Point-light facial displays enhance comprehension of speech in noise

Lawrence D. Rosenblum, Jennifer A. Johnson
University of California, Riverside

Helena M. Saldaña
Indiana University

Published as: Rosenblum, L.D., Johnson, J. A., & Saldaña, H.M. (1996). Visual kinematic information for embellishing speech in noise. Journal of Speech and Hearing Research 39(6), 1159-1170.

Abstract

Seeing a talker's face can improve the perception of speech in noise. There is little known about which characteristics of the face are useful for enhancing the degraded signal. In this study, a point-light technique was employed to help isolate the salient kinematic aspects of a visible articulating face. In this technique, fluorescent dots were arranged on the lips, teeth, tongue, cheeks, and jaw of an actor. The actor was videotaped speaking in the dark, so that when shown to observers, only the moving dots were seen. To test whether these reduced images could contribute to the perception of degraded speech, noise-embedded sentences were dubbed with the point-light images at various signal-to-noise ratios. It was found that these images could significantly improve comprehension for adults with normal hearing, and that the images became more effective as participants gained experience with the stimuli. These results have implications for uncovering salient visual speech information as well as the development of telecommunication systems for listeners who are hearing-impaired.

Visible articulatory information can be an important component for speech perception. Seeing the face of a speaker can be critical to individuals with hearing impairments as well as to those with normal hearing when faced with a noisy environment (e.g., MacLeod & Summerfield, 1987). Visual speech information can also help enhance[1] auditory speech spoken with a heavy foreign accent or which conveys difficult subject matter (Reisberg, McLean, & Goldfield, 1987).

A significant amount of research has measured the extent to which visual speech information can enhance auditory speech. Much of this work has tested whether visual speech can improve recovery of speech embedded in noise for normal hearing listeners (e.g., Erber, 1969; MacLeod & Summerfield, 1987; 1990; Middleweerd & Plomp, 1987; Sumby and Pollack, 1954). This research has revealed that visual speech information can enhance speech in noise so that it functionally improves signal to noise ratios (SNRs) as much as 22 dB (e.g., Sumby & Pollack, 1954). These studies have also revealed that this enhancement increases as the SNR of the auditory component decreases (e.g., Middleweerd & Plomp, 1987).

Although it has been established that seeing an articulating face improves speech perception in noise, little is known about which aspects of visual speech information lead to increased comprehension. In an attempt to address this issue, Summerfield (1979) tested degraded visual images in the speech-in-noise paradigm. Sentences embedded in noise were presented to observers with normal hearing without visual stimuli. These sentences were presented in synchrony with four different types of visual stimuli. The four types of visual stimuli included: (a) a fully illuminated face, (b) isolated lips (painted with luminous makeup) against an all black background, (c) four small, isolated glowing points on the lips (center of top and bottom lips and the corners of the mouth) against a black background, and (d) an annular display (shown on an

oscilloscope) whose diameter was correlated with the amplitude of the audio signal. Using a percentage difference criterion, Summerfield found that only the fully illuminated face and the isolated lips were able to significantly improve comprehension over the auditory alone condition. Most observers easily recognized the image with the four luminous points as a moving mouth. However, many observers also reported that the point locations did not precisely specify speaking lips even for fully-frontal bilabials.

Summerfield's (1979) study was an excellent first attempt to isolate the critical aspects of the visual information for speech-in-noise enhancement. However, we believe that a greater number of luminous points more strategically placed may provide more information about the critical aspects of the visual signal. Recent research has revealed that both speech and non-speech events can be identified with 'point-light faces'. For these demonstrations, luminous points were placed on facial features of a darkened face. Facial movements (e.g., articulations) were then video recorded under special lighting conditions. The videotapes were then displayed on a contrast-adjusted monitor so that only the moving points were seen. Recent research has shown that facial expressions (Bassili, 1978), age related 'person qualities' (Berry, 1990), and viseme categories (Rosenblum, Johnson, & Saldaña, 1995) can be recovered from these images. The point-light technique has also been applied to the recognition of full-body events (e.g., Bingham, 1987; Johansson, 1973; Kozlowski & Cutting, 1977; Runeson & Frykholm, 1981; Verfaillie, De Troy, & Van Rensbergen, 1994).

The point-light technique helps demonstrate the importance of kinematic dimensions of visual speech events (where kinematics refer to the time-bound dimensions of events such as displacement, velocity, and acceleration). Point-light displays specify events while portraying motion only: no 'pictorial' information such as skin texture or color is present. In fact, all of the aforementioned point-light stimuli are not recognizable when shown to observers statically (i.e., when the videotape is 'paused' to a single frame). This is an important point because visual speech information is most often described in terms of static features such as visual information for place of constriction; open, closed, or rounded lips; and visible teeth (e.g., McGrath, 1985; Montgomery and Jackson, 1983; Petajan, 1984; Summerfield & McGrath 1984). However, research in our laboratory suggests that the isolated kinematic information for speech provided by point-light images can also be salient. We have found that observers are able to distinguish between viseme categories (e.g. /b/ and /p/ were identifiably different from /v/ and /f/) with point-light images comprised of 19 dots (Rosenblum et al., 1995). In addition, Rosenblum and Saldaña (1996) found that point-light stimuli (comprised of 28 points) can visually influence discrepant auditory syllables so that they 'sound' like the visual syllables.

There are additional benefits of the point-light technique for lipreading research. First, the technique allows for the determination of salient visual speech features through precise manipulation of the visual features available in a test stimulus. Additionally, point-light images afford a more straight-forward kinematic analysis: it is much simpler to track the motion of a small point than it is to track a particular skin patch on, for example, the lips. Finally, the point-light technique might be useful for the development of telecommunication systems. It has been suggested that point-light images, unlike fully-illuminated faces, might afford transmission through the low bandwidth of a simple telephone line (Massaro, 1987; Pearson, 1981; see also, Tartter & Knowlton, 1981). If these images are found to convey rich linguistic information, they could be implemented to enhance noisy auditory signals and to help listeners with hearing impairment telecommunicate more effectively. In fact, Erber (1980) used a similar technique in placing reflective dots on the lips and testing 10 children with severe hearing impairment. Erber concluded that his results indicate ". . . that optical speech information may be drastically reduced with only moderate decrement in A-V efficiency and encourage development of a prototype system for transmitting optical components of speech over ordinary telephone lines" (p. 49).

