Audiovisual Asynchrony Detection and Speech Intelligibility in Noise With Moderate to Severe Sensorineural Hearing Impairment

Deniz Başkent[1,2] and Danny Bazo[3]

Objective: The objective of this study is to explore the sensitivity to intermodal asynchrony in audiovisual speech with moderate to severe sensorineural hearing loss. Based on previous studies, two opposing expectations were an increase in sensitivity, as hearing-impaired listeners heavily rely on lipreading in daily life, and a reduction in sensitivity, as hearing-impaired listeners tend to be elderly and advanced age could potentially impair audiovisual integration.

Design: Adults with normal (N = 11, ages between 23 and 50 yrs) and impaired hearing (N = 11, ages between 54 and 81 yrs, the pure-tone average between 42 and 67 dB HL) participated in two experiments. In the first experiment, the synchrony judgments were recorded for varying intermodal time differences in audiovisual sentence recordings. In the second experiment, the intelligibility of audiovisual and audio-only speech was measured in speech-shaped noise, and correlations were explored between the synchrony window and intelligibility scores for individual listeners.

Results: Similar to previous studies, a sensitivity window on the order of a few hundred milliseconds was observed with all listeners. The average window shapes did not differ between normal-hearing and hearing-impaired groups; however, there was large individual variability. Individual windows were quantified by Gaussian curve fitting. Point of subjective simultaneity, a measure of window peak shift from the actual synchrony point, and full-width at half-maximum, a measure of window duration, were not correlated with participant's age or the degree of hearing loss. Points of subjective simultaneity were also not correlated with speech intelligibility scores. A moderate negative correlation that was significant at most conditions was observed between the full-width at half-maximum values and intelligibility scores.

Conclusions: Contrary to either expectation per se, there was no indication of an effect of hearing impairment or age on the sensitivity to intermodal asynchrony in audiovisual speech. It is possible that the negative effects of aging were balanced with the positive effects of increased sensitivity due to reliance on visual cues with hearing impairment. The listeners, normal hearing or hearing impaired, who were more sensitive to asynchrony (with narrower synchrony windows) tended to understand speech in noise better, with both audio-only and audiovisual speech. The practical implication of the results is that delays in audio or video signals of communication systems would affect hearing-impaired listeners in a manner similar to normal-hearing listeners, and due to the importance of visual cues for the hearing-impaired listeners, special attention should be given to limit these delays.

(Ear & Hearing 2011;32;582–592)

[1]Starkey Hearing Research Center, Berkeley, California; [2]Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, Groningen, The Netherlands; and [3]Department of Electrical and Computer Engineering, University of California, Berkeley, California.

INTRODUCTION

Visual speech cues provide significant gain in speech understanding, especially in difficult listening conditions (Sumby & Pollack 1954; Erber 1969; Sanders & Goodrich 1971; Breeuwer & Plomp 1984; Helfer 1997; Schwartz et al. 2004; Helfer & Freyman 2005; Ross et al. 2007). An important factor in the perception of audiovisual speech is the cross-modal integration of information from audio and visual components. If the relative timing between the two is disrupted, integration may be negatively affected and the benefit from visual cues may be reduced (Campbell & Dodd 1980; Pandey et al. 1986; Munhall et al. 1996; Grant & Seitz 1998; Grant et al. 2004). Previous studies have shown that there is some tolerance to this disruption. In normal hearing, there is a synchrony window of a few hundred milliseconds, within which a time delay between the audio and visual speech signals cannot be detected, and integration occurs (McGrath & Summerfield 1985; Massaro et al. 1996; Munhall et al. 1996; Grant et al. 2004; Conrey & Pisoni 2006; van Wassenhove et al. 2007). The synchrony window is asymmetrical; it is more difficult for listeners to detect asynchrony when audio lags behind video. The asymmetry has been attributed mainly to the adaptation of the human auditory system to slower transmission of sound than light. As a result, audio and visual signals pertaining to the same event can be integrated efficiently, despite the difference in the time of arrival (Dixon & Spitz 1980; Summerfield 1992; Spence & Squire 2003; Sugita & Suzuki 2003; Kopinska & Harris 2004; Vatakis & Spence 2006).

Hearing-impaired listeners have also been shown to integrate audio and visual speech and benefit from lipreading (Middelweerd & Plomp 1987; Braida 1991; Bosman & Smoorenburg 1997; Grant et al. 1998; Bernstein & Grant 2009). However, little research has been done on the sensitivity to audiovisual asynchrony in this population. Intermodal asynchrony can be especially important for those hearing-impaired listeners who rely on visual cues as a crucial aid to speech understanding in daily life. Signal processing delays in hearing devices or asynchrony between audio and video in telecommunication devices, TV broadcast, and movies may have a disruptive effect on speech perception by this population (Summerfield 1992; Reeves & Voelker 1993; Liu & Sato 2009). Hearing impairment may affect sensitivity to audiovisual asynchrony in two opposing ways. Hearing-impaired listeners, especially with moderate to severe levels of impairment, rely heavily on visual cues in everyday life to compensate for the poorer speech intelligibility. The increased reliance may result in better use of visual cues and better integration of audio and visual speech (McGrath & Summerfield 1985;

0196/0202/11/3205-0582/0 • Ear & Hearing • Copyright © 2011 by Lippincott Williams & Wilkins • Printed in the U.S.A.

Mohammed et al. 2005; Auer & Bernstein 2007; Tye-Murray et al. 2007), even though hearing-impaired listeners have not always shown better lipreading skills than normal-hearing listeners (Lyxell & Rönnberg 1989). Due to the increased demand on the visual system, there could also be functional changes in the brain organization. Evidence for such cross-modal plasticity and cortical reorganization was shown with blind people (for example, with sharper auditory spatial tuning; Röder et al. 1999) and with congenitally deaf listeners (for example, with enhanced perception of motion and peripheral stimuli; Bavelier et al. 2000; Mitchell & Maslin 2007). With this reasoning, our first expectation would be an increased sensitivity to audiovisual asynchrony with moderate to severe degrees of hearing impairment. A similar reasoning also applies to cochlear implant users; they similarly rely on visual cues to compensate for the distorted speech input from their devices, and this reliance is considered to be one of the reasons for enhanced multisensory integration observed in implant users (Giraud & Truy 2002; Rouger et al. 2007; Strelnikov et al. 2009). However, despite the enhanced audiovisual integration and increased reliance on visual cues (Desai et al. 2008; Rouger et al. 2008), one study showed no difference in the sensitivity to intermodal asynchrony between implant users and normal-hearing listeners (Hay-McCutcheon et al. 2009). What made a difference was the age of the listeners. Elderly listeners in each listener group tended to have longer synchrony windows, which was linked to reduced audiovisual integration observed with the elderly in some studies (Musacchia et al. 2009). Because many hearing-impaired listeners are older, our second and opposing expectation would then be a reduced sensitivity to asynchrony in hearing-impaired listeners.

The general goal of the present study was to gain further insight on audiovisual speech perception in sensorineural hearing impairment and provide results that could be useful in designing communication devices for the hearing impaired. The specific goals were to explore whether sensorineural hearing loss of moderate to severe levels (and accompanying factor of aging) would have an effect on the sensitivity to bimodal asynchrony in audiovisual speech and whether individual sensitivity windows would be correlated to audiovisual speech perception, as the asynchrony sensitivity window seems to be closely related to the window of integration for audio and visual speech (Grant & Seitz 1998; Conrey & Pisoni 2006).

MATERIALS AND METHODS

In the first experiment, we measured sensitivity to audiovisual asynchrony using a synchrony judgment task. The stimulus onset asynchrony was varied systematically and the participants reported on perceived synchrony. In the second experiment, we measured speech perception in noise with audio-only and audiovisual sentences and explored the correlations between intelligibility scores and individual asynchrony sensitivity. The same group of listeners participated in both experiments, and there were some similarities in the methodology. Therefore, this section explains the methods for both experiments.

