Investigating Speechreading and Deafness

Edward T. Auer Jr.*

Abstract

Background: The visual speech signal can provide sufficient information to support successful communication. However, individual differences in the ability to appreciate that information are large, and relatively little is known about their sources.

Purpose: Here a body of research is reviewed regarding the development of a theoretical framework in which to study speechreading and individual differences in that ability. Based on the hypothesis that visual speech is processed via the same perceptual-cognitive machinery as auditory speech, a theoretical framework was developed by adapting a theoretical framework originally developed for auditory spoken word recognition.

Conclusion: The evidence to date is consistent with the conclusion that visual spoken word recognition is achieved via a process similar to auditory word recognition provided differences in perceptual similarity are taken into account. Words perceptually similar to many other words and that occur infrequently in the input stream are at a distinct disadvantage within this process. The results to date are also consistent with the conclusion that deaf individuals, regardless of speechreading ability, recognize spoken words via a process similar to individuals with hearing.

Key Words: Deafness, lipreading, speechreading

Abbreviation: NAM = Neighborhood Activation Model

Individual speechreading ability—that is, the understanding of spoken sentences solely by viewing the talker—is known to vary between 0 and close to 90% words correct in sentences (MacLeod and Summerfield, 1987; Bernstein et al, 2000; Auer and Bernstein, 2007). At a coarse level of description, spoken language understanding entails the encoding of the physical stimulus, followed by word recognition, and finally decoding of the intended message. It is likely that variations in one or more of these component perceptual and cognitive processes underlie individual differences in speechreading ability. A well-specified theoretical framework for speechreading would facilitate the discovery of those specific variations associated with individual differences. The approach my colleagues and I have taken to developing a theoretical framework for speechreading (Auer and Bernstein, 1997; Auer, 2002; Mattys et al, 2002) has been to focus on what we believe to be a central component of spoken language understanding, specifically the process of spoken word recognition (Bernstein and Auer, 1996; Bernstein et al, 2000). Based on the hypothesis that visual speech is processed via the same perceptual-cognitive machinery as auditory speech, the theoretical framework for visual spoken word recognition has been developed through the adaptation of a framework initially developed for auditory spoken word recognition. Here, the development of this framework and initial investigations of how it might be altered by the influence of deafness are reviewed. Three key elements in spoken word recognition are the primary focus in the current review: (1) lexical candidates for recognition are activated as a function of their perceptually defined similarity with incoming phonetic information; (2) ease of recognition is a function of the competition among active lexical candidates; (3) lexical knowledge, or vocabulary, defines the context in which the lexical activation and competition occur.

To successfully recognize a spoken word, the perceiver must decode the incoming sensory signal for

*Department of Speech-Language-Hearing, University of Kansas

Edward T. Auer Jr, Department of Speech-Language-Hearing, University of Kansas, Dole Human Development Center, 10000 Sunnyside Ave., Room 3001, Lawrence, KS 66045-7555; Phone: 785-864-1460; Fax: 785-864-3974; E-mail: auer@ku.edu

the purpose of isolating the intended word from the tens of thousands of words stored in his or her mental lexicon. Roughly 40 yr of research investigating auditory word recognition has converged on the common view that recognition can be conceptualized as an activation-competition process (Luce and McLennan, 2005). Though the specific details of implementation vary across contemporary theoretical models, general agreement exists that as the acoustic speech stimulus unfolds in time, multiple word candidates are activated as a function of their form-based, or perceptual, similarity with the stimulus. Thus, activation for any given word in the mental lexicon is hypothesized to increase as perceptual similarity with the bottom-up stimulus input increases. Recognition is achieved via a competition among the active candidates based on their stimulus-driven activation levels combined with a bias favoring words that occur more frequently in language input. The competition is typically won by the word in the lexicon with the best match to the perceptual input, with the speed and accuracy of these recognition processes being modulated by the number of words competing and their frequencies of occurrence in the language. In the remainder of this review, the three key elements within this framework are examined in relation to the investigation of speechreading.