Given the practical and conceptual utility of point-light stimuli, it seems useful to re-examine, using more than four points, whether these images can improve the comprehension of speech-in-noise. Because point-light stimuli can be informative with regard to viseme categories (Rosenblum et al., 1995) and can influence perception when auditory and visual information conflict

(Rosenblum & Saldaña, 1996), we expected that these stimuli would significantly enhance speech in noise. For our study, normal hearing participants were presented with sentences embedded in noise accompanied by one of four visual presentation conditions including a fully illuminated face and three different point-light configurations. The different point-light configurations were chosen to determine which array might be helpful in enhancing degraded speech.

In order to test visual improvement, sentences were presented in noise using MacLeod and Summerfield's (1987) ascending method of limits methodology[2]. This methodology involves presenting each sentence at increasing SNRs until listeners are able to recognize three keywords. From this, a Speech Reception Threshold in Noise (SRTN) is derived based on the minimal SNR at which all three key words are recognized. This methodology avoids many of the problems (e.g., ceiling and floor effects) inherent in past studies that measured functional SNR change with percentage scores. In adopting this presentation technique together with the use of more strategically placed points, the current study extends the work of Summerfield (1979) in determining whether point-light images can enhance speech in noise.

In addition, the effects of experience with the visual display will be evaluated. Even if there is sufficient information in visual speech kinematic displays, it might be that some degree of perceptual learning is needed for utilization of full point-light specification. Rosenblum and Saldaña (1996) found evidence that experience helps observers attune to the relevant kinematic information available in these displays (see McGrath, 1985, for a similar argument concerning reduced facial displays). Perceptual learning might also be an important component for the effectiveness of point-light stimuli in embellishing speech in noise. In order to examine this issue, a learning effects analysis will be conducted to determine if there was an increase in effectiveness from experience with the stimuli.

Method

Main Experiment

Participants

Twenty-five undergraduates (15 female and 15 male) were given class credit and were paid five dollars for their participation. They were native English speakers and were screened for normal hearing and normal vision. The screenings revealed that none of the participants had worse than 20-20 visual acuity (as assessed by a Snellen chart), and none had pure tone thresholds greater than 20 dB HL (re ANSI, 1989) at frequencies between 250 and 6000 Hz in the tested ear (as assessed by a Beltone audiometer). Their ages ranged from 18 to 24 years.

Materials

Sentences

Sixty short, simple sentences (6 to 7 syllables each) were selected as stimuli. These sentences were derived from sentence lists developed for speech in noise tests (MacLeod & Summerfield, 1987; Nilsson, Soli, & Sullivan, 1994) based on Bench & Bamford's (1979) BKB sentences[3]. Because five presentation conditions were tested (see below), the sentences were divided into five lists (12 sentences each) of equal levels of lipreading difficulty, based on an earlier pilot experiment (see Appendix). In the pilot experiment, each sentence was presented with only a video image of the same fully-illuminated articulating face used in the main experiment. Twenty native-English speaking undergraduates with self-reported normal hearing and vision were used as pilot participants. These participants were asked to write down any words they were able to lipread. Each sentence was scored using the Loose Keyword Scoring method (Bench & Bamford, 1979), where a point is given for each of the three key words recognized (morphological errors permitted). A total score was obtained for each sentence, equaling the total number of keywords recognized by 20 participants (out of 60 possible points). The five sentence lists were developed by distributing the sentences so that all lists had a nearly equal degree of total lipreading difficulty, as assessed from the total of the scores for sentences (see Appendix).

The sentence lists were based on stimuli that were different from those used in the actual experiment (see below). Although the pilot experiment used visual-only stimuli comprised of a fully-illuminated face, the stimuli of the actual experiment consisted of mostly audiovisual speech

in noise presentations often comprised of reduced (point-light) visual images. Furthermore, there is evidence that list equivalency varies from lipreading-only to audiovisual speech in noise contexts (Hinkle, 1979). Given these concerns, analyses were conducted to determine whether relative performance on the lists interacted with the various presentation conditions.

Visual stimuli

An American-English speaking Caucasian male actor was videotaped with a Panasonic PVS350 camcorder from a distance of five feet. The actor's head was secured by a metal brace to inhibit movement. In the recorded image, the actor's entire head was visible against a black cloth background.

The actor was told to maintain a normal and even intonation, tempo, and vocal intensity while producing the sentences. He was not however, instructed to enunciate in a way to make his articulations easy to speechread. During recording, the experimenters monitored the actor's speaking rate and intensity (using a sound level meter) in an effort to keep the sentences similar across the various visual conditions. Informal inspection of the auditory recordings indicated that the sentences did not differ substantially between visual conditions.

Four types of visual stimuli were recorded. For the Fully-Illuminated stimuli, the actor was recorded with full overhead (fluorescent) lighting along with a single spotlight situated about four feet in front of and focusing on the actor's face. No special effort was made to illuminate the inside of the actor's mouth; however, the spotlight was directed so that minimal shadowing occurred. For the three types of point-light stimuli, the actor's tongue was painted black with food coloring and his teeth were painted black with Mehron ToothFX tooth paint. Construction paper dots were made with a 3 mm hole punch and were painted with yellow Wildfire Ultra-Violet Sensitive Fluorescent Paint in Brilliant Yellow. These dots were applied to features of the face with medical adhesive (on the face) and dental adhesive (on the teeth and tongue). The dots were small enough so that they did not interfere with the actor's articulations. The actor was illuminated with two Black Light (Fluorescent) 24 inch, 10 watt bulbs held vertically three feet away and at a 45 degree angle to the side/front of his face: no other lighting was used. This lighting technique produced recorded images in which only the dots and their movements could be seen.

Three types of point-light stimuli were tested. The selection of point-light positions was based on previous research that used the technique for speech and nonspeech experiments (e.g., Bassili, 1978; Berry, 1990; Brooke & Summerfield, 1983; Summerfield, 1979; Rosenblum and Saldaña, 1996; Rosenblum et al., 1995). The three point-light stimulus types will be referred to as (a) Lips Lights, (b) Lips, Teeth, and Tongue Lights, and (c) All Lights. For the Lips Lights stimuli, 14 dots were arranged on the lips. A point was placed at each corner of the mouth. Six point-lights were placed around the inner edge of the mouth where the lips come together; the remaining 6 point-lights were placed in corresponding locations on the outer edges of the lips. For the Lips, Teeth, and Tongue Lights stimuli, 14 point-lights were applied as for the Lips Lights stimuli, with the addition of one point-light on the tip-of-the-tongue and 4 point-lights on the two front teeth on the bottom and top rows (19 points total). The point-lights on the teeth were applied toward the edge of the tooth on the inner corner. For the All Lights stimuli, point-lights were applied as for the Lips, Teeth, and Tongue Lights stimuli, with the addition of 4 point-lights on the chin, 8 point-lights outlining the jaw line, 4 point-lights on the cheeks, 2 point-lights on the forehead, one point-light on the tip of the nose, and one point-light on the bridge of the nose (39 points total).