Participants

As hearing impairment is more common among elderly, the hearing-impaired participants in the present study were older than the normal-hearing participants. Eleven normal-hearing listeners (six females and five males), aged between 23 and 50 yrs (average 36 yrs), and 11 hearing-impaired listeners (eight females and three males), aged between 54 and 81 yrs (average 71 yrs), participated in the study. The number of participants was determined based on the study by Hay-MacCutcheon et al. (2009), where an effect of age was observed on bimodal asynchrony detection with similar numbers of participants. All listeners were native monolingual speakers of American English. The pure-tone average (PTA; the average of hearing thresholds at the audiometric frequencies of 500, 1000, and 2000 Hz) was used in the selection of the listeners. The inclusion criteria for normal hearing were to have a PTA <15 dB HL (according to Table 5.4 in the handbook by Katz & Gabbay 1994) and to have hearing thresholds ≤ 20 dB HL at the audiometric frequencies ≤ 4000 Hz. The hearing impaired listeners had symmetrical and postlingual sensorineural hearing loss of moderate to severe levels, with PTAs ranging from 42 to 67 dB HL, with an average of 54 dB HL. Figure 1 shows the audiometric thresholds of the participants, in the left panel, and the ages of the listeners as a function of the PTA, in the right panel, and Table 1 shows further details on individual hearing-impaired listeners.

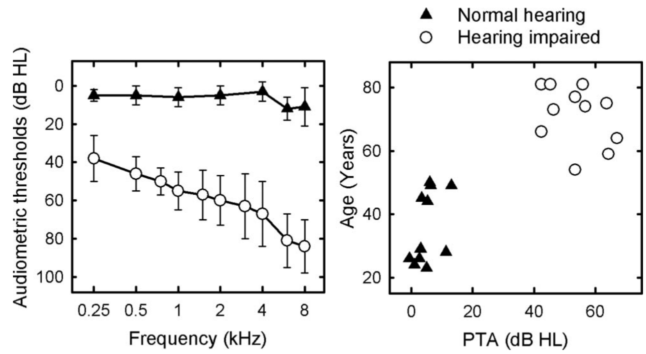

Fig. 1. Audiometric thresholds of the participants (left panel) and the age of the participants as a function of the pure tone average (PTA) (right panel). The error bars show 1SD.

Two hearing-impaired listeners, who originally qualified for participation, were ultimately excluded from the study. One listener was 90 yrs old and the other had a PTA of 83 dB HL, and despite acceptable speech understanding in quiet, the former scored near 0% in both audio-only and audiovisual tests and the latter scored near 0% in all audio-only tests of the second experiment.

All listeners were fully informed about the study, and written informed consent was collected before their participation. The study was carried out in accordance with the National Institutes of Health regulations and ethical guidelines on experimentation with human subjects.

Stimuli and Signal Processing

Audio and video signals of the stimuli were separated and converted into .wav file (44 kHz sampling rate) and Apple Quicktime movie (29.97 frames/sec, 720 × 480 frame size) formats, respectively. All audio stimuli were equalized in dB in root mean square, calculated using sentence portions only, without the silence before and after.

In the first experiment, for the asynchrony detection, 50 sentences from the Audiovisual Lexical Neighborhood Sen-

TABLE 1. Detailed information about the hearing-impaired participants

Participant Number	PTA (dB HL)	Age (yrs)	Age at the Onset of Deafness (yrs)	Duration of Deafness (yrs)	Etiology	Hearing Aid Use	Full-Width at Half-Maximum (msecs)
HI1	46	73	52	21	Unknown, progressive	None	278
HI2	53	77	49	28	Unknown, progressive	Bilateral BTEs	346
HI3	42	81	Unknown	>14	Unknown, progressive	Bilateral BTEs	386
HI4	64	59	53	6	Unknown, progressive	Bilateral CICs	432
HI5	45	81	Unknown	Unknown	Unknown, progressive	Bilateral BTEs	372
HI6	64	75	50	25	Viral infection, sudden	Bilateral BTEs	376
HI7	67	64	37	27	Noise exposure and presbycusis, progressive	Bilateral CICs	324
HI8	42	66	Unknown	>10	Presbycusis, progressive	Bilateral BTEs	210
HI9	53	54	27	27	Unknown, progressive	Bilateral BTEs	342
HI10	56	81	Unknown	>8	Unknown, progressive	Bilateral BTEs	328
HI11	57	74	Unknown	>9	Unknown, progressive	Bilateral BTEs	258

The rightmost column lists individual values for the full-width at half-maximum from Experiment 1.

tence Test (Reference Note 1), spoken by one female talker, were used. The sentences in this database are lexically controlled and the lists are equalized for difficulty. The asynchrony between the audio and video signals was produced by changing the onset time of the audio signal with respect to the video.

In the second experiment, for speech recognition in noise, sentences from build-a-sentence database (Tye-Murray et al. 2008), spoken by one female talker, were used. In each list, there are 12 meaningless sentences, forming a closed set of 36 words. A steady speech-shaped noise was produced with Matlab software by averaging the spectra of all audio speech stimuli in the build-a-sentence database and randomizing the phase.

Procedure

The entire procedure was completed in one session, in 2 to 2.5 hrs. Listeners were seated in a sound-treated booth. The video was presented on a 3M touchscreen monitor. The audio stimuli were routed through the SPDIF output of an M-Audio Delta AP soundcard and Lavry DA10 D/A converter and presented diotically over Sennheiser HD-580 headphones. Max/MSP software (from Cycling '74) was used for presenting the stimuli, collecting the responses, and storing the results for offline analysis. There was an internal asynchrony jitter in the experimental set up, which was minimized by optimization of software settings for best synchrony between audio and video signals, as well as turning off all automatic background processes of the computer. With this optimization, the jitter was minimized to values less than one frame length (33 msecs), with a SD of 15 msecs over a 100 measurements.

In the first experiment, sensitivity to intermodal asynchrony was measured with a subjective judgment task (Conrey & Pisoni 2006; Hay-McCutcheon et al. 2009). Stimulus onset asynchrony was introduced at 15 audio delay values, varying from −210 to 330 msecs, based on the results from previous studies (Dixon & Spitz 1980; Munhall et al. 1996; Grant et al. 2004; Conrey & Pisoni 2006; Jones & Jarick 2006; van Wassenhove et al. 2007; Hay-McCutcheon et al. 2009; Navarra et al. 2010). The step size was set to 30 msecs or larger to further minimize the effect of the internal asynchrony jitter of the computer on the results. The participant and the experimenter, seated inside and outside the booth, respectively, saw the same screen on each of their monitors. The participant controlled the pace of the experiment, while the experimenter monitored the progress. Participants were asked to watch and listen to one stimulus at a time and answer a single question, namely whether the audio and video of the stimulus were synchronized. They chose a "yes" or "no" response by pressing the appropriate button on the touchscreen monitor. No catch trials were included. In each block, all 15 temporal conditions were tested in random order and with one sentence each. There were 10 blocks, with a total of 150 trials (10 sentences per condition). One of the 50 sentences was randomly picked for each trial, and due to the greater number of trials than the number of sentences, some sentences were repeated during data collection.

In the second experiment, speech intelligibility was measured by counting the keywords that were correctly repeated by the listeners. Signal-to-noise ratios (SNRs) of −5 and −10 dB were used. These values were selected for a number of reasons. (1) Because we wanted to explore correlations with individual synchrony windows, variability in intelligibility scores caused by background noise was desirable. (2) Using relatively high noise levels, we intended to mask some portions of audio speech and so make the participants rely more on the visual cues. (3) Ross et al. (2007) showed that most benefit from lipreading is observed for moderate noise levels. (4) A pilot study showed that most normal-hearing and hearing-impaired listeners produced reasonable speech perception scores with least amount of floor and ceiling effects with these SNRs. In this part of the experiment, a split screen was used between the touchscreen monitor inside the booth and the monitor outside the booth. The participant only saw the video while the experimenter only saw the text of the words presented to the participant. The participant and the experimenter communicated via an audiometer. The participants verbally reported the words they heard after each sentence was presented, and the experimenter marked the correctly identified keywords. For familiarization, before data collection, a short run at a low-level noise (SNR = 10 dB) was completed with one list of 12 audiovisual sentences. During data collection, one list of 12 sentences was used for each of the four experimental noise conditions (audio-only speech at SNR = −5 dB and −10 dB and audiovisual speech with SNR = −5 dB and −10 dB), resulting in a total of 48 trials. In this experiment, no sentence was repeated.