The first key element is that the activation of lexical candidates is a function of the perceptually defined similarity between the phonetic information in the incoming stimulus and the mental representation. This is particularly relevant for speechreading because the stimulus information available from auditory and visual speech signals differs in both the availability of phonetic information and the patterns of perceptual similarity among speech segments. Under the best of perceptual conditions, the phonetic information needed to perceive some phonemic distinctions is not reliably visible, with the exact amount of perceptually available distinctiveness known to vary as a function of perceiver, talker, and segmental context (Kricos and Lesner, 1982, 1985; Montgomery and Jackson, 1983; Jackson, 1988; Auer et al, 1997). Typically, visually available phonetic information is characterized as extremely impoverished (Fisher, 1968; Owens and Blazek, 1985; Kuhl and Meltzoff, 1988; Massaro, 1998). Along with reduced phonetic distinctiveness, the patterns of similarity among speech segments are known to be different when viewing speech compared to listening to speech degraded by noise (Breeuwer and Plomp, 1985; Grant and Braida, 1991; Grant and Walden, 1996). For example, the consonants /p/, /t/, and /k/ are perceptually similar for auditory speech presented in noise, but these consonants remain visually distinct. In contrast, /b/, /p/, and /m/ are perceptually similar when viewing a talker, but these consonants remain auditorily distinct.

With respect to the reduction in available phonetic information, an intriguing implication of the spoken word recognition theoretical framework is that accurate word recognition can occur even in cases where it is not possible to uniquely identify all of the word's constituent consonants and vowels. That is, recognition can occur provided sufficient segmental information is available to uniquely identify the word from other words in the language. For example, if it was not possible to distinguish /p/, /t/, and /k/, the word "peace" would still be recognizable, because "teace" and "keace" are not words in English and do not enter into the competition for recognition. Thus, word recognition is predicted to be a function of both the segmental intelligibility and the distribution of words in the perceiver's mental lexicon.

To investigate the relation between the reduced phonemic distinctiveness and constraints provided by the English lexicon for speechreading, a computational modeling procedure was developed (Auer and Bernstein, 1997). The method involved four steps. First, rules are developed to retranscribe words so that their transcriptions represent only the segmental distinctions that are estimated to be visually perceptible. The retranscription rules are in the form of *phoneme equivalence classes (e.g., b, p, m)*. Second, retranscription rules are applied to the words in a phonemically transcribed, computer-readable lexicon. Third, the retranscribed words are sorted so that words rendered identical (no longer distinct) are placed in the same *lexical equivalence class (e.g., bat, pat, mat)*. Finally, quantitative measures are used to determine the information that remains in the retranscribed lexicon. Based on an analysis conducted using 12 phonemic equivalence classes, which represents the typical visually available phonemic distinctiveness, approximately 54% of words (frequency weighted) remain distinct. Using 19 phonemic equivalence classes, to represent highly skilled speechreaders, close to 90% of words (frequency weighted) were found to remain distinct. The results of this research provided evidence that stimulus-based lexical dissimilarity can contribute to reduction of the problem of phonetic impoverishment in visual spoken word recognition (Auer and Bernstein, 1997). Iverson, Bernstein, and I (1998) extended the analysis to compare mono- and multisyllabic words. We found that only 15% of all monosyllables preserve their visual uniqueness whereas over 75% of all longer words do, predicting that on average longer words should be easier to speechread.

Another prediction based on this computational modeling that derives from the manipulation of the number of phonemic equivalence classes in the computational analysis is that individuals with slightly increased ability in segment-level identification could reap large gains in word- and sentence-level recognition

accuracy. Bernstein and colleagues (2002) demonstrated slightly more accurate consonant identification for early-onset deaf perceivers compared to perceivers with normal hearing, suggesting that the early-onset deaf participants should be capable of significant increases in speechreading accuracy for words and sentences. Recent studies of visual speech perception by college-educated, congenitally deaf adults (Bernstein et al, 2001; Bernstein et al, 2002; Auer and Bernstein, 2007) have demonstrated that speechreading can be highly accurate, reaching greater than 80% words correct in isolated sentences. This level of performance has only been observed within deaf participants with early-onset losses. Taken together, the evidence suggests that segment-level ability may be one piece of the puzzle for understanding individual differences in speechreading ability.

A second key element in spoken word recognition is that recognition occurs through competition among active lexical candidates. This leads to the prediction that the ease of a word's recognition is influenced by its perceptually defined lexical context, or neighborhood density. Neighborhood density refers to the number of words perceptually similar to a given target word in the perceiver's mental lexicon (Luce and Pisoni, 1998). A target word's neighborhood is operationally defined, in most studies, as the number of real words that can be formed by a single phoneme substitution, addition, or deletion from the target word. Words similar to many other words are referred to as being in dense neighborhoods, whereas words similar to few other words are referred to as being in sparse neighborhoods. Words within dense neighborhoods are predicted to be harder to recognize than words within sparse neighborhoods. Additionally, words that occur frequently in the linguistic environment are afforded an advantage in the recognition process, such that low-frequency words are predicted to be harder to recognize than high-frequency words. Within the literature on auditory spoken word recognition, both of these predictions (neighborhood density and frequency) have now been verified with ample empirical evidence (Howes, 1957; Savin, 1963; Forster, 1976; Balota and Chumbley, 1984; McClelland and Elman, 1986; Marslen-Wilson, 1989; Norris, 1994; Luce and Pisoni, 1998; Gaskell and Marslen-Wilson, 2002; Norris and McQueen, forthcoming).