An additional set of stimuli, Audio Alone, showed only a black screen with no visual image.

Auditory stimuli

The actor's auditory speech was recorded onto videotape using a Shure SM57 microphone placed at a distance of 1.5 feet from the mouth. The 240 sentences (60 sentences x 4 visual stimulus types) were then sampled into an AMC 486/33 computer and attenuated so that they all had the same average energy value (approximately 72 dB SPL, A weighted). The sentences were then each replicated 10 times with each replication attenuated by 3 dB. Broadband White noise (constant at 72 dB SPL, A weighted, with a 20 KHz lowpass cut-off) was added to the sentences in the audiovisual dubbing process. This resulted in 10 SNRs (-27, -24, -21, -18, -15, -12, -9, -6, -3, and 0) (MacLeod & Summerfield, 1987). The computer, together with a video recorder and

video player, were used to dub the audiovideo signals. To dub each token, the original tape was played so that its video signal was output to the video recorder and its audio signal was output to a sound activated circuit that was interfaced with the computer. Upon sensing the audio signal, the sound activated circuit signaled the computer to output an attenuated auditory file to the video recorder. Thus, the video token of the original tape and the audio token from the computer file were recorded simultaneously onto a second tape resulting in a new synchronous audiovisual sentence. At the onset of the visual signal (20 frames before the beginning of the auditory stimulus), the noise was added from an Onkyo RX-FS400 dual cassette player. The introduction of the noise before the speech signal avoided a startle effect for participants. The noise lasted about 20 frames after the end of the speech signal. To generate the Audio-Alone presentations, the utterances produced for the Fully-Illuminated presentations were dubbed onto a black image. This image (along with the white noise) began 20 frames before and ended 20 frames after each sentence.

Design

Participant groups

The 25 participants were randomly assigned to five participant groups. All participants were presented with the same 60 sentences in the same order, but each participant group differed in the Presentation Condition assigned to each sentence. Thus, each participant group saw a different set of sentences for each of the presentation conditions (i.e., no one participant saw all 60 sentences under all presentation conditions). This manipulation was included to test whether the sentence lists differentially affected SRTNs across presentation conditions. The sentences were presented with alternating presentation conditions in the general order Audio-alone, Fully-illuminated, Lip-lights, Lips, Teeth, and Tongue lights, and All Lights. However, the condition which began this alternating order was different for each participant group.

Procedure

Participants were tested individually over three one-hour sessions, all occurring within one week, with no two sessions taking place in less than 24-hours. On the first day, participants were screened for normal hearing and good vision before beginning the experiment. On any given day, participants were allowed to complete as many trials as they wanted and the session was terminated based on their fatigue level. In general, participants were presented between 15 and 25 sentences a day and all participants easily finished the 60 trials across the three sessions.

Participants were seated at a table in a 11 foot x 9 foot room facing a 20-inch Panasonic Color Video Monitor (CT-2010Y) at a distance of 5 feet. The monitor was adjusted to maximize the contrast for the point-light images, and the color was turned off. The participant and experimenter wore Sony MDV600 headphones and the auditory stimuli were presented to the right ear only (MacLeod & Summerfield, 1987). The experimenter sat to the side, facing the participant. The participant was instructed to carefully attend to the video monitor.

Stimuli were presented to the participant using an ascending method of limits (MacLeod & Summerfield, 1987). After an initial video-alone presentation, each audiovisual sentence was presented at the -27 SNR. Sentences in the Audio Alone presentation condition began, of course, with the first audio presentation. After each sentence presentation, the tape was paused and the participant was asked to report any words recognized. Participants were told which words in their response were correct (MacLeod & Summerfield, 1987). The sentence was then repeated with increasing SNRs at 3 dB intervals until the participant was able to recognize all three key words. After the participant recognized the three key words, the next sentence was presented using the same ascending SNR method.

Relative contribution of sentence repetition and SNR manipulation

The repetition-with-feedback-presentation methodology was chosen because it avoids many of the shortcomings of prior research (e.g., ceiling and floor effects). It is also not unlike a natural setting for a listener in a noisy environment: a speaker who is not understood would usually be asked to repeat what was said. However, this presentation technique does present a potential confound. Specifically, it is difficult to determine the extent to which improvement is due to the SNR change versus the repetition-with-feedback-presentation technique itself. With regard to our

primary question, this issue is a minor concern because all of the compared presentation conditions were presented with this methodology. However, this issue is still worth exploring because the repetition methodology is somewhat novel.

In order to examine the relative contribution of the repetition methodology, five additional participants were run using only repetition-with-feedback-presentations. In other words, no SNR manipulation was used and SNR was maintained at the lowest value (-27dB). By calculating the number of repetitions needed for key-word identification for the sentences and then comparing these scores to those derived from participants in the main experiment, the contribution of the SNR manipulation can be assessed. In addition, this control will allow us to determine whether the observed differences between stimulus presentation conditions holds under conditions with and without SNR change.

For this control condition, the five participants were tested with the same stimulus list used for Group 2 in the main experiment. These five participants were recruited from the same pool, and had the same native language, hearing and vision characteristics as the original participants (based on the same assessment). The general procedure was the same including sentence ordering and blocking. The participants were also tested over a three-day period and received a combination of course credit and payment for their participation. The only procedural difference was in the way the sentences were presented. For the control participants, the SNR was set at a constant -27 across all repetitions of each sentence (except, of course, the initial visual-alone presentation). After a sentence was presented, participants were provided feedback in the same manner as the other 25 participants (they were told which of the three key-words were guessed correctly). Participants were again given as many as 11 repetitions of each sentence (including the visual-alone) in order to guess all three key-words.

Learning effects

In order to look at whether participants improved their performance during the experiment, SRTN scores will be compared between the first and second halves of the experiment. Analyses will focus on whether performance changes differently for the various presentation conditions.

Results

Main Experiment

Scoring

A Speech Reception Threshold in Noise (SRTN) score was obtained for each sentence for each participant based on the SNR value at which all three key words of the sentence were recognized. Due to a recording error, one sentence was left out of one of the presentation tapes so that the total number of judged sentences was 1495 (5 participant groups x 5 participants per group x 60 sentences - 5 missing sentences). Sentences that were fully recognized with the initial video-alone presentation were given a score of -30, one step below the starting SNR for the ascending series (MacLeod & Summerfield, 1987). A total of 50 of the 1495 responses were scored with a value of -30 SNR. There were 40 such responses for the Full Illumination condition, 1 for the Lips Lights condition, 4 for the Lips, Teeth and Tongue condition and 5 for the All Lights presentation condition. MacLeod and Summerfield (1987) also used this scoring method as an approximation and warned that it has the potential to overestimate the signal level for the SRTN. Thus, the resulting average range of visual improvement may be an underestimate of the actual increase in the intelligibility due to the addition of visual information.