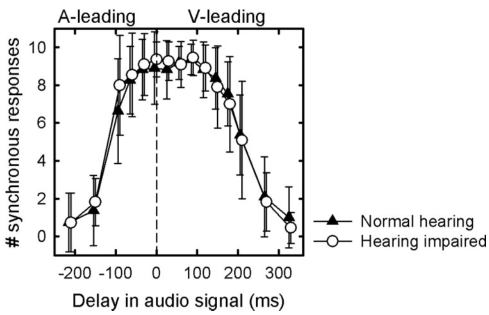

Fig. 2. Synchrony windows, averaged for normal-hearing and hearing-impaired listeners and shown as a function of the temporal conditions. The vertical dash bar shows the reference for the perfect synchronous condition. The error bars show 1SD.

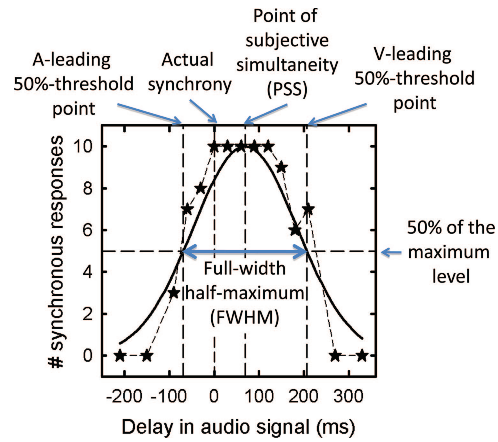

Fig. 3. Gaussian curve fitting (solid line) shown with data from one listener (stars connected with short dashes).

Presentation Levels

For normal-hearing listeners, the presentation level for the speech stimuli was 60 dB SPL (measured at the output of the headphones with an artificial ear coupler). For hearing-impaired listeners, 60 dB SPL was the level before linear amplification was applied within the Max/MSP software. The individual frequency shaping was based on each participant's audiometric thresholds and the NAL-R formula (Dillon 2001). A simple linear frequency shaping was preferred to minimize potential distortions from nonlinear amplification. Uncomfortable loudness was prevented by first limiting the gain in each frequency to 40 dB and then asking the listeners to adjust the overall volume to a loud but comfortable level. During this adjustment, participants were given ample time to listen to a variety of stimuli, all similar but not identical to the experimental ones, with varying background noise conditions. Once the appropriate presentation level was found, it was no longer changed throughout the data collection.

RESULTS

Experiment 1: Sensitivity to Audiovisual Asynchrony

Figure 2 shows the average subjective judgment results for asynchronous audiovisual stimuli as a function of the stimulus onset asynchrony. The negative and positive delays respectively refer to the conditions where the audio preceded the video signal (A-leading) and where the video preceded the audio signal (V-leading). The y axis shows the number of the times that the listener reported the audio and video signals to be synchronous (out of 10 trials per condition).

The curves in the figure show that there was a synchrony window of a few hundred milliseconds, within which the listeners could not detect asynchrony. The window is asymmetrically situated around the 0-msec delay, the actual point of synchrony, indicating that asynchrony was more difficult to detect in A-leading conditions. The main interest of the present study was, however, the comparison of the results between normal-hearing and hearing-impaired listeners. Visually, the curves from the two groups seemed to be similar. A two-way mixed analysis of variance with one within-subject factor (delay; 15 levels) and one between-subject factor (listener group; two levels) confirmed that there was no significant difference in the curves. There was a significant main effect of the delay ($F[14,280] = 117.794$, $p < 0.001$, Cohen's $f = 2.43$, indicating a large effect size) but no significant main effect of listener group ($F[1,20] = 0.035$, $p = 0.854$, Cohen's $f = 0.042$, indicating a negligible effect size) and no significant interaction between the factors ($F[14,280] = 0.640$, $p = 0.831$, Cohen's $f = 0.179$, indicating a small effect size). At the true synchrony point of 0-msec delay, the average number of "synchronous" responses was not 10, the theoretical maximum, as was expected. A close inspection of individual scores showed that listeners had the peak values of 10 at different delays, not necessarily always at 0 msec. As a result, when the scores were averaged at 0-msec delay, they were 8.9 and 9.4 for normal-hearing and hearing-impaired listeners, respectively. The peaks in the curves averaged for each listener group occurred at the 90-msec delay, with values of 9.4 and 9.5. A post hoc Tukey HSD test showed that the scores from −60 to 150 msecs did not differ significantly from the scores at 0 msec or from the peak values at 90 msecs.

For further analysis, synchrony windows were quantified with a method described by Conrey and Pisoni (2006), and Hay-McCutcheon et al. (2009). First, a Gaussian curve was fitted to each listener's individual results by minimizing root mean square error (Fig. 3). The delay where the maximum value of each curve occurred (the mean of the Gaussian curve) produced the point of subjective simultaneity (PSS). Then, both on the increasing (A-leading) and decreasing (V-leading) sides, the 50% levels with respect to the maximum curve value determined the 50% threshold points. The distance between the 50% threshold points was the full-width at half-maximum (FWHM), the quantitative measure of the synchrony sensitivity window. Table 2 shows the A-leading and V-leading 50% threshold points, and PSS and FWHM values, averaged for normal-hearing and hearing-impaired listeners. As the last row of the table shows, there was no significant difference in these values between the listener groups. Despite the similarity in average scores between the groups, however, there was a large variation in individual results within the groups, as shown by the range of the individual values presented in each cell of the table.

PSS is a measure of deviation from the actual synchrony point of 0 msec. FWHM is a measure of the synchrony window. These values were accepted as measures of sensitivity to asynchrony, and the potential factors that may have caused the variability were explored with correlational analysis.

TABLE 2. Estimated values from Gaussian curve fitting

Listener Group	A-Leading 50% Threshold Point (msecs)	Point of Subjective Simultaneity (PSS; msecs)	V-Leading 50% Threshold Point (msecs)	Full-Width at Half-Maximum (FWHM; msecs)
Normal hearing	−196 to −69	11–69	141–280	258–476
	−108 (40)	46 (18)	203 (39)	311 (69)
Hearing impaired	−179 to −42	23–66	152–253	210–432
	−122 (35)	47 (14)	210 (34)	332 (63)
Comparison: two-tailed *t* test	$t(20) = 0.874$	$t(20) = 0.053$	$t(20) = 0.451$	$t(20) = 0.745$
	$p = 0.393$	$p = 0.958$	$p = 0.657$	$p = 0.465$
	Cohen's $d = 0.391$ (small effect size)	Cohen's $d = 0.065$ (negligible effect size)	Cohen's $d = 0.201$ (small effect size)	Cohen's $d = 0.333$ (small effect size)

In each cell, the first line shows the range of the values (from minimum to maximum), and the second line shows the average and the standard deviation (in parentheses). The bottom row shows the results from a two-tailed t test that compared the estimates between normal-hearing and hearing-impaired groups.

Age, Hearing Loss, and Asynchrony Sensitivity

Age (in years) and hearing loss (in PTA [dB HL]) were explored as potential factors causing the variability in individual audiovisual asynchrony sensitivity with a correlation analysis. Age was analyzed within each group as well as both groups combined. PTA was analyzed for hearing-impaired listeners only. Figure 4 shows the corresponding regression lines superimposed with individual PSS and FWHM values. Table 3 shows the Pearson Product Moment correlations. Combined, they show that the correlations between asynchrony sensitivity measures and hearing loss or age were weak or nonexistent and not significant.

FWHM values for individual hearing-impaired listeners were listed in the rightmost column of Table 1 for further inspection. Visually, there seemed to be no relationship between the FWHM values and age, PTA, etiology of deafness, or hearing-aid usage of hearing-impaired participants. One listener (participant 4) with the relatively short duration of deafness (6 yrs) had the longest FWHM value; however, as there were many other listeners who were not certain about the onset of their hearing loss, it is difficult to further speculate on the effect of duration of deafness on audiovisual asynchrony sensitivity.