An initial behavioral investigation of the hypothesis that visual spoken word recognition is influenced by neighborhood density was reported in Auer (2002). In that study the Neighborhood Activation Model (NAM) was adapted to model visual spoken word recognition. Neighborhood density (Luce and Pisoni, 1998) was computed for a set of monosyllabic visual spoken words using visual phonetic similarity estimates derived from visual nonsense syllable perceptual confusion matrices. Twelve participants with normal hearing and 12 deaf participants visually identified isolated spoken words that had either sparse or dense neighborhoods. Words from sparse neighborhoods were identified considerably more accurately than dense-neighborhood words by both participant groups. The results were interpreted as evidence that competition based on perceptually defined similarity between words and input constitutes a reliable spoken word recognition principle irrespective of input modality or participant hearing status. The results also provided evidence that a word's competitor environment is defined on the basis of form-based, or perceptual, similarity among words rather than being an abstract lexical property. Specifically, when auditory confusion matrices were used to compute lexical density, the output of the NAM was no longer predictive of identification accuracy. This is consistent with the conclusion that the specific set of words entering into competition for recognition depends on the phonetic information available in the stimulus. Thus a word's neighborhood appears to be dynamically defined in the recognition process. More recently, an adaptation of the NAM using an operational definition of similarity analogous to using phonemic equivalence classes for the substitution metric has also been shown to be predictive of word identification (Tye-Murray et al, 2007).

Mattys, Bernstein, and I (2002) used lexical equivalence class size, as defined above, rather than the NAM output values to operationally define the size of a stimulus word's neighborhood and manipulated word frequency. The use of lexical equivalence class size expanded the generality of the test of lexical density by facilitating the inclusion of both mono- and multisyllabic words as stimuli. Eight participants with normal hearing and eight deaf participants visually identified isolated spoken words that were selected to be in large, medium, or unique lexical equivalence classes. For each of the lexical equivalence class sizes, separate sets of words were selected that were either high or low in frequency and were either mono- or disyllabic. Accuracy of mono- and disyllabic word recognition was found to be a function of both a word's lexical equivalence class size and its frequency of occurrence. Word-identification accuracy increased as stimulus word lexical equivalence class size decreased and frequency of occurrence increased. I (forthcoming) replicated the Mattys and colleagues study and extended the observed frequency and lexical equivalence class effects to a larger pool of deaf participants who varied over a wide range in their speechreading ability. In my study, the overall levels of identification accuracy rose and fell as a function of participants' speechreading abilities; however, the pattern of lexical equivalence class size and frequency effects remained stable over the range of speechreading abilities. Thus, the results to date are consistent with the conclusion that an activation-competition framework is a general mechanism for spoken word recognition

regardless of input modality and that lexical similarity is form based (Auer, forthcoming).

A third key element in spoken word recognition is that lexical knowledge, or vocabulary, is an important factor in determining the ease of word recognition. Lexical knowledge is particularly important in investigating word recognition by deaf participants because it is a function of linguistic experience, which clearly differs for individuals with early-onset deafness compared to individuals with hearing. In studies of auditory spoken word recognition, the adult perceiver's lexical knowledge is estimated using objective measures of word knowledge and exposure (Gernsbacher, 1984). Objective measures are taken by analyzing large-scale linguistic corpora hypothesized to be representative of the linguistic experience of the participant pool. However, objective corpora do not currently exist that are designed to represent the linguistic experience of early-onset deaf perceivers. We have examined the use of subjective estimates of word experience as an alternative approach to estimating lexical knowledge (Auer et al, 2000; Auer and Bernstein, 2008).

