# LETTERS TO THE EDITOR

## Some normative data on lip-reading skills (L)

Nicholas A. Altieri[a]
*Department of Psychology, The University of Oklahoma, 3100 Monitor Avenue, 2 Partners Place, Suite 280, Norman, Oklahoma 73072*

David B. Pisoni and James T. Townsend
*Department of Psychological and Brain Sciences, Indiana University, 1101 E. 10th Street, Bloomington, Indiana 47405*

(Received 16 July 2010; revised 19 April 2011; accepted 4 May 2011)

The ability to obtain reliable phonetic information from a talker's face during speech perception is an important skill. However, lip-reading abilities vary considerably across individuals. There is currently a lack of normative data on lip-reading abilities in young normal-hearing listeners. This letter describes results obtained from a visual-only sentence recognition experiment using CUNY sentences and provides the mean number of words correct and the standard deviation for different sentence lengths. Additionally, the method for calculating T-scores is provided to facilitate the conversion between raw and standardized scores. This metric can be utilized by clinicians and researchers in lip-reading studies. This statistic provides a useful benchmark for determining whether an individual's lip-reading score falls within the normal range, or whether it is above or below this range. © *2011 Acoustical Society of America*. [DOI: 10.1121/1.3593376]

PACS number(s): 43.71.Sy, 43.71.Rt, 43.71.Gv, 43.71.Lz [MSS] Pages: 1–4

### I. INTRODUCTION

Evidence from studies in audiovisual speech perception has shown that visual speech cues, provided by optical information in the talker's face, has facilitatory effects in terms of accuracy across a wide range of auditory signal-to-noise ratios (Grant and Seitz, 1998; Sumby and Pollack, 1954). In their seminal study, Sumby and Pollack reported that the obtained benefit from the visual speech signal is related to the quality of the auditory information, with a more noticeable gain observed for lower signal-to-noise ratios. Their findings are theoretically important as they also show that visual information provides benefits across many signal-to-noise ratios.

In a more recent study involving neural measures of visual enhancement, van Wassenhove *et al.* (2005) compared peek amplitudes in an EEG task and observed that lip-reading information "speeds up" the neural processing of auditory speech signals. The effects of visual enhancement of speech continue to be explored using more recent methodologies and tools (for an analysis using fMRI, see also Bernstein *et al.*, 2002). Although numerous studies have demonstrated that visual information about speech enhances and facilitates auditory recognition of speech in both normal-hearing and clinical populations (Bergeson and Pisoni, 2004; Kaiser *et al.*, 2003), there is currently a lack of basic information regarding more fundamental aspects of visual speech processing. Most surprising perhaps, is that even after decades of research there are no normative data on lip-reading ability available to researchers and clinicians to serve as benchmarks of performance.

When a researcher obtains a cursory assessment of lip-reading ability, how does the score compare to the rest of the population? Simply put, what exactly constitutes a "good" or otherwise above-average lip-reader? Although there is a growing body of literature investigating perceptual and cognitive factors associated with visual-only performance (e.g., Bernstein *et al.*, 1998; Feld and Sommers, 2009) exactly what constitutes superior, average, and markedly below-average lip-reading ability has yet to be quantified in any precise manner. Auer and Bernstein (2007) did report lip-reading data from sentence recognition tasks using normal-hearing and hearing-impaired populations, and provided some initial descriptive statistics from both populations. In this letter, we go a step further by reporting standardized T-scores, and, additionally, recognition scores for sentences of different word lengths.

#### A. Visual-only sentence recognition

To answer the question of what accuracy level makes a good lip reader, we carried out a visual-only sentence

[a] Author to whom correspondence should be addressed. Electronic mail: nick.altieri@ou.edu

recognition task designed to assess lip-reading skills in an ecologically valid manner. Eighty-four young normal-hearing undergraduates were presented with 25 CUNY sentences (with the auditory track removed) of variable length spoken by a female talker (Boothroyd *et al.*, 1988). The use of CUNY sentence materials provides a more ecologically valid measure of language processing than the perception of words or syllables in isolation. One potential objection to using sentences is their predictability. However, language processing requires both sensory processing in addition to the integration of contextual information over time. Therefore, the use of less predictable or anomalous sentences might be of interest in future studies, although it remains beyond the scope of our present report. We shall now describe the details of the study, and provide the results and method for converting raw scores to T-scores.