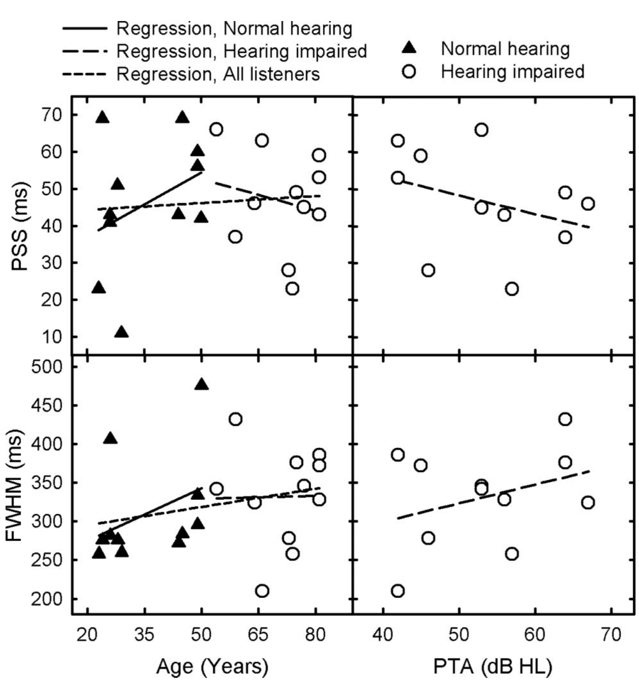

Fig. 4. Individual values for point of subjective simultaneity (PSS) and full-width at half-maximum (FWHM), shown in the upper and lower panels, respectively, as a function of all listener's ages (left panels) and hearing-impaired listeners' pure tone averages (PTAs; right panels).

Hay-McCutcheon et al. (2009) showed no effect of age on PSS but an effect of age on FWHM. For a direct comparison with their results, we have further analyzed our data. The age ranges for their middle-aged and elderly groups were between 41 to 55 and 65 to 81 yrs, respectively. We have similarly picked listeners inside these age brackets disregarding hearing modality. Thus, six and eight listeners were in each age group with mostly normal hearing in the six-person group and all hearing impaired in the other group. When we averaged the FWHMs for the groups separately, there was no significant difference (t test with unequal samples; $t[12] = 0.401$, $p = 0.698$, Cohen's $d = 0.24$, indicating a small effect size): 334 ± 75 msecs and 319.25 ± 64 msecs for middle-aged and elderly listeners, respectively.

Experiment 2: Speech Intelligibility in Noise and Its Correlation to Audiovisual Asynchrony

We have explored correlations between speech perception in noise and asynchrony sensitivity for both listeners groups. Due to many scores at floor and ceiling, the percent correct scores were transformed to rationalized arcsine units (RAU; Studebaker 1985). Figures 5 and 6 show the RAU scores for speech intelligibility in background noise as a function of individual PSS and FWHM, respectively. In each figure, the top and bottom panels show the scores with audio-only and audiovisual speech, and the left and right panels show the scores for background noise at SNR = −10 and −5 dB, respectively. In each panel, the solid and dashed lines show the regression for normal-hearing and hearing-impaired listeners. Figure 5 additionally shows the average speech perception scores for each listener group and for each listening condition on the left side of each panel. Tables 4 and 5 show the Pearson Product Moment correlations between PSS or FWHM, respectively, and RAU scores.

The average results, shown in the left side of the panels in Figure 5, indicate that the speech intelligibility in noise was significantly higher with normal-hearing listeners than hearing impaired listeners in all conditions (t test; $t[20] > 4.411$, $p \leq 0.001$, Cohen's $d > 0.80$, indicating a large effect size). The difference in average RAU scores varied from 26 to 43. A further inspection of individual scores showed that only two

TABLE 3. Pearson Product Moment correlations for the age and pure-tone average (PTA) factors

Factor	Correlation With Point of Subjective Simultaneity (PSS)		Correlation With Full-Width at Half-Maximum (FWHM)	
	r	*p*	*r*	*p*
Age (normal hearing only)	0.366	0.268	0.372	0.259
Age (hearing impaired only)	−0.195	0.565	0.021	0.950
Age (normal hearing and hearing impaired combined)	0.084	0.710	0.250	0.261
Hearing impairment (PTA) (hearing impaired only)	−0.335	0.314	0.345	0.299

hearing-impaired participants (HI1 and HI8) had audio-only scores at both noise levels similar to that of the normal-hearing control group. Both listeners had low PTAs (46 and 42 dB HL, respectively), but they were not among the youngest of the hearing-impaired group (73 and 66 yrs old, respectively). Interestingly, regardless of the differences in performance, both groups of listeners showed substantial improvement in intelligibility when visual cues were added; even the hearing impaired listeners who were at floor level at SNR = −10 dB with audio-only speech showed an improvement of 50 to 70 RAU with audiovisual speech.

The regression lines in Figure 5, combined with correlations in Table 4, show that there was no correlation between the PSS measure of asynchrony and speech perception in noise. The regression lines in Figure 6, combined with correlations in Table 5, show that there was a moderate negative correlation between the FWHM measure of asynchrony (synchrony window duration) and speech perception in noise. The correlation was significant in most listening conditions. At SNR = −5 dB, with both audio and audiovisual speech, the correlation was not significant with normal-hearing listeners, but this may have been caused by the ceiling effect.

Note that the hearing-impaired group was 54 yrs and older, while the normal-hearing control group was 50 yrs and younger (Fig. 1). Due to this separation in age, the interpretations of Figures 5 and 6 and Tables 4 and 5 would be the same if the results were analyzed for younger and older listeners, instead of normal-hearing and hearing-impaired listeners.

DISCUSSION

In the present study, we have explored sensitivity to audiovisual asynchrony with audiovisual sentences and speech perception in noise with audio-only and audiovisual speech by listeners with moderate to severe sensorineural hearing loss. A group of young normal-hearing listeners served as the control. In the first experiment, sensitivity to intermodal asynchrony was measured with a synchrony judgment task. Average and (estimated) individual synchrony windows were analyzed for hearing impairment and age. In the second experiment, the correlations were explored between individual synchrony windows and speech intelligibility.

Sensitivity to Audiovisual Asynchrony

Similar to previous findings, overall results showed that asynchronies up to a few hundred milliseconds could not be detected (Dixon & Spitz 1980; Munhall et al. 1996; Grant et al. 2004; Conrey & Pisoni 2006; Jones & Jarick 2006; van Wassenhove et al. 2007; Hay-McCutcheon et al. 2009; Navarra et al. 2010).

The main interest of the present study was whether the asynchrony sensitivity would differ with hearing impairment of moderate to severe levels. There were two opposing arguments for why this difference should occur. On one hand, the heavier reliance of hearing-impaired listeners on visual cues could make them more sensitive to asynchrony (Tye-Murray et al. 2007). On the other hand, as hearing-impaired listeners tend to

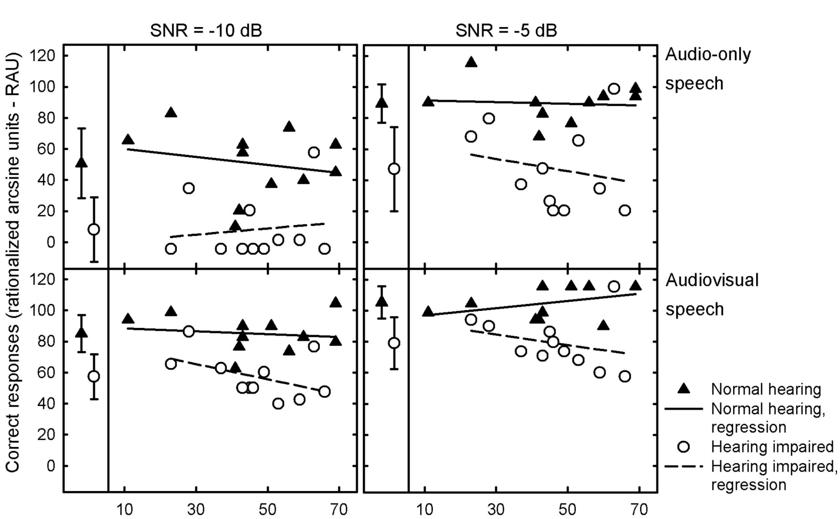

Fig. 5. Individual speech intelligibility scores in arcsine transformed units (RAU), superimposed with regression lines and shown as a function of the point of subjective simultaneity (PSS), for audio-only and audiovisual speech (top and bottom panels) and for varying background noises (left and right panels). The scores to the left of each panel show the average intelligibility scores for each listener group with 1SD.