Sentences that were never fully recognized were given a score of +3, one step above the last SNR for the ascending series. This applied to 59 of the 1495 responses scored. There were 25 such responses for the Auditory Alone condition, 5 for the Full Illumination condition, 5 for the Lips Lights condition, 11 for the Lips, Teeth, and Tongue Lights condition, and 13 of these scores for the All Lights presentation condition. It is acknowledged that this scoring could produce underestimations of the signal level for the SRTN, thus overestimating the increase in intelligibility with the introduction of visual information. However, it was most often the case that when participants were unable to recognize all three key words, they had already recognized two. Thus,

participants may have been very close to recognizing all three key words at the 0 SNR level so that the +3 SRTN scores may approximate these thresholds.

Analyses

Mean SRTNs (and standard deviations) for Presentation Conditions were -4.7 (4.0), -15.5 (9.1), -8.4 (5.5), -9.6 (6.7), and -10.0 (7.2) for Auditory Alone, Full Illumination, Lips Lights, Lips, Teeth, and Tongue Lights, and All Lights conditions, respectively. Thus, the average visual benefit over auditory speech (measured as the difference between the Audiovisual and Audio-alone SRTNs) for the Fully-Illuminated condition was 10.8 dB whereas the improvement for the point-light conditions ranged between 3.7 and 5.3 dB. (Informal surveying of the raw data suggests that the relative performance between presentation groups was reflected in individual participant performance.) Mean SRTNs (and standard deviations) for Participant Groups 1 through 5 were -9.8 (7.9), -10.6 (7.3), -9.6 (7.8), -8.4 (7.1), and -9.6 (7.4), respectively.

Each sentences' SRTN score was used as a data point in an omnibus ANOVA of Presentation Condition (5) X Participant Groups (5). The Greenhouse-Geisser epsilon correction was used to adjust for high degrees of freedom. Presentation Condition was found to be significant, $F(4, 1160) = 105.79$, $p<.0001$, as was Participant Group, $F(4, 1160) = 3.41$, $p<.01$. However, no interaction was found $F(16, 1160) = .94$, $p = .52$, suggesting that the difference between Participant Groups was not influenced by Presentation Condition. This indicates that the five sentence lists were not differentially effective across the presentation conditions.

Simple means comparisons were tested between each of the Presentation Conditions, with a Bonferroni Test Correction ($p<.02$) due to the high number of comparisons made. All of the 10 possible comparisons were found to be significant, except for Lips, Teeth, and Tongue Lights vs. All Lights conditions. All visual conditions were significantly different from the Auditory-Alone presentation condition.

The SRTN score ranges of the presentation conditions are listed in Table 1. These scores were computed by calculating participant group means for each sentence and then each presentation condition. The ranges were then derived by pooling these scores for each presentation condition and then finding the difference between these scores for the audiovisual conditions and audio-alone condition. The overall range of 0 to 15.8 dB is consistent with the improvement ranges found in previous research.

Table 1

Ranges of Improvement for Speech in Noise Identifications With Four Audiovisual Presentation Conditions

Presentation Condition	Mean SRTN	Range of Improvement of SRTN
Lips Lights	-8.4	0 - 7.3 dB
Lips, Teeth, and Tongue Lights	-9.6	.3 - 12.8 dB
All Lights	-10.0	1.3 - 10.5 dB
Full Illumination	-15.5	5.9 - 15.8 dB

For the participant group means, a Fisher's Protected LSD means comparison found three significant comparisons for participant groups at the $p<.05$ level. This test revealed that participants in Group 4 had higher SRTNs overall than participants in Groups 1, 2, and 5. Because participants were randomly assigned to the five participant groups, it is unclear why participants in Group 4 performed less well. However, surveying the audio-alone scores for the five participant groups revealed that participants in Group 4 did have a higher mean SRTN score than the others (Group 1 = -4.75 (3.7); 2 = -5.5 (3.8); 3 = -4.9 (4.2); 4 = -3.5 (4.0); 5 = -4.55 (4.2)). This could indicate that for some reason, participants in Group 4 were generally worse at extracting speech from noise.

Relative contribution of sentence repetition and SNR manipulation

In order to evaluate the performance of the control group, a repetition score was obtained for each sentence for each participant based on the sentence repetition number at which all 3 key words were recognized. If the three keywords were never recognized, then a score of 11 was recorded, and if they were recognized during the video-alone trial, the recorded score was 0. The mean repetition scores (and standard deviations) for the Presentation Conditions were 11 (0), 6.45 (4.88), 10.48 (2.03), 9.57 (3.30), and 9.65 (3.0) for Auditory Alone, Full Illumination, Lips Lights, Lips, Teeth, and Tongue Lights, and All Lights conditions, respectively.

In order to determine whether the SNR increase added substantially to performance, a comparison was conducted between the control group data and the data of the original Group 2 participants. For this comparison, raw scores from the Group 2 participants were converted to repetition scores. Thus, the -30, -27,. . .+3 scores were converted to 0, 1, . . . 11 scores respectively. The converted Group 2 mean repetition scores (and standard deviations) were 8.03 (1.51), 4.67 (2.97), 6.93 (2.57), 6.6 (2.55), and 6.28 (1.74) for Auditory Alone, Full Illumination, Lips Lights, Lips, Teeth, and Tongue Lights, and All Lights conditions, respectively.

An omnibus ANOVA was conducted including the variables of Participant Group (original Group 2 Participants and Control Participants) x Presentation Condition. Again, the Greenhouse-Geisser epsilon correction was used to adjust for high degrees of freedom. This ANOVA revealed significant effects for Participant Group, $F(1,472)=145.7$, $p<.0001$, and for Presentation Condition, $F(4,472)=36.65$, $p<.0001$, but no significant interaction between these two factors, $F(4,472)=1.98$, $p>.05$.

Although assessing the relative contribution of the SNR and repetition/feedback manipulations for the original data is difficult, a rough estimate can be made. An index of relative improvement was derived by calculating the absolute difference between a baseline score of '11' and the means for the Control and Group 2 values. The mean improvement score for the Control group was 1.57 and the score for Group 2 was 4.72.

