Speechreading—Not Just for Deafies

Julie Fitzer [a]

Introduction

People who have hearing loss often talk about speechreading as if they have a special skill. My superior speechreading abilities have earned me the nickname of “undercover CIA agent” on my college campus, and I do not hesitate to brag a little when I have an entire silently speechread conversation with an unsuspecting stranger. Furthermore, our normal hearing friends tend to think that the deaf are the only ones who speechread and speechread well; I often get comments from hearing people along the lines of “I couldn’t possibly speechread all day in a classroom like you do.” Once, in the fifth grade, my hoarse laryngitis-stricken school librarian, assuming that the other students—all hearing—could not speechread her, told me to repeat to the rest of the class what she said as she taught a lesson.

However, as I discovered this past semester while working on a research project on the topic, speechreading is not a privileged skill of the deaf. Even people with normal hearing speechread to some extent, provided that they have the visual capacities to do so and that the relevant visual information is provided simultaneously with the auditory input. The McGurk effect illustrates this point. McGurk and MacDonald (1976) reported that presenting phonemes visually could alter the perception of heard phonemes[1], even if the visual phoneme—or viseme—differs from the heard phoneme. Subjects perceive “da” when presented with a visual “ga” and an auditory “ba” simultaneously; to a lesser degree, subjects perceive “ta” when presented with an auditory “ka” and a visual “pa” simultaneously. The “da” and “ta” represent fused perceptions; when the conflicting stimuli are reversed, such that visual “ba” is presented with auditory “ga” and visual “ka” is presented with auditory “pa,” subjects perceive combinations such as “paka” and “baga”. Both fusions and combinations demonstrate that speechread information can and does influence speech perception, even when subjects are instructed to report only what they hear. The robust McGurk effect does not habituate[2] over time and does not disappear even when the subject being tested is aware of the phenomenon.

When I first learned of the McGurk effect early in my career as a psychology major at Washington and Lee University, I began to ask questions. How do I speechread so well? How does the visual and auditory information join together into understandable speech? Why did I have to turn off my hearing aids when speechreading oral transliteraters who were inevitably a few beats behind the spoken message? Why, above all, would people with normal hearing need to have the ability to speechread? Dr. David Elmes, my wise faculty advisor who encouraged me to use my personal experiences with hearing loss to guide my intellectual curiosity[3], suggested two years ago that I make speechreading the topic of my senior project that is required of all psychology majors. Hence, when it came time to pick a topic and start the project last September, I listened to Dr. Elmes and drove the interlibrary loan staff person crazy with numerous requests for books and articles that my college, which does not have a Communication Sciences and Disorders department, does not carry in its library. In the end, I produced a 45 page paper which reviews the literature on how auditory and visual information is integrated when people with normal hearing use speechreading along with auditory information in speech reception processes.

Rather than bore the non-specialist reader of this journal with arcane descriptions of the cognitive science of speech perception—it takes a strange breed of person, of which I am one, to get excited about minute details of information processing in the brain—I extract from my research some basic information that would be of interest to those who have hearing loss. The present article explores the question of why Nature[4], in all of Her wisdom, would have wired human beings to speechread, and thus makes human beings vulnerable to the McGurk illusion. I first present additional evidence for the natural development of speechreading skill even in people with normal hearing, particularly children. Second, I explain the difficulties of auditory-only speech perception that even those with normal hearing encounter and how speechreading may help people overcome these difficulties.

Speechreading Happens Naturally

Developmental studies of young children demonstrate that speechreading is a skill acquired early in life. Presumably, if speechreading were merely a compensatory mechanism for those with hearing loss, then there would be little need for non-disabled children to acquire much in the way of speechreading skill. Although children's speechreading skills develop at a slower pace than their auditory speech perception skills, and they do not demonstrate the same level of speechreading proficiency as adults (McGurk & MacDonald, 1976), they do speechread familiar words by 19 months of age and are aware of the congruence between lip movements and heard words by 3 months of age (Dodd, 1987).

One may wonder, then, what happens to blind children who obviously do not have access to speechread information. Blind children appear to have more difficulties in language acquisition than sighted children. They demonstrate more articulatory problems and make a different pattern of articulation errors than sighted children. Whereas sighted children tend to learn visually distinct articulations fastest, blind children do not show this trend. Sighted children's errors tend to confuse visually similar phonemes while blind children tend to make more errors between acoustically similar phonemes. Blind children may also be affected by the inability to see that sound originates from a talking face, which normally helps infants make the connection between language sounds they hear and their relevance as a form of human communication. They also may be disadvantaged by the inability to imitate adults' oral motor movements and receive visual reinforcement for correct imitations, an important step in learning articulation skills (Mills, 1987).