Bernstein and colleagues (Auer et al, 2000) used a subjective familiarity measure to investigate the lexical knowledge of skilled deaf speechreaders. Specifically, 50 deaf and 50 hearing individuals, all of whom were average or better speechreaders, rated individual words on a seven-point scale that ranged from "never seen, heard, or read the word" to "know the word and am confident of its meaning." As expected, deaf participants consistently judged words to be less familiar compared to hearing participants. However, the pattern of group average item (word) familiarity ratings was similar across the groups ($r = 0.90$). But more detailed analysis investigating item correlations within and across participant groups demonstrated the existence of subtle differences. Specifically, the patterns of familiarity were more similar within a participant group compared to across participant groups. Thus, a deaf participant's familiarity with words was more similar to that of other deaf participants than to that of hearing participants. The significance of these differences for spoken word recognition remains to be examined. On the whole, the pattern of familiarity with words is similar for both participant groups tested, suggesting similar lexical knowledge with only minor differences emerging upon detailed analysis.

More recently, Bernstein and I (Auer and Bernstein, 2008) used two additional subjective measures to further investigate the influence of linguistic experience on the lexicon of early-onset deaf participants. Substantial behavioral evidence exists of a reliable relationship between the age at which words are subjectively estimated to have been acquired and the efficiency with which those words are recognized (for reviews, see Morrison and Ellis, 1995; Juhasz, 2005). Typically, early-

learned words are recognized more quickly and easily that late-learned words. Prior to investigating whether the same relationship holds for visual spoken word recognition, Bernstein and I (Auer and Bernstein, 2008) collected initial estimates of age of acquisition for a set of words from deaf and hearing participants. It also was hypothesized that the deaf participants could have acquired their vocabularies through a diverse set of communication channels, including reading, speech-reading, and/or the use of some type of English-based manual sign system. To compare their experience with the experience of those with lifelong hearing, a new lexical experience measure was collected: acquisition channel. Thus in this study, subjective measures of when (age of acquisition) and how (acquisition channel, i.e., spoken, printed, or signed) words were learned were collected from 50 deaf and 50 hearing individuals, all of whom were average or better speechreaders. The age-of-acquisition ratings within each participant group were highly correlated with the normative acquisition order taken from the Peabody Picture Vocabulary Test—revised (deaf group, $r = 0.950$; hearing group, $r = 0.946$ [Dunn and Dunn, 1981]). The average age-of-acquisition ratings for stimulus items were also highly correlated across participant groups ($r = 0.971$). However, the two participant groups differed in when words were rated as learned, with the deaf participants words rated as learned later. The groups also differed in their rating of how words were learned, with the deaf participants relying more on the print and manual channels for acquisition. Intriguingly, those deaf participants who subjectively reported relying more on the spoken channel for acquisition were also better speechreaders as measured with an objective measure. The directionality of this relationship remains to be investigated. Taken together, these subjective measures are important in that they provide a means to look at lexical experience and look back at the development of the lexicon in adult participants for whom no direct developmental data are available.

FUTURE DIRECTIONS

The perceiver's task during spoken word recognition is to select a word from the tens of thousands of words he or she knows on the basis of the incoming perceptual information. Research to date is consistent with the claim that this task is accomplished via a process of activation followed by competition regardless of the modality of input. Here, the adaptations of three key elements associated within this theoretical framework were reviewed in relation to speechreading and deafness. This approach has proved successful for developing a theoretical framework, but much work remains both in developing the framework and in investigating individual differences.

For example, to understand spoken language, the perceiver must recognize words spoken not only in isolation, as investigated here, but also in a continuous speech stream. An issue that arises associated with continuous speech is the need to determine where words begin and end. Detection of meaningful boundaries is crucial to successful speech comprehension. A single consistent acoustic-phonetic event that signals the presence of junctures or the boundaries between words has not been found (Lehiste, 1970; Nakatani and Dukes, 1977). Numerous studies investigating boundary cues have resulted in evidence suggesting that word segmentation is aided by multiple cues, including phonotactic cues (Jusczyk et al, 1994), transitional probabilities (Saffran et al, 1996), and rhythmic cues (Cutler and Norris, 1988; Cutler, 1990).

Successful juncture perception is particularly important for speech comprehension under conditions of reduced phonetic distinctiveness. Extension of the lexical effects detailed above implicitly assumes that the perceiver can reliably segment words from the continuous speech stream. If accurate segmentation is not achievable, the benefit of lexical structure is likely reduced, if not completely mediated. In a computational study, Harrington and Johnstone (1987) computed the number of possible word-level parses of an input utterance when it was transcribed using transcription unit with reduced phonetic distinctiveness. For utterances that were between 7 and 10 words in length, the number of alternative parses extended into the millions. Recent evidence suggests that lexical effects are still present but reduced within sentence-length stimuli (Auer and Reed, 2008). Thus, future speechreading studies should investigate the segmentation strategies used in sentence-length materials.