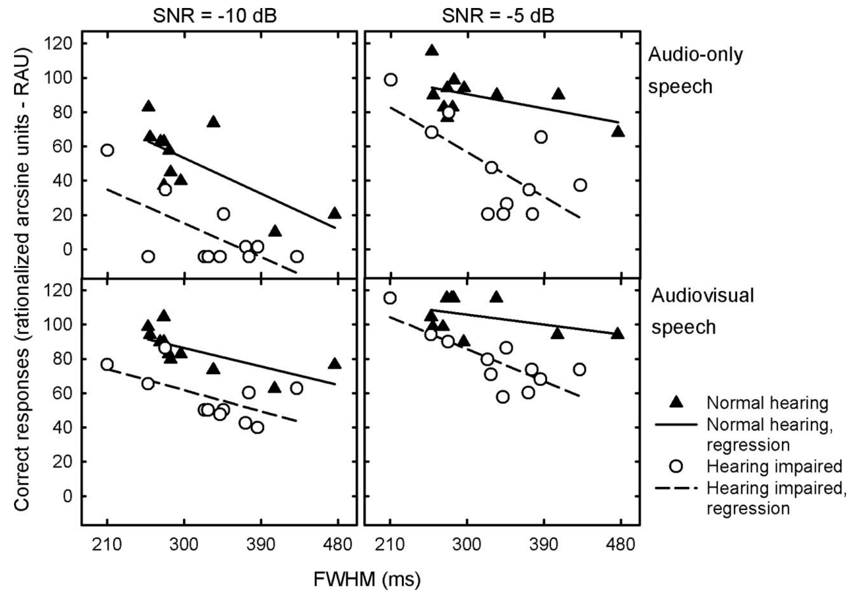

Fig. 6. Individual speech intelligibility scores and regression lines shown as a function of the full-width at half-maximum (FWHM).

be elderly and some of the previous studies have shown deficiency in audiovisual integration with aging, they could be less sensitive (Tanaka et al. 2007; Hay-McCutcheon et al. 2009; Musacchia et al. 2009). Contrary to each argument per se, the present study showed no difference in average asynchrony sensitivity between young normal-hearing and relatively older hearing-impaired listeners. For further analysis, individual synchrony window shapes were quantified with a Gaussian curve fitting (Conrey & Pisoni 2006; Hay-McCutcheon et al. 2009). The PSSs, a measure of perceived synchrony, were similar to those reported by Conrey and Pisoni (2006) and smaller than those reported by Hay-McCutcheon et al. (2009). The A-leading and V-leading 50% threshold points and the FWHMs, a measure of synchrony window width, were comparable to, but slightly smaller than, that measured by Conrey and Pisoni (2006) and Hay-McCutcheon et al. (2009). The small differences in results between the present study and the previous ones could be due to differences in the speech materials and experimental set-up. First, our listeners were instructed specifically to look for an asynchrony and were pointed out to cases where it may be easier to detect (such as observing “plosives,” which are easy to see on the video and to detect in the audio due to the sharp onset). We have used sentences (in contrast to isolated words used in previous studies), which may have given the participants more such samples to judge the asynchrony. Second, the fewer temporal conditions of the present study may have further contributed to narrower Gaussian curve fits.

Between normal-hearing and hearing-impaired groups, the average values of the 50% thresholds, PSSs, and FWHMs were similar, but the individual values varied greatly within each group. Hearing loss and age were considered as potential experimental factors contributing to the variability; however, a correlational analysis indicated no effect of either factor. In a similar study with cochlear implant users, Hay-McCutcheon et al. (2009) had observed no effect of hearing modality but a significant effect of age on FWHM values and asynchrony curves. For a direct comparison with this study, a final analysis was conducted with a subgroup of listeners, selected to replicate the age ranges reported by Hay-McCutcheon et al. (2009), which also failed to show an effect of age. The finding should be interpreted with caution, however, as the number of data points used in this analysis was very low. These results are somewhat puzzling, as the present study design was based on the study by Hay-McCutcheon, for example, in choosing the number of participants and using a similar subjective judgment task. There are a number of factors that may have affected the results of both studies; the potential response bias (due to the lack of catch trials) and the low statistical power. In the present study, the effect sizes were small or negligible; hence, even if

TABLE 4. Correlations between the percent correct scores (in RAU) for perception of speech in noise and the PSS (in msec), shown for both audio-only and audiovisual speech stimuli, and for both normal hearing and hearing impaired listeners (listed separately)

	SNR = −10 dB		SNR = −5 dB	
	r	*p*	*r*	*p*
Audio only				
Normal hearing	−0.207	0.541	−0.076	0.824
Hearing impaired	0.129	0.705	−0.198	0.559
Audiovisual				
Normal hearing	−0.136	0.691	0.399	0.225
Hearing impaired	−0.458	0.157	−0.275	0.412

TABLE 5. Correlations between the percent correct scores (in RAU) for perception of speech in noise and the FWHM (in msec)

	SNR = −10 dB		SNR = −5 dB	
	r	*p*	*r*	*p*
Audio only				
Normal hearing	−0.715	0.013	−0.523	0.099
Hearing impaired	−0.666	0.025	−0.678	0.022
Audiovisual				
Normal hearing	−0.696	0.018	−0.429	0.188
Hearing impaired	−0.599	0.051	−0.790	0.038

the power was increased and a significance was shown with very large numbers of participants, the practical consequences of such small effects could be minimal. Moreover, these two factors should have affected the two studies in similar ways and could not explain the differing findings. A third factor, however, is that using sentences as stimuli may have made the task of detecting asynchrony relatively easier for the elderly listeners of the present study. Nevertheless, mixed effects of aging have been observed with other studies as well; while a number of methods have shown a negative effect of age, others have shown none (Cienkowski & Carney 2002; Sommers et al. 2005; Tanaka et al. 2007; Tye-Murray et al. 2007, 2008; Musacchia et al. 2009).

Our results, in short, indicated no effect of hearing loss or age on asynchrony sensitivity. We had purposely selected listeners with relatively severe hearing loss (moderate to severe levels instead of mild to moderate) who would have difficulty understanding speech in daily life and would therefore have to rely on visual cues heavily. It is possible that this reliance at more severe degrees of hearing loss compensates for the negative effects of aging. The present study was not designed to show the effects from hearing loss and age separately, but it showed that (relatively) elderly and hearing-impaired listeners were as sensitive to asynchrony as young and normal-hearing listeners.

Note that there were a number of factors produced by the experimental design that may have inadvertently affected the present results. One such confounding factor was the potential subject bias in the subjective judgment task. As there were no catch trials included in this yes-no task, the measurements were not bias-free, and they most likely reflected both the individual sensitivity to asynchrony and the individual response criterion. Previous studies indicated that elderly people may have different response criteria than younger people (Gordon-Salant 1986; Ratcliff et al. 2001). Differing biases in choosing "yes/no" answers between the (younger) normal-hearing and (elderly) hearing-impaired listeners of the present study may have hidden potential differences in asynchrony sensitivity. With the current design of the subjective judgment task, the effects from sensitivity and bias could not be separated. A second confounding factor was the inherent asynchrony jitter of the experimental set-up, which could have additionally contributed to the variability in individual scores. However, this should have affected the measurements similarly for each group and have therefore minimal to no effect on the comparisons of the group data.

Speech Intelligibility in Noise

On average, speech perception in noise was better with normal-hearing listeners than hearing impaired (and elderly) listeners, consistent with the literature (Dubno et al. 1984; Horst 1987; Jerger et al. 1991). Hearing-impaired listeners benefited substantially from visual cues. Even the listeners who had no speech understanding with audio-only speech (e.g., at the SNR = −10 dB level) had relatively high scores with the audiovisual speech—much higher than would be expected from lipreading alone. Hence, even when the audio signal by itself provided no intelligibility, its combination with visual cues has produced a synergistic effect, possibly due to the complementary nature of the audio and visual speech cues (Binnie et al. 1974; Erber 1979; Summerfield & Assmann 1987; Grant et al. 1998; Robert-Ribes et al. 1998; Grant & Seitz 2000).