Given that normally developing children acquire and use speechreading skills, one may wonder what purpose these skills serve when auditory-only speech perception is typically nearly perfect. The answer lies in understanding the complexity of speech understanding and the ease with which the process may be disrupted.

Difficulties of Auditory Speech Perception

Listening to speech entails an amazing perceptual feat. Somehow the brain converts conversational or running speech, which is qualitatively the same as non-speech sounds, into phonemes and words. How does the brain "know" that a sound comes from speech, and how does the brain classify sounds into phonemes? At one time, psychologists thought that the brain learns to associate specific sounds with specific phonemes (Lieberman, 1996). Researchers sought the salient features of each phoneme in hopes of describing a straightforward transcription of sound into words. For example, a "p" might be identified by a certain pattern of pitch, which the brain could recognize as the "p" phoneme.

Cataloging the salient features of each phoneme, however, turned out to be impossible. Phonemes do not have consistent patterns on spectrograms[5]. The sound of a phoneme, particularly consonants, depends heavily on its neighboring phonemes. In other words, the "d"s in "duck" and "dot" have different spectrogram patterns. However, in spite of this phenomenon—known in the literature as coarticulation—listeners identify the "d" in "duck" and "dot" as being the same phoneme, "d". The categorical perception of "d" operates on an "either-or" basis; listeners do not judge the "d-ness" of the heard phoneme. Furthermore, categorical perception extends across phoneme production by various speakers who differ greatly in terms of voice, rate of speech, and care in pronunciation (Goldstein, 2002). How, then, can the brain deal with all of these possible variations and in the end recognize only "d" or "not d"?

Researchers have proposed several theories of speech perception that recognize that speech perception does not operate on a mere one-to-one learned correspondence of sound to phoneme (e.g., Liberman & Mattingly, 1985). I will not go into the details of them here, except to say that the finding that phoneme identification does not correspond directly to sound qualities has allowed researchers to entertain the notion that speechreading can visually provide some of same phonemic information present in heard speech, and that the information from heard and seen speech could be combined for speech understanding.

More importantly, for the purposes of this discussion, coarticulatory effects reveal that speech packs a great deal of information into every second. At one moment, one may hear at the same time the ending of an "r," the middle of an "a, and the very beginning of a "t". Because the pronunciation of one phoneme changes depending on its neighbors, the acoustic slice at any moment contains information about not only the present phoneme, but also predicts its neighbors. This allows speakers of all languages to pack a large amount of phonemic information into relatively short periods of time and for listeners to extract that information at the rate of speech. If speakers did not coarticulate, the rate of speech would be too slow for efficient communication (Liberman & Cooper, 1967). Coarticulation also helps listeners keep track of the time dimension of speech. Running speech consists of sequential information whose order must be remembered in the decoding process. A phoneme's phonic modification from coarticulation tags its place in a sequence of phonemes (McClelland & Elman, 1986).

Speechreading to the Rescue

Auditory speech perception, then, is not a simple task for even the person who has normal hearing. Note, however, that the person must also accomplish this task in a variety of non-optimal listening situations. Background noise, attentional demands, and speaker variability all potentially degrade the quality of the information-packed auditory speech signal. Speechreading can, through redundant information, help the listener segment the continuous speech signal into discrete phonemes. In particular, speechreading can help the listener identify quiet-but-important consonant sounds, which tend to be easily disrupted by noise (Summerfield, 1987).

Summerfield (1987) notes that the consonant sounds which are most easily confused by audition alone are often the sounds which are most easily distinguished by vision alone, and vice versa. For example, the "k" and "p" sounds are very easily confused by audition, but are quite distinct visually. In another example, "th" and "dh" look very much alike but sound very different. The inverse relationship between auditory and visual distinctiveness suggests that audio-visual speech perception may have naturally evolved to allow for speech understanding in noisy conditions. Interestingly, even though speechreading is a non-auditory skill, adding speechreading information to auditory input can improve signal-to-noise ratio by about 15 dB (Summerfield, 1987).

