg for their technical support; Mathias Hällgren and Helena Torlofson for their assistance during this study; and Katarina Marjanovic for speaking the recorded stimuli. The authors also thank Prof. Andrew Oxenham and two anonymous reviewers for their comments on this manuscript.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a grant from the Swedish Research Council (349-2007-8654).

References

- Ahlstrom, J. B., Horwitz, A. R., & Dubno, J. R. (2014). Spatial separation benefit for unaided and aided listening. *Ear & Hearing, 35*, 72–85.
- Arditi, A. (2005). Improving the design of the letter contrast sensitivity test. *Investigate Ophthalmology & Visual Science, 46*, 2225–2229.
- Aydelott, J., Leech, R., & Crinion, J. (2010). Normal adult aging and the contextual influences affecting speech and meaningful sound perception. *Trends in Amplification, 14*, 218–232.
- Baart, M., Vroomen, J., Shaw, K., & Bortfeld, H. (2014). Degrading phonetic information affects matching of audiovisual speech in adults, but not in infants. *Cognition, 130*, 31–43.
- Baskent, D., & Bazo, D. (2011). Audiovisual asynchrony detection and speech intelligibility in noise with moderate to severe sensorineural hearing impairment. *Ear & Hearing, 32*, 582–592.
- Bernstein, J. G. W., & Grant, K. W. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 125*, 3358–3372.
- Best, V., Ozmeral, E. J., & Shinn-Cunningham, B. G. (2007). Visually-guided attention enhances target identification in a complex auditory scene. *Journal of Association for Research in Otolaryngology, 8*, 294–304.
- Boothroyd, A., & Nitttrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *Journal of the Acoustical Society of America, 84*, 101–114.
- Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancement for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America, 121*, 2339–2349.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.
- Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology, 5*(7), e1000436.
- Corthals, P., Vinck, B., De Vel, E., & Van Cauwenberg, P. (1997). Audiovisual speech reception in noise and self-perceived hearing disability in sensorineural hearing loss. *Audiology, 36*, 46–56.
- de la Vaux, S. K., & Massaro, D. W. (2004). Audiovisual speech gating: Examining information and information processing. *Cognitive Processing, 5*, 106–112.
- Dimitrijevic, A., John, M. S., & Picton, T. W. (2004). Auditory steady-state responses and word recognition scores in normal-hearing and hearing-impaired adults. *Ear & Hearing, 25*, 68–84.
- Dufor, S., & Frauendelder, U. H. (2010). Phonological neighborhood effects in French-spoken word recognition. *Quarterly Journal of Experimental Psychology, 63*, 226–238.
- Elliott, L. L., Hammer, M. A., & Evan, K. E. (1987). Perception of gated, highly familiar spoken monosyllabic nouns by children, teenagers, and older adults. *Perception & Psychophysics, 42*, 150–157.
- Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech and Hearing Research, 12*, 423–425.
- Frisina, D. R., & Frisina, R. D. (1997). Speech recognition in noise and presbycusis: Relations to possible neural mechanisms. *Hearing Research, 106*, 95–104.
- Friston, K. J., & Kiebel, S. J. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 364*, 1211–1221.
- Gatehouse, S., & Gordon, J. (1990). Response time to speech stimuli as measure of benefit from amplification. *British Journal of Audiology, 24*, 63–68.
- Grant, K. W., & Seitz, P. F. (2000). The recognition of isolated words and words in sentences: Individual variability in the use of semantic context. *Journal of the Acoustical Society of America, 107*, 1000–1011.
- Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *Journal of the Acoustical Society of America, 103*, 2677–2690.
- Grosjean, F. (1980). Spoken word recognition processes and gating paradigm. *Perception & Psychophysics, 28*, 267–283.
- Hardison, D. M. (2005). Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Applied Psycholinguistics, 26*, 579–596.
- Houben, R., van Doorn-Bierman, M., & Dreschler, W. A. (2013). Using response time to speech as a measure of listening effort. *International Journal of Audiology, 52*, 753–761.
- Huckvale, M., & Leak, J. (2009). Effect of noise reduction on reaction time to speech in noise. *Proceedings of the 10th Annual Conference of the International Speech Communication Association* (pp. 1–4). Brighton, UK.
- Järpsten, B. (2002). *DLSTM handledning*. Stockholm, Sweden: Hogrefe Psykofoforlaget AB.
- Jesse, A., & Janse, E. (2012). Audiovisual benefit for recognition of speech presented with single-talker noise in older listeners. *Language and Cognitive Processes, 27*, 1167–1191.
- Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in psychtoolbox-3? *Perception, 36*.
- Lidestam, B. (2014). Audiovisual presentation of video-recorded stimuli at a high frame rate. *Behavior Research Methods, 46*, 499–516.
- Lidestam, B., Moradi, S., Petterson, R., & Ricklefs, T. (2014). Audiovisual training is better than auditory-only training for auditory only speech-in-noise identification. *Journal of the Acoustical Society of America, 136*, EL142–EL147.
- Lively, S. E., Pisoni, D. B., & Goldinger, S. D. (1994). Spoken word recognition: Research and theory. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 265–301). San Diego, CA: Academic Press.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing, 19*(1), 1–36.
- Lyxell, B., & Rönnerberg, J. (1989). Information-processing skill and speech-reading. *British Journal of Audiology, 23*, 339–347.