Learning effects

In order to determine whether there were learning effects during the experiment, SRTN scores between the first and second halves of the main experiment were compared. The means for each presentation condition for each half of the main experiment are depicted in the top panel of Figure 1. To test the statistical significance of these effects, an ANOVA was conducted including the factors of Experiment Half (2) and Presentation Condition (5). Again, the Greenhouse-Geisser epsilon correction was used to adjust for high degrees of freedom. This test revealed a significant interaction of Experiment Half and Presentation Condition, $F(4, 96)=4.59$, $p=.0019$, as well as main effects of Experiment Half, $F(1,96)=34.57$, $p<.0001$, and Presentation Condition, $F(4,96)=131.71$, $p<.0001$. Post-hoc comparisons were conducted to test whether there was an effect of Experiment Half for each presentation condition. Results of these analyses indicate that while there was a significant experience effect for all audiovisual presentations at the $p<.01$ level, there was no significant effect for the audio-alone condition ($F(1,96)=.181$, $p=.67$).

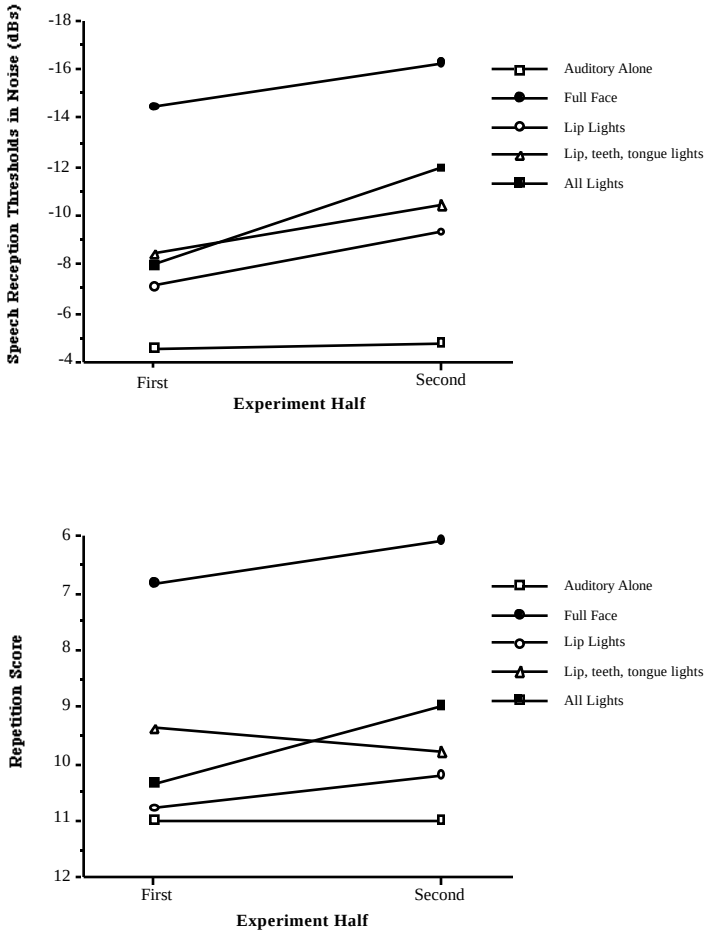

Figure 1. Change in performance across experiment halves for five presentation conditions for the main experiment (top panel) and control condition (bottom panel) (see text for details).

An analogous learning effect analysis was conducted on the data of the five participants from the repetition control experiment. The means for each presentation condition for each half of this control experiment are depicted in the bottom panel of Figure 1. An ANOVA was conducted on these data again including the factors of Experiment Half (2) and Presentation Condition (5). This test revealed significant main effects of Experiment Half, $F(1,16)=2.65$, $p=.036$, and Presentation Condition, $F(4,16)=125.22$, $p<.0001$, but no interaction between these factors, $F(4,16)=4.7$, $p=.177$.

Discussion

Our main question concerned the relative effectiveness of the point-light conditions for enhancing recovery of speech in noise. Results show that all of the point-light conditions induced a significant increase in performance over the Auditory Alone condition. This suggests that the kinematic information provided by the point-lights was able to improve recognition of speech in noise. This result is even more compelling when it is noted that a conservative measure was used; the unrecognized sentences were scored with an SRTN value of +3, with the Auditory Alone condition receiving the most +3 scores. Thus, the auditory alone scores might be disproportionately underestimated such that the visual enhancement was even greater than is suggested by these scores.

The current results invite some speculation about what aspects of the moving face help enhance speech in noise. It is noted that our design precludes precise determination of the salient visual features. For this to be accomplished, specific facial features (lips; tongue) would need to be directly pitted against one-another. Furthermore, a thorough test of features would involve systematic, phoneme-by-phoneme manipulation of the material tested. In that our experiment involved only three point-light conditions and sentence length material, only gross evaluations of the contributions of visual features to comprehension of speech in noise can be accomplished. With this caveat in mind, some general conclusions can be drawn.

Visual features for enhancing speech in noise

The current results showed evidence that lip-lights alone could enhance recovery of speech in noise. This finding contrasts with those of Summerfield (1979) who found that his lip-point configuration did not provide significant enhancement. Summerfield mentioned that the specific placement of points prompted participants to comment that the lips never really seemed to close. Without lip closure, many consonantal segments (e.g., bilablials) would not be specified. Summerfield also suggested that certain vowel distinctions might have been disguised with his point configuration. He cites the work of Jackson, Montgomery, and Binnie (1976) that indicates that to lipread vowels, the dimensions of horizontal lip extension, vertical lip separation, and overall oral area must be specified. Summerfield speculated that while his point configuration might have specified the first of these dimensions, it likely did not specify the latter two.

In contrast to Summerfield's (1979) results, we did observe significant enhancement with our lip-light display. A number of factors distinguish our experiment from Summerfield's, including the type of noise used, the participant task, and the analysis criterion. However, the most profound difference lies with the stimuli themselves. In contrast to Summerfield's displays, our Lip-Lights condition involved points on the inner and outer lip margins, as well as the lip corners. To us, the lips clearly looked like they were closed during bilabials, and numerous vowel distinctions were clear. Thus, the relative success of our displays can most easily be attributed to the increased number and more informative placement of points on the lips. In commenting on Summerfield's failure to find significant enhancement with lip-lights, Campbell (1989) has stated that ". . .the major gain in lipreading does not come from a simple computation of the movements associated with the lips. . ." (p.195). However, this conclusion seems premature in light of the current findings; the significant enhancement provided by our Lip-Light condition suggests that there can be substantial information in lip kinematics.