The main interest of the present study was in the correlations between the asynchrony sensitivity and speech intelligibility in noise, particularly in hearing impairment. Synchrony window can be perceived as a measure of the time during which the bimodal information is bound together perceptually as an individual event (Munhall et al. 1996; Kopinska & Harris 2004). Grant and Seitz (1998) assumed that good lipreaders must be more attentive to efficiently extract information from visual cues and combine it with the associated speech movements, and therefore, they would be more susceptible to a disruption in the timing due to asynchrony.

Consequently, asynchrony sensitivity can be closely linked with lipreading benefit and audiovisual speech intelligibility, an idea partially supported in previous studies. While McGrath & Summerfield (1985) observed better asynchrony sensitivity by good lipreaders, Grant and Seitz (1998) and Conrey and Pisoni (2006) observed no correlation between asynchrony sensitivity and visual speech perception with hearing-impaired listeners and normal-hearing listeners tested with temporally distorted speech. However, both studies, as well as the study by Hay-McCutcheon et al. (2009) with middle-aged normal-hearing and cochlear implant listeners, indicated a negative correlation between asynchrony sensitivity and audiovisual speech perception.

Our results are in partial agreement with previous studies. We similarly observed moderate negative correlation between asynchrony sensitivity and audiovisual speech perception in noise, within each group of younger normal-hearing and older hearing-impaired listeners. What differed was that we also observed this correlation with audio-only speech perception. One could argue that the nonoptimal audibility due to reduced dynamic range in hearing impairment produced this observation; not being able to hear speech adequately may affect both tasks of (audio or audiovisual) speech recognition and asynchrony detection (Grant & Seitz 2000). However, we rule out this possibility, as the results with normal-hearing listeners showed the same trend, except for the conditions under which the performance was at ceiling levels. The assumption by Grant and Seitz (1998), mentioned earlier, can only explain the results with the audiovisual speech. Our results are more consistent with an alternative idea that Grant and Seitz (1998) mentioned, namely that there is a single underlying construct for speech perception that potentially uses the same resources and mechanisms, such as linguistic knowledge, context, cognitive function, and attention. This idea is further supported by newer studies that showed that audiovisual and audio-only speech perception is more tightly coupled than previously thought (Von Kriegstein et al. 2008; Bishop & Miller 2009). Therefore, sensitivity and performance in all speech-recognition tasks should be correlated (Bilger 1984; Watson et al. 1996; Olsen et al. 1997; Auer 2010). In the present study, several factors related to the experimental design of the intelligibility test, such as the lack of context in the sentences and using moderate and varying levels of noise, may have emphasized the variability in individual intelligibility scores—it is possible that these findings would not be observed in easier speech recognition tests (such as with highly contextual materials and conducted in quiet). The variability in scores could have contributed to the establishment of stronger correlations of the present study. In addition, due to the elimination of context and minimized speech redundancy, it is possible that the scores truly reflected an inherent ability of speech recognition system to decode audio or visual information, with minimal help from such resources.

There are other factors that could have caused a longer synchrony window with listeners who had poorer speech intelligibility in noise. The unitary speech perception mechanism idea (Bilger 1984) implies that poorer performers of the present study would have more difficulty understanding speech in general. Reduced asynchrony sensitivity can in fact be a lengthening in audiovisual integration window to compensate for this difficulty. With training, both audiovisual speech intelligibility and audiovisual asynchrony detection can be improved (Montgomery et al. 1984; Richie & Kewley-Port 2008; Kawase et al. 2009; Powers et al. 2009). Therefore, it is possible that as the difficulty that the listeners experience in understanding speech decreases due to training, the synchrony window becomes narrower. In addition, increased effort due to difficulty understanding audio speech may reduce the effort and attention that the listener can put into detecting visual cues necessary for detection of asynchrony (Summerfield 1987, 1991). Contrary to the idea of audiovisual integration being preattentive (McGurk & MacDonald 1976; Soto-Faraco et al. 2004), a number of studies show an affect of attention on multimodal integration. For example, Lesner and Hardick (1982) implied a correlation between visual attention and lipreading skill, and recent studies have shown the effects of selective attention and attentional load on audiovisual integration (Alsius et al. 2005; Talsma & Woldorff 2005; Fujisaki & Nishida 2008; Talsma et al. 2009; Navarra et al. 2010), indirectly supporting our explanation.

As a final note, the correlations observed in the present study and the aforementioned interpretations should be taken cautiously. Despite our efforts to make a careful experimental design, one should remember that the asynchrony sensitivity depends on many factors, such as the experimental method used (van Eijk et al. 2008), recalibration due to awareness of the source distance (Stone et al. 2001; Sugita & Suzuki 2003; Fujisaki et al. 2004; Kopinska & Harris 2004; Vroomen et al. 2004; Vroomen & Keetels 2010) or due to continuous exposure to asynchronous audiovisual stimuli (Navarra et al. 2005), shortening due to training (Powers et al. 2009), or changes due to certain disorders (Virsu et al. 2003; Hamilton et al. 2006; Foucher et al. 2007; Giersch et al. 2009). Therefore, to establish generality, as well as to reveal underlying neural mechanisms of the behavioral observations (Stevenson et al. 2010), more studies with different experimental designs are needed.

Practical Implications

Audiovisual speech tests have not gained widespread popularity as part of routine procedures used in audiology clinics, even though hearing-impaired listeners and users of hearing devices are often in situations where visual speech is available (Woodhouse et al. 2009). There is a need to understand audiovisual speech perception with hearing-impaired listeners, so that appropriate procedures can be developed and a more realistic assessment of device usage can be made. Our data indirectly suggest that the hearing device benefit could be underestimated with audio-only testing.

The main contribution of the present study to practical considerations, however, is showing the similarity in sensitivity to intermodal delays in audiovisual speech between normal-hearing and hearing-impaired listeners. One of the motivations in studying these delays is that, if hearing impaired listeners were more sensitive, then the delays in hearing aids, cochlear implants, or communication or multimedia devices would have to be re-evaluated, especially because visual cues may carry more importance for hearing-impaired listeners for understanding speech (McGrath & Summerfield 1985; Pandey et al. 1986). The processing delays in hearing aids and cochlear implants are small in modern devices and certainly within the synchrony window duration shown in this and previous studies (Stone & Moore 1999). In multimedia applications and video communication devices, however, there could still be considerable intermodal asynchrony that cannot be entirely eliminated due to technical limitations (Bloom 1985; Shah & Marshall 1994; Chen et al. 1995; Chen & Rao 1998; Finger & Davis 1998; Zonja et al. 2006) to the degree that special standards had to be developed. For example, the ITU-T (1990) standard specifies an audio lead and lag of no more than 45 and 125 msecs, respectively. Our data suggest that hearing-impaired (and elderly) listeners are as sensitive to intermodal asynchrony as young and normal-hearing listeners, and special attention needs to be given to control for such delays.

ACKNOWLEDGMENTS

The authors thank Rachael Holt, Brent Spehar, and Matt Burk for audiovisual speech materials; Erin Riley and Nazanin Nooraei for contacting and screening the participants; Dan Steele for help in developing the testing program; Wiebe Horst and Anastasios Sarampalis for comments on an earlier version of the article; and the listeners for their participation.

Danny Bazo is currently at the Media Arts and Technology Graduate Program, University of California, Santa Barbara, 3309 Phelps Hall, Santa Barbara, CA 93106, USA.

This work was supported by Starkey Labs.

Presented in part at the 2009 Midwinter Meeting by the Association for Research in Otolaryngology, Baltimore, MD, USA; and the 2009 International Forum for Hearing Instrument Developers, Oldenburg, Germany.

Address for correspondence: Dr. Deniz Başkent, Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, P.O. Box 30.001, 9700 RB Groningen, The Netherlands. E-mail: d.baskent@med.umcg.nl.

Received May 18, 2010; accepted January 4, 2011.

REFERENCES

Alsius, A., Navarra, J., Campbell, R., et al. (2005). Audiovisual integration of speech falters under high attention demands. *Curr Biol*, *15*, 839–843.

Auer, E. T. (2010). Investigating speechreading and deafness. *J Am Acad Audiol*, *21*, 163–168.

Auer, E. T., & Bernstein, L. E. (2007). Enhanced visual speech perception in individuals with early-onset hearing impairment. *J Speech Lang Hear Res*, *50*, 1157–1165.

Bavelier, D., Tomann, A., Hutton, C., et al. (2000). Visual attention to the periphery is enhanced in congenitally deaf individuals. *J Neurosci*, *20*, 1–6.