In conclusion, the evidence to date suggests that the process of visual spoken word recognition is achieved via a process similar to auditory word recognition provided differences in perceptual or form-based similarity are taken into account. Words perceptually similar to many other words and that occur infrequently in the input stream are at a distinct disadvantage within this process. The results to date also are consistent with the conclusion that deaf individuals, regardless of speechreading ability, recognize spoken words via a similar process. Work is currently ongoing to further develop and adapt the theoretical framework as well as examine whether variations in its component perceptual and cognitive processes are responsible for the individual differences in speechreading.

REFERENCES

Auer ET, Jr. (2002) The influence of the lexicon on speech read word recognition: contrasting segmental and lexical distinctiveness. *Psychon Bull Rev* 9(2):341–347.

Auer ET, Jr. (2009) Spoken word recognition by eye. *Scand J Psychol* 50(5):419–425.

Auer ET, Jr, Bernstein LE. (1997) Speechreading and the structure of the lexicon: computationally modeling the effects of reduced phonetic distinctiveness on lexical uniqueness. *J Acoust Soc Am* 102(6):3704–3710.

Auer ET, Jr, Bernstein LE. (2007) Enhanced visual speech perception in individuals with early-onset hearing impairment. *J Speech Lang Hear Res* 50(5):1157–1165.

Auer ET, Jr, Bernstein LE. (2008) Estimating when and how words are acquired: a natural experiment on the development of the mental lexicon. *J Speech Lang Hear Res* 51(3):750–758.

Auer ET, Jr, Bernstein LE, Tucker PE. (2000) Is subjective word familiarity a meter of ambient language? A natural experiment on effects of perceptual experience. *Mem Cognit* 28(5):789–797.

Auer ET, Jr, Bernstein LE, Waldstein RS, Tucker PE. (1997, September 26–27). *Effects of phonetic variation and the structure of the lexicon on the uniqueness of words.* Paper presented at the ESCA/ESCOP Workshop on Audio-Visual Speech Processing, Rhodes, Greece.

Auer ET, Jr, Reed R. (2008) Investigating lexical influences on the accuracy of speechreading words presented in isolation and in sentence context. Abstract. *J Acoust Soc Am* 124(4):2459.

Balota DA, Chumbley JI. (1984) Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *J Exp Psychol Hum Percept Perform* 10(3):340–357.

Bernstein LE, Auer ET, Jr. (1996) Word recognition in speechreading. In: Stork DG, Hennecke ME, eds. *Speechreading by Humans and Machines.* NATO ASI Series F: Computer and System Sciences, vol. 150. Berlin: Springer-Verlag.

Bernstein LE, Auer ET, Jr, Moore JK, Ponton CW, Don M, Singh M. (2002) Visual speech perception without primary auditory cortex activation. *Neuroreport* 13(3):311–315.

Bernstein LE, Auer ET, Jr, Tucker PE. (2001) Enhanced speechreading in deaf adults: can short-term training/practice close the gap for hearing adults? *J Speech Lang Hear Res* 44(1):5–18.

Bernstein LE, Demorest ME, Tucker PE. (2000) Speech perception without hearing. *Percept Psychophys* 62(2):233–252.

Breeuwer M, Plomp R. (1985) Speechreading supplemented with formant-frequency information from voiced speech. *J Acoust Soc Am* 77(1):314–317.

Cutler A. (1990) Exploiting prosodic probabilities in speech segmentation. In: Altmann GTM, ed. *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives.* Cambridge, MA: MIT Press, 105–121.

Cutler A, Norris DG. (1988) The role of strong syllables in segmentation for lexical access. *J Exp Psychol Hum Percept Perform* 14(1): 113–121.

Dunn LM, Dunn LM. (1981) *Peabody Picture Vocabulary Test—Revised.* Circle Pines, MN: American Guidance Service.

Fisher CG. (1968) Confusions among visually perceived consonants. *J Speech Hear Res* 11:796–804.

Forster KI. (1976) Accessing the mental lexicon. In: Wales RJ, Walker E, eds. *New Approaches to Language Mechanisms.* Amsterdam: North-Holland, 257–284.

Gaskell MG, Marslen-Wilson WD. (2002) Representation and competition in the perception of spoken words. *Cognit Psychol* 45(2): 220–266.