Next, the current results reveal that increasing the number of points improved enhancement of speech in noise. This is apparent in the significant improvement with both the Lips, Teeth, and Tongue Light condition and All Light condition over the Lip Lights condition. Increasing the number of point-lights coincides with an increase in the amount of available kinematic information and articulators specified. Clearly, adding information about the moving tongue adds to articulatory specification. Adding lights to additional articulators could augment information in more general ways as well. For example, additional points on relatively slow-moving articulators might help provide better references against which movement of the more animated points are seen. This might be particularly true of each set of teeth lights. Although the upper and lower sets move vertically, the two points within each set do not move relative to one-another. This arrangement might provide a stable backdrop against which the other more mobile lip and tongue lights are seen. Next, adding point-lights to additional articulators likely provides more depth information about the location and movement of the articulators. Although some depth information might be available in the shape deformation of the points themselves, it is likely that the occlusion and disocclusion of, say, the teeth and tongue lights by the lips, provide additional information. This additional depth information could help to more precisely specify the subtleties of lip and tongue movements.

The significant improvement with the addition of teeth and tongue lights is compatible with past research. For example, by comparing the enhancement provided by the luminous lips (in which the lips were painted with luminous makeup) and full-face conditions, Summerfield (1979) speculated on the relative contribution of extra-lip features such as the teeth, tongue, and jaw. He found that the isolated lips were not as enhancing as the fully-illuminated face. In explaining these

results, Summerfield cites the comments of his participants that the absence of the teeth and tongue—more than the missing jaw and 'facial frame'—was responsible for their poorer performance. This intuition is supported by recent observations of McGrath and his colleagues (McGrath, 1985; McGrath, Summerfield, & Brooke, 1984; Summerfield, MacLeod, McGrath, & Brooke, 1989) who found that inclusion of the teeth in displays improved lip-reading performance. The current speech-in-noise results further hint at the importance of teeth and tongue visibility. Adding these points to the display significantly improved enhancement over the Lip-Lights alone condition.

Although it was generally the case that more points allowed for improvement in SRTNs, there was no significant difference between the Lips, Teeth, and Tongue Lights and the All Lights conditions. Apparently, adding points to the brow, cheeks, jaw, and nose-tip did not add substantially to performance when seeing points on lips, teeth, and tongue. These results also bear on past research. In fact, there is some evidence that the full facial frame itself can afford viseme recovery. Greenberg and Bode (1968) found that seeing the entire face improved consonant identification over seeing the mouth (lips, teeth, tongue) alone (see also Larr, 1959; and Stone, 1957). However, recent research by IJsseldijk (1992) found no significant improvement with full-facial displays for lipreading words, phrases, and sentences. The current speech in noise results also fail to show a benefit of seeing extra-mouth dimensions: no significant improvement was found over the mouth points display when adding points to the chin, cheeks, nose, and brow. As mentioned, our methodology allows only gross speculation about the relative importance of features. Future research should determine whether the observed lack of enhancement was due to our particular placement of points, or if extra-mouth dimensions in general, provide little benefit for speech in noise recovery.

Overall, the current results reveal that point-light images can significantly enhance speech recognition in noise. It is noted however, that no point-light presentation condition provided as much enhancement as the Full Illumination condition. These results are reminiscent of other point-light speech findings. In the audiovisually discrepant speech experiment of Rosenblum and Saldaña (1996), most conditions revealed a significantly greater visual influence with the fully-illuminated face than the point-light stimuli. The superior enhancement of the fully-illuminated condition was also evident in our point-light lipreading experiments (Rosenblum et al., 1995).

There are a number of reasons why point-light speech stimuli might not be as effective as fully-illuminated faces (see Rosenblum & Saldaña, 1996, for a lengthy discussion of this issue). First, it could be that the pictorial features provided in fully-illuminated displays are indispensable. Perhaps seeing the contrasting textures of the surfaces of the teeth, tongue, lips, and facial skin is critical for full speech-in-noise enhancement. Alternatively, pictorial features might not be necessary for optimal enhancement. Perhaps, the abstracted kinematic information provided by point-light stimuli can be equally effective. Potentially, the current results could be a consequence of a point arrangement which is less than optimal for capturing all salient kinematic dimensions (cf. Rosenblum & Saldaña, 1996). It is possible that another arrangement of points would provide enhancement equivalent to the fully-illuminated condition.

Relative contribution of sentence repetition and SNR manipulation

The control condition was implemented to determine the extent to which the repetition with feedback presentation methodology influenced performance over the SNR manipulation. The results revealed a significant difference between the control and original Group 2 participants indicating that the SNR manipulation did account for a significant amount of the original performance. Based on the derived index, the Control group score was about 35% that of the Group 2 score. This provides a rough estimate that 65% of the original performance was based on the SNR manipulation over and above the influence of the repetition with feedback technique. Furthermore, the fact that no interaction was found between Participant Group and Presentation Condition factors indicates that the relative performance of the presentation conditions was maintained regardless of whether SNR was manipulated.

Learning effects

Analyses showed that participants improved their performance during the experiment. While these learning effects were observed for both the main experiment group and control group data,

discussion will concentrate on the main experiment group since this data was derived from five times as many subjects, and included the correct sentence ordering control.

There are a number of possible reasons why these learning effects occurred. First, it could be that the observed improvement was a result of simply learning to extract speech from noise. Secondly, the improvement could be a result of learning the cognitive-linguistic aspects of the task. Participants were presented linguistically similar sentences and were given feedback upon each presentation. It could be then, that participants' strategies in taking cues from the semantic context of the sentences, as well as from the experimenter's feedback, improved across the experiment.

However, if improvement was due to either simple extraction of speech in noise or general cognitive-linguistic factors, then all of the presentation conditions should show similar learning effects. The main experiment results show this not to be the case: although SRTNs improved for each of the audiovisual conditions, no significant improvement was observed for the audio-alone condition. A more tenable explanation is that the improvement was a result of perceptual learning with the various visual conditions. It is likely that some of this improvement was based on increasing familiarity with the particular speaker used in this experiment. This can be inferred from the fact that participants showed improved performance for the fully-illuminated face condition even though they have had life-long experience with similar stimuli. Still, the improvement with the point-light conditions is striking. In the second half of the main experiment, the All-Points condition induced a substantial mean 7.1 dB SRTN improvement over the second half's Audio-Alone condition.

This improved ability lends further support to the notion that there is useful information in the kinematics, and that pictorial features are not necessary for enhancing speech in noise[4]. Perhaps the differential effectiveness of fully-illuminated and point-light displays is simply a consequence of experience with the two types of stimuli. Clearly, observers have infinitely more experience with fully-illuminated speaking faces. With regard to what the nature of the learning might be, experience might help observers perceive the correspondence between auditory and visual information in the point-light stimuli. Alternatively, and as Rosenblum and Saldaña (1996) have suggested, observers might simply need to learn to attend to the appropriate kinematic dimensions in these displays. In fact, learning to detect salient kinematic dimensions might not be peculiar to point-light displays. It could be that learning to attend to these dimensions is critical for general lipreading.