Bernstein, J., & Grant, K. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *J Acoust Soc Am*, *125*, 3358–3372.

Bilger, R. (1984). Speech recognition test development. *ASHA Reports*, *14*, 2–7.
Binnie, C. A., Montgomery, A. A., Jackson, P. L. (1974). Auditory and visual contributions to the perception of consonants. *J Speech Lang Hear Res*, *17*, 619–630.
Bishop, C. W., & Miller, L. M. (2009). A multisensory cortical network for understanding speech in noise. *J Cogn Neurosci*, *21*, 1790–1804.
Bloom, P. (1985). High-quality digital audio in the entertainment industry: An overview of achievements and challenges. *IEEE Acoust Speech Signal Process Mag*, *2*, 2–25.
Bosman, A., & Smoorenburg, G. (1997). Speechreading supplemented with auditorily presented speech elements in the profoundly hearing impaired. *Int J Audiol*, *36*, 29–45.
Braida, L. (1991). Crossmodal integration in the identification of consonant segments. *Q J Exp Psychol A*, *43*, 647–677.
Breeuwer, M., & Plomp, R. (1984). Speechreading supplemented with frequency-selective sound-pressure information. *J Acoust Soc Am*, *76*, 686–691.
Campbell, R., & Dodd, B. (1980). Hearing by eye. *Q J Exp Psychol*, *32*, 85–99.
Chen, T., Graf, H., Wang, K. (1995). Lip synchronization using speech-assisted video processing. *IEEE Sig Proc Letters*, *2*, 57–59.
Chen, T., & Rao, R. (1998). Audio-visual integration in multimodal communication. Proc IEEE, *86*, 837–852.
Cienkowski, K., & Carney, A. (2002). Auditory-visual speech perception and aging. *Ear Hear*, *23*, 439–449.
Conrey, B., & Pisoni, D. B. (2006). Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *J Acoust Soc Am*, *119*, 4065–4073.
Desai, S., Stickney, G., Zeng, F. (2008). Auditory-visual speech perception in normal-hearing and cochlear-implant listeners. *J Acoust Soc Am*, *123*, 428.
Dillon, H. (2001). *Hearing Aids.* New York, NY: Thieme.
Dixon, N., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception*, *9*, 719–721.
Dubno, J., Dirks, D., Morgan, D. (1984). Effects of age and mild hearing loss on speech recognition in noise. *J Acoust Soc Am*, *76*, 87–96.
Erber, N. (1979). Auditory-visual perception of speech with reduced optical clarity. *J Speech Lang Hear Res*, *22*, 212–223.
Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *J Speech Hear Res*, *12*, 423–425.
Finger, R., & Davis, A. (1998). Measuring video quality in videoconferencing systems. Wainhouse Research Whitepaper.
Foucher, J., Lacambre, M., Pham, B., et al. (2007). Low time resolution in schizophrenia: Lengthened windows of simultaneity for visual, auditory and bimodal stimuli. *Schizophr Res*, *97*, 118–127.
Fujisaki, W., & Nishida, S. (2008). Top-down feature-based selection of matching features for audio-visual synchrony discrimination. *Neurosci Lett*, *433*, 225–230.
Fujisaki, W., Shimojo, S., Kashino, M., et al. (2004). Recalibration of audiovisual simultaneity. *Nat Neurosci*, *7*, 773–778.
Giersch, A., Lalanne, L., Corves, C., et al. (2009). Extended visual simultaneity thresholds in patients with schizophrenia. *Schizophr Bull*, *35*, 816–825.
Giraud, A., & Truy, E. (2002). The contribution of visual areas to speech comprehension: A PET study in cochlear implants patients and normal-hearing subjects. *Neuropsychologia*, *40*, 1562–1569.
Gordon-Salant, S. (1986). Effects of aging on response criteria in speech-recognition tasks. *J Speech Hear Res*, *29*, 155–162.
Grant, K., & Seitz, P. (1998). Measures of auditory-visual integration in nonsense syllables and sentences. *J Acoust Soc Am*, *104*, 2438–2450.
Grant, K., & Seitz, P. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *J Acoust Soc Am*, *108*, 1197–1208.
Grant, K., van Wassenhove, V., Poeppel, D. (2004). Detection of auditory (cross-spectral) and auditory-visual (cross-modal) synchrony. *Speech Commun*, *44*, 43–53.
Grant, K., Walden, B., Seitz, P. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *J Acoust Soc Am*, *103*, 2677–2690.
Hamilton, R., Shenton, J., Coslett, H. (2006). An acquired deficit of audiovisual speech processing. *Brain Lang*, *98*, 66–73.
Hay-McCutcheon, M., Pisoni, D., Hunt, K. (2009). Audiovisual asynchrony detection and speech perception in hearing-impaired listeners with cochlear implants: A preliminary analysis. *Int J Audiol*, *48*, 321–333.
Helfer, K., & Freyman, R. (2005). The role of visual speech cues in reducing energetic and informational masking. *J Acoust Soc Am*, *117*, 842–849.
Helfer, K. S. (1997). Auditory and auditory-visual perception of clear and conversational speech. *J Speech Lang Hear Res*, *40*, 432–443.
Horst, J. W. (1987). Frequency discrimination of complex signals, frequency selectivity and speech perception in hearing-impaired subjects. *J Acoust Soc Am*, *82*, 874–885.
ITU-T (1990). Television and sound transmission: Tolerances for transmission time differences between the vision and sound components of a television signal. International Telecommunication Union, Telecommunication standardization sector of ITU, Recommendation J.100, CMTT 717 in CCIR Recommendations, 12.
Jerger, J., Jerger, S., Pirozzolo, F. (1991). Correlational analysis of speech audiometric scores, hearing loss, age, and cognitive abilities in the elderly. *Ear Hear*, *12*, 103–109.
Jones, J., & Jarick, M. (2006). Multisensory integration of speech signals: The relationship between space and time. *Exp Brain Res*, *174*, 588–594.
Katz, J., & Gabbay, W. L. (1994). *Handbook of Clinical Audiology.* Baltimore, MD: Williams & Wilkins.
Kawase, T., Sakamoto, S., Hori, Y., et al. (2009). Bimodal audio-visual training enhances auditory adaptation process. *Neuroreport*, *20*, 1231–1234.
Kopinska, A., & Harris, L. (2004). Simultaneity constancy. *Perception*, *33*, 1049–1060.
Lesner, S., & Hardick, E. (1982). An investigation of spontaneous eye blinks during lipreading. *J Speech Hear Res*, *25*, 517–520.
Liu, Y., & Sato, Y. (2009). Recovery of audio-to-video synchronization through analysis of cross-modality correlation. *Pattern Recogn Lett*, *31*, 696–701.
Lyxell, B., & Rönnberg, J. (1989). Information-processing skill and speech-reading. *Br J Audiol*, *23*, 339–347.
Massaro, D., Cohen, M., Smeele, P. (1996). Perception of asynchronous and conflicting visual and auditory speech. *J Acoust Soc Am*, *100*, 1777–1786.
McGrath, M., & Summerfield, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *J Acoust Soc Am*, *77*, 678–685.
McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.
Middelweerd, M., & Plomp, R. (1987). The effect of speechreading on the speech-reception threshold of sentences in noise. *J Acoust Soc Am*, *82*, 2145–2147.
Mitchell, T., & Maslin, M. (2007). How vision matters for individuals with hearing loss. *Int J Audiol*, *46*, 500–511.
Mohammed, T., Campbell, R., MacSweeney, M., et al. (2005). Speechreading skill and visual movement sensitivity are related in deaf speechreaders. *Perception*, *34*, 205–216.
Montgomery, A., Walden, B., Schwartz, D., et al. (1984). Training auditory-visual speech reception in adults with moderate sensorineural hearing loss. *Ear Hear*, *5*, 30–36.
Munhall, K., Gribble, P., Sacco, L., et al. (1996). Temporal constraints on the McGurk effect. *Percept Psychophys*, *58*, 351–362.
Musacchia, G., Arum, L., Nicol, T., et al. (2009). Audiovisual deficits in older adults with hearing loss: Biological evidence. *Ear Hear*, *30*, 505–514.
Navarra, J., Alsius, A., Soto-Faraco, S., et al. (2010). Assessing the role of attention in the audiovisual integration of speech. *Inform Fusion*, *11*, 4–11.
Navarra, J., Alsius, A., Velasco, I., et al. (2010). Perception of audiovisual speech synchrony for native and non-native language. *Brain Res*, *1323*, 84–93.
Navarra, J., Vatakis, A., Zampini, M., et al. (2005). Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Brain Res Cogn Brain Res*, *25*, 499–507.
Olsen, W., Van Tasell, D., Speaks, C. (1997). Phoneme and word recognition for words in isolation and in sentences. *Ear Hear*, *18*, 175–188.
Pandey, P., Kunov, H., Abel, S. (1986). Disruptive effects of auditory signal delay on speech perception with lipreading. *J Aud Res*, *26*, 27–41.