Gernsbacher MA. (1984) Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *J Exp Psychol Gen* 113(2):256–281.

Grant KW, Braida LD. (1991) Evaluating the articulation index for auditory-visual input. *J Acoust Soc Am* 89(6): 2952–2960.

Grant KW, Walden BE. (1996) Evaluating the articulation index for auditory-visual consonant recognition. *J Acoust Soc Am* 100(4, Pt. 1):2415–2424.

Harrington J, Johnstone A. (1987) The effects of equivalence classes on parsing phonemes into words in continuous speech recognition. *Comput Speech Lang* (2):273–288.

Howes D. (1957) On the relation between the intelligibility and frequency of occurrence of English words. *J Acoust Soc Am* 29(2): 296–305.

Iverson P, Bernstein LE, Auer ET, Jr. (1998) Modeling the interaction of phonemic intelligibility and lexical structure in audiovisual word recognition. *Speech Commun* 26(1–2):45–63.

Jackson PL. (1988) The theoretical minimal unit for visual speech perception: visemes and coarticulation. *Volta Rev* 90 (5):99–115.

Juhasz BJ. (2005) Age-of-acquisition effects in word and picture identification. *Psychol Bull* 131(5):684–712.

Jusczyk PW, Luce PA, Charles-Luce J. (1994) Infants' sensitivity to phonotactic patterns in the native language. *J Mem Lang* 33: 630–645.

Kricos PB, Lesner SA. (1982) Differences in visual intelligibility across talkers. *Volta Rev* 84(4):219–225.

Kricos PB, Lesner SA. (1985) Effect of talker differences on speechreading by hearing-impaired teenagers. *Volta Rev* 87(1): 5–17.

Kuhl PK, Meltzoff AN. (1988). Speech as an intermodal object of perception. In: Yonas A, ed. *Perceptual Development in Infancy* (The Minnesota Symposia on Child Psychology, Vol. 20). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 235–266.

Lehiste I. (1970) *Suprasegmentals*. Cambridge, MA: MIT Press.

Luce PA, McLennan CT. (2005) Spoken word recognition: the challenge of variation. In: Pisoni DB, Remez RE, eds. *The Handbook of Speech Perception*. Malden, MA: Blackwell, 591–609.

Luce PA, Pisoni DB. (1998) Recognizing spoken words: the Neighborhood Activation Model. *Ear Hear* 19(1):1–36.

MacLeod A, Summerfield Q. (1987) Quantifying the contribution of vision to speech perception in noise. *Br J Audiol* 21(2):131–141.

Marslen-Wilson WD. (1989) Access and integration: projecting sound onto meaning. In: Marslen-Wilson WD, ed. *Lexical Access and Representation*. Cambridge, MA: Bradford Books, 3–24.

Massaro DW. (1998) *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. Cambridge, MA: MIT Press.

Mattys SL, Bernstein LE, Auer ET, Jr. (2002) Stimulus-based lexical distinctiveness as a general word-recognition mechanism. *Percept Psychophys* 64(4):667–679.

McClelland JL, Elman JL. (1986) The TRACE model of speech perception. *Cognit Psychol* 18(1):1–86.

Montgomery AA, Jackson PL. (1983) Physical characteristics of the lips underlying vowel lipreading performance. *J Acoust Soc Am* 73(6):2134–2144.

Morrison CM, Ellis AW. (1995) Roles of word frequency and age of acquisition in word naming and lexical decision. *J Exp Psychol Learn Mem Cogn* 21(1):116–133.

Nakatani LH, Dukes KD. (1977) Locus of segmental cues for word juncture. *J Acoust Soc Am* 62(3):714–719.

Norris D. (1994) Shortlist: a connectrionist model of continuous speech recognition. *Cognition* 52:189–234.

Norris D, McQueen JM. (2008) Shortlist B: a Bayesian model of continuous speech recognition. *Psychol Rev* 115(2):357–395.

Owens E, Blazek B. (1985) Visemes observed by hearing-impaired and normal hearing adult viewers. *J Speech Hear Res* 28:381–393.

Saffran JR, Aslin RN, Newport EL. (1996) Statistical learning by 8-month-old infants. *Science* 274(5294):1926–1928.

Savin HB. (1963) Word-frequency effect and errors in the perception of speech. *J Acoust Soc Am* 35(2):200–206.

Tye-Murray N, Sommers M, Spehar B. (2007) Auditory and visual lexical neighborhoods in audiovisual speech perception. *Trends Amplif* 11(4):233–241.