## II. EXPERIMENT

### A. Participants

Eighty-four college-age participants were recruited at Indiana University and were either given course credit or paid for their participation. All participants reported normal hearing and had normal or corrected vision at the time of testing.

### B. Stimulus materials

The stimulus set consisted of 25 sentences obtained from a database of pre-recorded audiovisual sentences (CUNY sentences) (Boothroyd *et al.*, 1988) spoken by a female talker. The auditory track was removed from each of the 25 sentences using Final Cut Pro HD. The set of 25 sentences was then subdivided into the following word lengths: 3, 5, 7, 9, and 11 words with five sentences for each length. We did this because sentence length naturally varies in everyday conversation. Sentences were presented randomly for each participant and we did not provide any cues with regard to sentence length or semantic content. The sentence materials are shown in the Appendix.

### C. Design and procedure

Data from the 25 visual-only sentences were obtained from a pre-screening session in two experiments designed to test hypotheses related to visual-only sentence recognition abilities. The stimuli were digitized from a laser video disk and rendered into a 720 × 480 pixel movie at a rate of 30 frames/s. The movies were displayed on a Macintosh monitor with a refresh rate of 75 Hz. Participants were seated approximately 16–24 in. from the computer monitor. Each trial began with the presentation of a fixation cross (+) for approximately 500 ms followed by a video of a female talker, with the sound removed, speaking one of the 25 sentences listed in the Appendix. After the talker finished speaking the sentence, a dialog box appeared in the center of the screen instructing the participant to type in the words they thought the talker said by using a keyboard. Each sentence was given to the participant only once. No feedback was provided on any of the test trials.

Scoring was carried out in the following manner: If the participant correctly typed a word in the sentence, then that word was scored as "correct." The proportion of words correct was scored across sentences. For the sentence "Is your sister in school," if the participant typed in "Is the…" only the word "Is" would be scored as correct. In this example, one out of five words would be correct, making the proportion correct = 1/5 = 0.20. Word order was not a critical criterion for a word to be scored as accurate. However, upon inspection of the data, participants almost never switched word order in their responses. Subject responses were manually corrected for any misspellings. These visual-only word-recognition scores provide a valuable benchmark for assessing overall lip-reading ability in individual participants and can be used as normative data for other research purposes.

## III. RESULTS

The results revealed that the mean lip-reading score in visual-only sentence recognition was 12.4% correct with a standard deviation of 6.67%. Figure 1 shows a box plot of the results where the lines indicate the mean, 75th and 25th percentile, as well as 1.5 times the interquartile range. Two outliers denoted by open circles, each close to 30% correct, are also plotted. The proportion of words identified correctly was not identical across sentence length. The mean and standard deviation of the accuracy scores for each sentence length are provided in Table I. Correct identification across sentence length differed, with increased accuracy for longer sentences (up to nine words) before decreasing again for sentence lengths of 11 words [F(4,83) = 21.46, $p < 0.001$]. This interesting finding is consistent with the hypothesis that language processing involves the use of higher-order cognitive resources to integrate semantic context over time. Hence, shorter sentences might not provide enough contextual information, whereas longer sentences may burden working memory capacity (see Feld and Sommers, 2009). Although

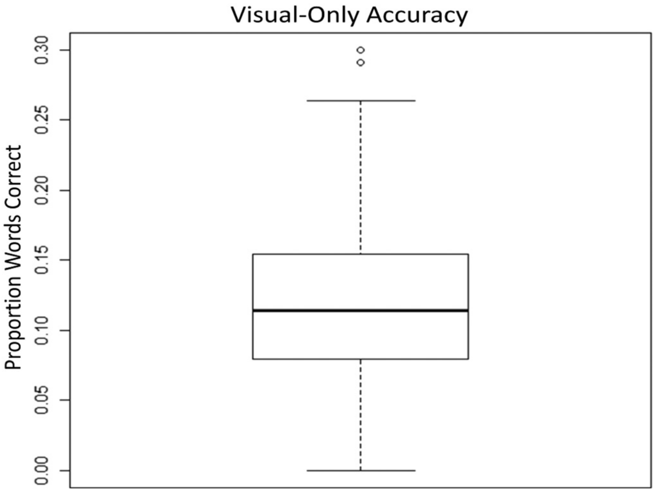

FIG. 1. The line in the middle of the box shows the mean visual-only sentence recognition score across all 84 participants. The 75th and 25th percentile are represented by the line above and below the middle line, respectively. The small bars on the top and bottom denote a value of 1.5 times the interquartile range.