Powers, A., III, Hillock, A., Wallace, M. (2009). Perceptual training narrows the temporal window of multisensory binding. *J Neurosci*, *29*, 12265–12274.
Ratcliff, R., Thapar, A., McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychol Aging*, *16*, 323–341.
Reeves, B., & Voelker, D. (1993). Effects of audio-video asynchrony on viewer's memory, evaluation of content and detection ability. Research Report Prepared for Pixel Instruments, Los Gatos, CA, USA.
Richie, C., & Kewley-Port, D. (2008). The effects of auditory-visual vowel identification training on speech recognition under difficult listening conditions. *J Speech Lang Hear Res*, *51*, 1607–1619.
Robert-Ribes, J., Schwartz, J., Lallouache, T., et al. (1998). Complementarity and synergy in bimodal speech: Auditory, visual, and audio-visual identification of French oral vowels in noise. *J Acoust Soc Am*, *103*, 3677–3689.
Röder, B., Teder-Sälejärvi, W., Sterr, A., et al. (1999). Improved auditory spatial tuning in blind humans. *Nature*, *400*, 162–166.
Ross, L., Saint-Amour, D., Leavitt, V., et al. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cereb Cortex*, *17*, 1147–1153.
Rouger, J., Fraysse, B., Deguine, O., et al. (2008). McGurk effects in cochlear-implanted deaf subjects. *Brain Res*, *1188*, 87–99.
Rouger, J., Lagleyre, S., Fraysse, B., et al. (2007). Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proc Natl Acad Sci USA*, *104*, 7295–7300.
Sanders, D., & Goodrich, S. (1971). The relative contribution of visual and auditory components of speech to speech intelligibility as a function of three conditions of frequency distortion. *J Speech Lang Hear Res*, *14*, 154–159.
Schwartz, J., Berthommier, F., Savariaux, C. (2004). Seeing to hear better: Evidence for early audio-visual interactions in speech identification. *Cognition*, *93*, B69–B78.
Shah, D., & Marshall, S. (1994). Multi-modality coding system for videophone application. WIASIC '94, Berlin, Germany, October 1994.
Sommers, M., Tye-Murray, N., Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear Hear*, *26*, 263–275.
Soto-Faraco, S., Navarra, J., Alsius, A. (2004). Assessing automaticity in audiovisual speech integration: Evidence from the speeded classification task. *Cognition*, *92*, B13–B23.
Spence, C., & Squire, S. (2003). Multisensory integration: Maintaining the perception of synchrony. *Curr Biol*, *13*, 519–521.
Stevenson, R., Altieri, N., Kim, S., et al. (2010). Neural processing of asynchronous audiovisual speech perception. *Neuroimage*, *49*, 3308–3318.
Stone, J., Hunkin, N., Porrill, J., et al. (2001). When is now? Perception of simultaneity. *Proc R Soc Lond B Biol Sci*, *268*, 31–38.
Stone, M. A., & Moore, B. C. J. (1999). Tolerable hearing-aid delays. I. Estimation of limits imposed by the auditory path alone using simulated hearing losses. *Ear Hear*, *20*, 182–192.
Strelnikov, K., Rouger, J., Barone, P., et al. (2009). Role of speechreading in audiovisual interactions during the recovery of speech comprehension in deaf adults with cochlear implants. *Scand J Psychol*, *50*, 437–444.
Studebaker, G. (1985). A "rationalized" arcsine transform. *J Speech Hear Res*, *28*, 455–462.
Sugita, Y., & Suzuki, Y. (2003). Audiovisual perception: Implicit estimation of sound-arrival time. *Nature*, *421*, 911.
Sumby, W. H., & Pollack, I. (1954). Visual contributions to speech intelligibility in noise. *J Acoust Soc Am*, *26*, 212–215.
Summerfield, Q. (1987). Some Preliminaries to a Comprehensive Account of Audio-Visual Speech Perception. In B. Dodd & R. Campbell. *Hearing by Eye: The Psychology of Lip-Reading* (pp. 3–51). London: Lawrence Eribaum.
Summerfield, Q. (1991). Visual Perception of Phonetic Gestures. In I. Mattingly & M. Studdert-Kennedy. *Modularity and the Motor Theory of Speech Perception* (pp. 117). Hillsdale, NJ: Lawrence Erlbaum.
Summerfield, Q. (1992). Lipreading and audio-visual speech perception. Philos Trans R Soc Lond B Biol Sci, *335*, 71–78.
Summerfield, Q., & Assmann, P. (1987). Auditory Enhancement in Speech Perception. In M. E. H. Schouten (Ed). *The Psychophysics of Speech Perception* (pp. 140–150). Dordrecht: Martinus Nijhoff.
Talsma, D., Senkowski, D., Woldorff, M. (2009). Intermodal attention affects the processing of the temporal alignment of audiovisual stimuli. *Exp Brain Res*, *198*, 313–328.
Talsma, D., & Woldorff, M. (2005). Selective attention and multisensory integration: Multiple phases of effects on the evoked brain activity. *J Cogn Neurosci*, *17*, 1098–1114.
Tanaka, A., Sakamoto, S., Tsumura, K., et al. (2007). Effects of intermodal timing difference and speed difference on intelligibility of auditory-visual speech in younger and older adults. Proceedings of the International Conference on Auditory-Visual Speech Processing 2007, 258–263.
Tye-Murray, N., Sommers, M., Spehar, B. (2007). Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing. *Ear Hear*, *28*, 656–668.
Tye-Murray, N., Sommers, M., Spehar, B., et al. (2008). Auditory-visual discourse comprehension by older and young adults in favorable and unfavorable conditions. *Int J Audiol*, *47*, 31–37.
van Eijk, L., Kohlrausch, A., Juola, J., et al. (2008). Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Percept Psychophys*, *70*, 955–968.
van Wassenhove, V., Grant, K., Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, *45*, 598–607.
Vatakis, A., & Spence, C. (2006). Audiovisual synchrony perception for music, speech, and object actions. *Brain Res*, *1111*, 134–142.
Virsu, V., Lahti-Nuuttila, P., Laasonen, M. (2003). Crossmodal temporal processing acuity impairment aggravates with age in developmental dyslexia. *Neurosci Lett*, *336*, 151–154.
Von Kriegstein, K., Dogan, O., Grüter, M., et al. (2008). Simulation of talking faces in the human brain improves auditory speech recognition. *Proc Natl Acad Sci USA*, *105*, 6747–6752.
Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Atten Percept Psychophys*, *72*, 871–884.
Vroomen, J., Keetels, M., de Gelder, B., et al. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cogn Brain Res*, *22*, 32–35.
Watson, C., Qiu, W., Chamberlain, M., et al. (1996). Auditory and visual speech perception: Confirmation of a modality independent source of individual differences in speech recognition. *J Acoust Soc Am*, *100*, 1153–1162.
Woodhouse, L., Hickson, L., Dodd, B. (2009). Review of visual speech perception by hearing and hearing-impaired people: Clinical implications. *Int J Lang Commun Disord*, *44*, 253–270.
Zonja, S., Livun, N., Jambrosic, K. (2006). Audio-visual interaction: Multimedia applications. 48th International Symposium ELMAR-2006 focused on Multimedia Signal Processing and Communications, 143–146.

REFERENCE NOTE

1. Holt, R., Kirk, K., Hay-McCutcheon, M. (in press). Assessing multimodal spoken word-in-sentence recognition in children with normal hearing and children with cochlear implants. *J Speech Lang Hear Res*.