- Mars Perceptrix. (2003). *The mars letter contrast sensitivity test: User manual*. Chappaqua, NY: Author.
- Marslen-Wilson, W. D. (1993). Issues of process and representation in lexical access. In G. Altmann, & R. Shillcock (Eds.), *Cognitive models of language processes: The second sperlonga meeting*. Hove, England: Erlbaum.
- Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word-recognition in continuous speech. *Cognitive Psychology*, *10*, 29–63.
- Metsala, J. L. (1997). An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Memory & Cognition*, *25*, 47–56.
- Molis, M. R., Kappel, S. D., McMillan, G. P., Gallun, F. J., Dann, S. M., Konrad-Martin, D. (2015). Effects of hearing and aging on sentence-level time-gated word recognition. *Journal of Speech, Language, and Hearing Research*, *58*, 481–496.
- Moradi, S., Lidestam, B., Hällgren, M., & Rönnerberg, J. (2014a). Gated auditory speech perception in elderly hearing aid users and elderly normal-hearing individuals: Effects of hearing impairment and cognitive capacity. *Trends in Hearing*. doi:10.1177/2331216514545406.
- Moradi, S., Lidestam, B., & Rönnerberg, J. (2013). Gated audiovisual speech identification in silence vs. noise: Effects on time and accuracy. *Frontiers in Psychology*, *4*, 359. doi:10.3389/fpsyg.2013.00359.
- Moradi, S., Lidestam, B., Saremi, A., & Rönnerberg, J. (2014). Gated auditory speech perception: Effects of listening conditions and cognitive capacity. *Frontiers in Psychology*, *5*, 531. doi:10.3389/fpsyg.2014.00531.
- Pelli, D. G. (1997). The video toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Pichora-Fuller, M. K. (2008). Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. *International Journal of Audiology*, *47*(suppl.2): S72–S82.
- Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, *97*, 593–608.
- Picou, E. M., Ricketts, T. A., & Hornsby, B. W. Y. (2013). How hearing aids, background noise, and visual cues influence objective listening effort. *Ear & Hearing*, *34*, e52–e64.
- Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, *15*, 285–290.
- Rogers, C. S., Jacoby, L. L., & Sommers, M. S. (2012). Frequent false hearing by older adults: The role of age differences in metacognition. *Psychology and Aging*, *27*, 33–45.
- Salasoo, A., & Pisoni, D. (1985). Interaction of knowledge source in spoken word identification. *Journal of Memory and Language*, *24*, 210–231.
- Shams, L., Wozny, D. R., Kim, R., & Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing. *Frontiers in Psychology*, *2*, 264. doi:10.3389/fpsyg.2011.00264.
- Sheffield, B. M., Schuchman, G., & Bernstein, J. G. (2015). Trimodal speech perception: How residual acoustic hearing supplements cochlear-implant consonant recognition in the presence of visual cues. *Ear & Hearing*, *36*, e99–112.
- Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *Journal of the Acoustical Society of America*, *123*, 489–499.
- Singer, M., Bronstein, D. M., & Miles, J. M. (1981). Effect of noise on priming in a lexical decision task. *Bulletin of the Psychonomic Society*, *18*, 187–190.
- Smeele, P. M. T. (1994). *Perceiving speech: Integrating auditory and visual speech* (PhD dissertation). Delft University of Technology, The Netherlands.
- Sommers, M. S., & Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. *Psychology and Aging*, *14*, 458–472.
- Sommers, M. S., Tye-Murray, N., & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear & Hearing*, *26*, 263–275.
- Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, *26*, 212–215.
- Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audiovisual speech perception. In B. Dodd, & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 3–51). Hillsdale, NJ: Lawrence.
- Tye-Murray, N., Sommers, N. S., & Spehar, B. (2007a). Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing. *Ear & Hearing*, *28*, 656–668.
- Tye-Murray, N., Sommers, N. S., & Spehar, B. (2007b). Auditory and visual lexical neighborhoods in audiovisual speech perception. *Trends in Amplification*, *11*, 233–241.
- Van Engen, K. J., Phelps, J. E., Smiljanic, R., & Chandrasekaran, B. (2014). Enhancing speech intelligibility: Interactions among context, modality, speech style, and masker. *Journal of Speech, Language, and Hearing Research*, *57*, 1908–1918.
- van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of United States of America*, *102*, 1181–1186.
- Walden, B. E., Grant, K. W., & Cord, M. T. (2001). Effects of amplification and speechreading on consonant recognition by persons with impaired hearing. *Ear & Hearing*, *22*, 333–341.
- Walden, B. E., Prosek, R. A., & Worthington, D. W. (1975). Auditory and audiovisual feature transmission in hearing-impaired adults. *Journal of Speech, Language, and Hearing Research*, *18*, 272–280.
- Walley, A. C., Michela, V. L., & Wood, D. R. (1995). The gating paradigm: Effects of presentation format on spoken word recognition by children and adults. *Attention, Perception, & Psychophysics*, *57*, 343–351.