Much research has shown that lip-reading skill can improve with training (see Massaro, Cohen, & Gesi, 1993, for a review). It is also known that the utility of visual speech in enhancing speech in noise can improve with experience, and that this experience can carry-over to other speech tasks (Danz & Binnie, 1983; Montgomery, Walden, Schwartz, & Prosek, 1984). This past research used fully-illuminated visual stimuli, as well as explicit and substantial training. Still, it could be that the improvements we have observed with our point-light stimuli are based on similar perceptual strategies. If so, it is likely that explicit training would further improve the power of the point-light stimuli for speech in noise enhancement. This training could involve describing where the points are located on the face, and presenting a full repertoire of viseme segments together with performance feedback. Experiments are planned to determine just how much improvement can occur in perception of point-light stimuli given explicit and lengthy training.

Implications for telecommunications

The current results also have implications for the development of telecommunication systems for listeners with hearing impairment. Point-light displays, unlike fully-illuminated images, can more easily be transmitted through existing telephone lines. In fact, there have been successful demonstrations of point-light telecommunication of hand signing (Pearson, 1981; Tartter & Knowlton, 1981). Point-light faces could help in the development of a telecommunication system to convey lipread information allowing the deaf to telecommunicate with those who do not know sign-language (Massaro, 1987; Pearson, 1981; Rosenblum & Saldaña, 1996). Erber (1980) has shown that point-light images can be useful for children with severe hearing impairment. The current speech-in-noise results are important in suggesting that point-light images could also improve comprehension of degraded speech signals for those with mild-to-moderate hearing

impairments. The current results further hint that relatively few points will be needed for enhancement; the mouth point configuration was as effective as the configuration with points on the mouth, chin, cheeks, nose and brow.

Finally, these results also provide evidence that simple experience can make these stimuli more effective. It is likely that explicit training would increase the stimuli's effectiveness even further, allowing them to be useful for telecommunication purposes. Clearly, these results are preliminary and significantly more research will be needed to determine the most efficient point placement. Also, testing individuals with hearing impairments with the current stimuli will be critical for these ends. Still, the current results provide good first evidence that facial point-light stimuli could be effective for enhancing noisy speech in telecommunication systems.

References

American National Standards Institution (1989). Specification for audiometers. (ANSI S3.6 - 1989)

Bassili, J.N. (1978). Facial motion in the perception of faces and of emotional expression. Journal of Experimental Psychology: Human Perception and Performance, 4, 373-379.

Bench, J. & Bamford, J. (1979). Speech-Hearing Tests and the Spoken Language of Hearing Impaired Children. London: Academic Press.

Berry, D.S. (1990). What can a moving face tell us? Journal of Personality and Social Psychology, 58, 1004-1014.

Bingham, G.P. (1987). Scaling and kinematic form: Further investigations on the visual perception of lifted weight. Journal of Experimental Psychology: Human Perception and Performance, 13, 155-177.

Bingham, G.P., Rosenblum, L.D., & Schmidt, R.C. (1995). Dynamics and the orientation of kinematic forms in visual event recognition. Journal of Experimental Psychology: Human Perception and Performance. 21, 1473-1493.

Brooke, N.M., & Summerfield, A.Q. (1983). Analysis, synthesis and perception of visible articulatory movements. Journal of Phonetics, 11, 63-76.

Campbell, R. (1989). Lipreading. In A.W. Young & H.D. Ellis (Eds.), Handbook of Research on Face Processing, North-Holland: Elsevier.

Danz, A.D. & Binnie, C.A. (1983). Quantification of the effects of training the auditory-visual recognition of connected speech. Ear and Hearing, 4, 146-151.

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.

Erber, N.P. (1980). Central Institute for the Deaf Research Department, Periodic Progress Report, 23.

Greenberg, H.J. & Bode, D.L. (1968). Visual discrimination of consonants. Journal of Speech and Hearing Research, 11, 869-874.

Hinkle, R.R. (1979). An investigation of list equivalency for auditory, visual and audiovisual performance using revised CID sentences, Unpublished doctoral dissertation, Purdue University.

IJsselddijk, F.J. (1992). Speechreading performance under different conditions of video image, repetition, and speech rate. Journal of Speech and Hearing Research, 35, 466-471.

Jackson, P.L., Montgomery, A.A., & Binnie, C.A. (1976). Perceptual dimensions underlying vowel lipreading performance. Journal of Speech and Hearing Research, 19, 796-812.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. Perception and Psychophysics, 14, 201-211.

Kozlowski, L.T., & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point light display. Perception and Psychophysics, 21, 575-580.

Larr, A.L. (1959). Speechreading through closed-circuit television. Volta Review, 61, 19-21.

MacLeod, A. & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. British Journal of Audiology, 21, 131-141.

MacLeod, A. & Summerfield, Q. (1990). A procedure for measuring auditory and audio-visual speech reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use. British Journal of Audiology, 24, 29-43.

Massaro, D.W. (1987). Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry. Hillsdale, NJ: Earlbaum.

Massaro, D.W., Cohen, M.M., & Gesi, A.T. (1993). Long-term training, transfer and retention in learning to lipread. Perception and Psychophysics, 53, 549-562.

McGrath, M. (1985). An examination of cues for visual and audio-visual speech perception using natural and computer-generated faces. Unpublished doctoral dissertation, University of Nottingham, Nottingham, U.K.

McGrath, M., Summerfield, Q. & Brooke, N.M. (1984). Roles of lips and teeth in lipreading vowels. Proceedings of the Institute of Acoustics, 6, 401-408.

Middleweerd, M. J., & Plomp, R. (1987). The effect of speechreading on the speech reception threshold of sentences in noise. Journal of the Acoustical Society of America, 82, 2145-2146.

Montgomery, A.A. & Jackson, P.L. (1983). Physical characteristics of the lips underlying vowel lipreading performance. Journal of the Acoustical Society of America, 73, 2134-2144.

Montgomery, A.A., Walden, B.E., Schwartz, D.M., & Prosek, R.A. (1984). Training auditory-visual speech reception in adults with moderate sensorineural hearing loss. Ear and Hearing, 5, 30-36.

Nilsson, M.J., Soli, S.D., & Sullivan, J. (1994). Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. Journal of the Acoustical Society of America, 95, 1085-1099.

Pearson, D. (1981). Visual communication systems for the deaf. IEEE Transactions on Communications, COM-29, 1986-1992.