Speechreading can also assist in phoneme identification more indirectly by guiding the listener's visual attention to the lips. By seeing the speaker, listeners can more effectively use binaural localization cues, likely voice characteristics based on speaker gender, and timing characteristics of the speech as aids in extracting the speech signal (Summerfield, 1987). The auditory signal, however, does not necessarily have to be degraded for vision to play an important role in speech perception. Reisberg et al. (1987) found that subjects perform better on shadowing tasks[6] involving a second language (reading French for novice speakers with four years of experience) or a semantically complex message (reading Kant) when they are able to see the speaker than when they must rely on audition alone. Other pieces of experimental evidence suggest that visual information can influence the perception of auditory information even before the phoneme has been identified. For example, speechread information—such as speaking rate—can cause boundary shifts in judgments of voice onset time (VOT), which distinguish between voiceless and voiced initial consonants (Green, 1998). Some word-initial consonants, such as "p" and "b", may be distinguished from each other partly in terms of the amount of time between the beginning of the consonant and the time that one turns on one's voice. Hence, "p" is "silent" with a long VOT, while "b" is "voiced" with a short VOT. The VOT is a continuous variable, which may be manipulated in steps in a laboratory; however, subjects change their "p" identification to "b" or vice versa along a narrow range of VOT values known as the "boundary." Subjects always identify a "p" or "b" on either side of that boundary; subjects, no matter how close the VOT is to the boundary, do not hear a "p-b" blend. This boundary could shift if subjects are predisposed by some means to be more likely to respond "p" such that they report "p" to shorter VOT times than before. Or, subjects could be disposed to more likely to respond "b" such that they report "b" to longer VOT times than before. Green (1998) found that speechread information for speaking rate could cause boundary shifts in auditory phoneme identification. In other words, subjects took into account speechread information at the phonemic feature level of cognitive processing, even before the phoneme itself had been identified.

Conclusion

The McGurk effect introduced a new era in speech perception and speechreading research. Speech perception researchers began to incorporate a role for vision in their models of speech understanding; correspondingly, speechreading researchers recognized that their object of study is not solely of interest to people with hearing losses. Although people with hearing loss may rely more heavily on speechreading than the normal population, it is important to recognize that auditory speech understanding is a complex task even for those lucky enough to score better than 15 dB across all frequencies on an audiogram. Therefore, even people with normal hearing benefit from the information provided by the visual modality. Nature, in Her wisdom, built humans to use that information as a supplement to audition.

For additional information:

McGurk effect:
http://www.psych.ucr.edu/avspeech/VSMcGurk.html
Spectrograms:
http://www.essex.ac.uk/speech/material/kate/general/spectrog.html
Coarticulation:
http://www.ccp.uchicago.edu/overview/language/speech/coarticulation.html

[a] Julie Fitzer is a senior psychology major at Washington and Lee University in Lexington, Virginia, who is now applying to Ph.D. programs in cognitive psychology. Born with a hearing loss, Julie wore hearing aids from age 3 to 20, and then received a Clarion cochlear implant last summer. The full text of her senior project on speech perception and speechreading may be read at http://www.wlu.edu/~fitzerj/403. *Julie thanks her advisor, Dr. David Elmes, Professor of Psychology at Washington and Lee University, for his generous support of her senior project.*

Footnotes

[1] A phoneme is the smallest meaningful unit of contrast in a language. For example, the word "bat" contains the phonemes "b", "a", and "t". If any of these phonemes were changed, a different word would result. To Japanese hearers, the "r" and "l" sounds of English are not meaningfully different. Thus, "r" and "l" are not two different phonemes in Japanese.
[2] Habituation is "getting used" to a stimulus such that one no longer responds to the stimulus as strongly as before.
[3] I still, even now, feel a little uncomfortable discussing my hearing loss in an academic context, for I do not wish my entire personality defined by the fact that I cannot be counted on to wake up in the middle of a nighttime thunderstorm that frightens a bunch of screaming Brownie Girl Scouts left in my charge.
[4] I capitalize "Nature" in the style of Ralph Waldo Emerson.
[5] A spectrogram is a visual display of sound that shows the intensities of various frequencies over time.
[6] In a shadowing task, the subject repeats what he or she hears as the message is played.

References

Dodd, B. (1987). The acquisition of lip-reading skills by normally hearing children. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 163-175). London: Erlbaum.

Green, K. P. (1998). The use of auditory and visual information during phonetic processing: Implications for theories of speech perception. In R. Campbell, B. Dodd, & D. Burnham, D (Eds.), *Hearing by eye II: Advances in the psychology of speechreading and auditory-visual speech* (pp. 3-22). East Sussex, UK: Psychology Press.

Goldstein, E. B. (2002). *Sensation and perception* (6th ed.) Pacific Grove, CA: Wadsworth.

Liberman, A. M. & Cooper, F. S. (1967). Perception of the speech code. *Psychological Review, 74,* 431-461. Reprinted in A. M. Liberman (1996), *Speech: A special code* (pp. 203-232). Cambridge, MA: The MIT Press.

Liberman, A. M. & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1-36. Reprinted in A. M. Liberman (1996), *Speech: A special code* (pp. 237-265). Cambridge, MA: The MIT Press.

Liberman, A. M. (1996). Speech: A special code. Cambridge, MA: The MIT Press.

McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1-86.

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

Mills, A. E. (1987). The development of phonology in the blind child. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 145-161). London: Erlbaum.

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* (pp. 97-113). London: Erlbaum.

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 3-51). London: Erlbaum.