TABLE I. The mean and standard deviation of words correctly identified for each sentence length, including the mean and standard deviation collapsed across all lengths ("Overall").[a]

| Word length | Mean | SD |
|---|---|---|
| Overall | 12.4 | 6.7 |
| 3 | 8 | 8.0 |
| 5 | 8.4 | 6.6 |
| 7 | 15 | 8.9 |
| 9 | 17 | 10 |
| 11 | 13 | 8.3 |

[a] Scores are given in percent correct. The scores indicate that lip-reading accuracy increases as a function of sentence length, but decreases again for sentences longer than nine words.

sentences do provide contextual cues, such information is quite difficult to obtain in the visual modality, especially when sentence length increases. Incidentally, this reasoning explains why auditory-only accuracy, but not visual-only accuracy, improves for sentence recognition compared to single-word recognition in isolation.

### A. Conversion of raw scores to T-scores

In order to determine individual performance relative to a standard benchmark, the method and rationale for calculating T-scores will be provided. Standardized T-scores have a mean of 50 and a standard deviation of 10. These standardized scores are generally preferred by clinicians and psychometricians over Z-scores due to the relative ease of their interpretability and appeal to intuition. For example, T-scores are positive, whereas Z-scores below the mean yield negative numbers, which does not make intuitive sense for visual-only accuracy scores.

The T-scores were computed in the following manner: The overall mean was subtracted from each individual raw score and divided by the standard deviation, thereby converting the raw score into a Z-score. Taking this score, multiplying it by a factor of 10, and then adding 50 provides us with the T-score for that individual:

$$T_i = 10\left[\frac{x_i - \mu}{\sigma}\right] + 50. \quad (1)$$

For example, the mean score of 12.4% correct-word recognition gives us a T-score of 50, whereas an accuracy level of just over 2% yields a T-score of 35 (1.5 standard deviations below the mean) and an accuracy level of just over 22% correct yields a T-score of 65 (1.5 standard deviations above the mean). Computing T-scores is quite convenient, and can be utilized to convert a raw CUNY lip-reading score obtained from an open set sentence recognition test into an interpretable standardized score. This can inform clinicians and researchers where an individual stands relative to the population of young healthy participants.

## IV. CONCLUSIONS

Qualitatively, the scores reflect the difficulty of lip reading in an open-set sentence recognition task. Mean-word-recognition accuracy scores were barely greater than 10% correct. Further, any individual who achieved a CUNY lip-reading score of 30% correct is considered an outlier, giving them a T-score of nearly 80—three times the standard deviation from the mean. A lip-reading recognition accuracy score of 45% correct places an individual 5 standard deviations above the mean.

These results quantify the inherent difficulty in visual-only sentence recognition. One potential concern is that CUNY sentences tend to yield lower *V*-only accuracy than other sentence materials (see, e.g., Auer and Bernstein, 2007). However, the major contribution of our study is that it provides clinicians and researchers with a valuable benchmark for assessing lip-reading skills using a database that has a well-established history in the clinical and behavioral research community (see, e.g., Bergeson and Pisoni, 2004; Boothroyd *et al.*, 1988; Kaiser *et al.*, 2003). CUNY sentences are used widely in sentence perception tasks using normal hearing, elderly, and patients with cochlear implants as subjects. With these results, it is now possible to quantify the lip-reading ability of an individual participant relative to a normal-hearing population.

One potentially fruitful application would be to determine where a specific hearing-impaired listener falls on a standardized distribution. Research on visual-only speech recognition, for example, has shown that individuals with a progressive, rather than sudden, hearing loss have higher lip-reading recognition scores (Bergeson *et al.*, 2003). Other research has also demonstrated that lip-reading ability serves as an important behavioral predictor of who will benefit from a cochlear implant (see Bergeson and Pisoni, 2004). How might the scores from each individual in these populations compare with the standard scores from a normal-hearing population? These examples and numerous other scenarios suggest the importance of having some normative data on lip-reading ability readily available for the speech research community including basic researchers, as well as clinicians, who work with hearing-impaired listeners to determine strengths, weaknesses, and milestones.