Petajan, E.D. (1984). Automatic lipreading to enhance speech recognition. Proceedings of the Global Communications Conference, Atlanta, Georgia, USA, IEEE Communications Society, 265-272.

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds), Hearing by eye: The psychology of lip reading. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Rosenblum, L. D., Johnson, J. A., & Saldaña. H. M. (1995). Determining the kinematic features for visual speech perception. Colloquium presented at the University of California, Riverside, May, 10.

Rosenblum, L. D. & Saldaña. H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception. Journal of Experimental Psychology: Human Perception and Performance, 22(2), 318-331.

Runeson, S. (1977). On the visual perception of dynamic events. Acta Universitatis Upsaliensis: Studia Psychologica Upsaliensia. (Series No. 9).

Runeson, S. & Frykholm, G. (1981). Visual perception of lifted weight. Journal of Experimental Psychology: Human Perception and Performance, 7, 733-740.

Stone, L. (1957). Facial clues of context in lip reading. John Tracy Clinic, Los Angeles Research Papers, 5. Los Angeles: John Tracy Clinic.

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., MacLeod, P., McGrath, M., & Brooke, N.M. (1989). Lips, teeth, and the benefits of lipreading. In Handbook of Research on Face Processing, A.W. Young and H.D. Ellis (eds.) New Holland: Elsevier, pp. 223-233.

Summerfield, Q. & McGrath, M. (1984). Detection and resolution of audio-visual incompatibility in the perception of vowels. Quarterly Journal of Experimental Psychology, 36, 51-74.

Summerfield, Q.A. (1979). Use of visual information for phonetic perception. Phonetica, 36, 314-331.

Tartter, V.C. & Knowlton, K.C. (1981). Perception of sign language from an array of 27 moving spots. Nature, 289, 676-678.

Verfaillie, K., De Troy, A., & Van Rensbergen, J. (1994). Transsaccadic integration of biological motion. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20, 649-670.

Appendix

Five Sentence Lists Equalized for Ease of Lip Reading

List	Sentence	Lipreading Score (see text)
1	The **football game** is **over**	46
	The **boy forgot** his **book**	30
	The **family bought** a **house**	29
	The **letter fell** on the **floor**	21
	The **cat drank** from the **saucer**	21
	They **watched** the **scary movie**	14
	The **police help** the **driver**	14
	The **lady sits** in her **chair**	8
	The tall **man tied** his **shoes**	7
	The **grocer sells butter**	4
	They went on **vacation**	3
	She **lost** her **credit card**	0
		total =197
2	A **boy fell** from the **window**	43
	The **boy got** into **trouble**	30
	The **cows** are **in** the **pasture**	29
	Someone's crossing the **road**	22
	The **wife helped** her **husband**	20
	He really **scared** his **sister**	16
	The **cat lay** on the **bed**	14
	The **shirts** were **in** the **closet**	9
	The **broom stood** in the **corner**	7
	The **lady wore** a **coat**	4
	The **sun melted** the **snow**	3
	There was a **bad train wreck**	0
		total =197
3	The **cat jumped** over the **fence**	42
	The **cups** are **on** the **table**	32
	Somebody stole the **money**	27
	The **mailman shut** the **gate**	22
	The **cherry pie** was **good**	20
	The **ball bounced** very **high**	16
	The **dog played** with a **stick**	13
	She argued with her **sister**	10
	He **broke** his **leg again**	7
	The **rancher has** a **bull**	4
	The **scissors** are **quite sharp**	3
	They wanted some **potatoes**	0
		total =196

List	Sentence	Lipreading Score (see text)
4	The **car** is **going fast**	39
	The **tree fell** on the **house**	32
	He **wore** his **yellow shirt**	26
	Flowers grow in the **garden**	24
	He **found** his **brother hiding**	19
	The **children** are **walking home**	16
	The **chicken laid** some **eggs**	12
	She looked in her **mirror**	10
	The **three girls** are **listening**	6
	Potatoes grow in the **ground**	5
	The **new road's** on the **map**	3
	The **young people** were **dancing**	2
		total =194
5	The **fire** was **very hot**	36
	The **boy's running away**	33
	The **oven door** was **open**	25
	Mother picked some **flowers**	25
	The **boy ran** down the **path**	18
	He played with his toy **train**	17
	The **two farmers** are **talking**	11
	She stood near her **window**	10
	The **match boxes** are **empty**	6
	The **children washed** the **plates**	6
	The **baby broke** his **cup**	3
	Her **shoes** were **very dirty**	3
		total =193

Author Notes

We gratefully acknowledge the assistance of Chantelle Bosely, Julie Garcia, Sunny Moore, Kristina R. Schillberg, Rebecca Vasquez.

This research was supported by NSF Grant DBS-9212225 awarded to Lawrence D. Rosenblum.

Requests for reprints should be sent to Lawrence D. Rosenblum, Department of Psychology, University of California, Riverside, Riverside, California, 92521, rosenblu@citrus.ucr.edu.

Footnotes

[1]Throughout this paper, the terms 'enhance' and 'enhancement' are used to denote how visual information embellishes a degraded auditory signal. It is acknowledged that it is also possible for auditory information to enhance a degraded visual image. The distinction between these scenarios guides research designs and has important implications for understanding types and magnitude of hearing impairments. (We thank an anonymous reviewer for pointing-out these issues.) In the present experiment, both degraded visual and auditory stimuli are tested. We have chosen to follow the lead of others who have used similar methodologies (MacLeod & Summerfield, 1987; 1990; Summerfield, 1979) and conceptualize the experiment as a test of how visual information enhances degraded auditory speech.

[2]In a more recent paper, MacLeod and Summerfield (1990) improved on their methodology by using an up-down adaptive sentence presentation procedure. This allowed SRTNs to be estimated by on-line manipulation of SNRs based on listener performance for each sentence. Our decision to

replicate the methodology of the 1987—rather than 1990—paper was based on equipment constraints.

[3]It is acknowledged that there is some repetition among the keywords across the sentences (e.g., 'boy'). Word repetition often occurs in studies that use the BKB lists. This fact does not present a large problem in the current study because all sentence lists were tested under all presentation conditions.

[4]It could be argued that our stimuli contain more than kinematic dimensions: i.e., there might be 'pictorial' information in the array of points. However when shown statically, our stimuli are not recognizable as faces and the points are not seen as facial features (see also Rosenblum & Saldaña, 1996). Thus, the points are not pictorial features in any usual sense of the term. It is in this way that we maintain our point-light images demonstrate the salience of kinematic dimensions. Similar arguments have been made regarding point-light images outside the domain of speech (e.g., Bingham, Rosenblum, & Schmidt, 1995; Runeson, 1977).