Although this study only employed CUNY sentences, the use of sentences from this well-established and widely used database provided a generalized measure of language processing ability, and a T-score conversion method that should be applicable to other open-set sentence identification tasks. Future studies might consider establishing norms for visually presented anomalous sentences, isolated words, and syllables. It will also be worthwhile to obtain normative data for elderly listeners who have been found to have poorer lip-reading skills than younger listeners (Sommers *et al.*, 2005).

## ACKNOWLEDGMENTS

This study was supported by the National Institute of Health (Grant No. DC-00111) and the National Institute of Health Speech Training (Grant No. DC-00012) and by NIMH (057717-07) and AFOSR (FA9550-07-1-0078) grants to J.T.T. We would like to acknowledge the members of the Speech Research Laboratory and James T. Townsend's

Psychological Modeling Laboratory at Indiana University for their helpful and insightful discussion. We also wish to thank two anonymous reviewers for their valuable input on an earlier version of this manuscript.

## APPENDIX

What will we make for dinner when our neighbors come over

Is your sister in school

Does your boss give you a bonus every year

Do not spend so much on new clothes

What is your recipe for cheesecake

Is your nephew having a birthday party next week

What is the humidity

Let the children stay up for Halloween

He plays the bass in a jazz band every Monday night

How long does it take to roast a turkey

Which team won

Take your vitamins every morning after breakfast

People who invest in stocks and bonds now take some risks

Those albums are very old

Aren't dishwashers convenient

Is it snowing or raining right now

The school will be closed for Washington's Birthday and Lincoln's Birthday

Your check arrived by mail

Professional musicians must practice at least three hours everyday

Are whales mammals

Did the basketball game go into overtime

When he went to the dentist he had his teeth cleaned

We'll plant roses this spring

I always mail in my loan payments on time

Sneakers are comfortable

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

Bergeson, T. R., and Pisoni, D.B. (**2004**). "Audiovisual speech perception in deaf adults and children following cochlear implantation," in *The Handbook of Multisensory Processes*, edited by G. A. Calvert, C. Spence, and B. E. Stein (The MIT Press, Cambridge, MA), pp. 153–176.

Bergeson, T. R., Pisoni, D. B., Reese, L., and Kirk, K. I. (**2003**). "Audiovisual speech perception in adult cochlear implant users: Effects of sudden vs. progressive hearing loss," Poster presented at the Annual Midwinter Research Meeting of the Association for Research in Otolaryngology, Daytona Beach, FL.

Bernstein, L. E., Auer, E. T., Moore, J. K., Ponton, C., Don, M., and Singh, M. (**2002**). "Visual speech perception without primary auditory cortex activation," NeuroReport. **13**, 31–315.

Bernstein, L. E., Demorest, M. E., and Tucker, P. E. (**1998**). "What makes a good speechreader? First you have to find one," in *Hearing by Eye II: Advances in the Psychology of Speechreading and Audio-Visual Speech*, edited by R. Campbell, B. Dodd, and D. Burnham (Psychology Press, Erlbaum, UK), pp. 211–227.

Boothroyd, A., Hnath-Chisolm, T., Hanin, L., and Kishon-Rabin, L. (**1988**). "Voice fundamental frequency as an auditory supplement to the speechreading of sentences," Ear Hear. **9**, 306–312.

Feld, J. E., and Sommers, M. S. (**2009**). "Lipreading, processing speed, and working memory in younger and older adults," J. Speech Lang. Hear. Res. **52**, 1555–1565.

Grant, K. W., and Seitz, P. F. (**1998**). "Measures of auditory-visual integration in nonsense syllables and sentences," J. Acoust. Soc. Am. **104**, 2438–2450.

Kaiser, A., Kirk, K., Lachs, L., and Pisoni, D. (**2003**). "Talker and lexical effects on audiovisual word recognition by adults with cochlear implants," J. Speech Lang. Hear. Res. **46**, 390–404.

Sommers, M., Tye-Murray, N., and Spehar, B. (**2005**). "Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults," Ear Hear. **26**, 263–275.

Sumby, W. H., and Pollack, I. (**1954**). "Visual contribution to speech intelligibility in noise," J. Acoust. Soc. Am. **26**, 12–15.

van Wassenhove, V. Grant, K., and Poeppel, D. (**2005**). "Visual speech speeds up the neural processing of auditory speech," Proc. Natl. Acad. Sci. USA **102**, 1181–1186.