Published in final edited form as:

J Am Acad Audiol. 2010 March ; 21(3): 163–168. doi:10.3766/jaaa.21.3.4.

Investigating Speechreading and Deafness

Edward T. Auer Jr.

Department of Speech-Language-Hearing, University of Kansas

Abstract

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

Purpose—Here a body of research is reviewed regarding the development 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, the development of the theoretical framework has focused on adapting a theoretical framework originally developed for auditory spoken word recognition.

Conclusions—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 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 are also consistent with the conclusion that deaf individuals, regardless of speechreading ability, recognize spoken words via a process similar to individuals with hearing.

Individual speechreading ability—that is the understanding of spoken sentences solely by viewing the talker—is known to vary between zero and close to ninety percent words correct in sentences (Auer & Bernstein, 2007; Bernstein, Demorest, & Tucker, 2000; MacLeod & Summerfield, 1987). At a coarse level of description, spoken language understanding entails the encoding of the physical stimulus, followed by word recognition, and finally decoding the intended message. It is likely that variations in one or more of the component perceptual and cognitive processes involved in this chain of events 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 we have taken to developing a theoretical framework for speechreading (Auer, 2002; Auer & Bernstein, 1997; Mattys, Bernstein, & Auer, 2002) has been to focus on a central component of spoken language understanding, specifically the process of spoken word recognition (Bernstein & 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. These elements are: 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 amongst active of lexical candidates; 3) lexical knowledge, or vocabulary, defines the context in which the lexical activation and competition occurs.

To successfully recognize a spoken word, the perceiver must decode the incoming sensory signal for the purpose of isolating the intended word from the tens of thousands of words

stored in their mental lexicon. Roughly forty years of research investigating auditory word recognition has converged on the common view that recognition is the result of an activation-competition process (Luce & 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 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 this recognition processes being modulated by the number of words competing and their frequencies of occurrence in the language. In the following paragraphs, the three key elements within this framework are examined in relation to the investigation of speechreading.

One key element in spoken word recognition 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 differ 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 (Auer, Bernstein, Waldstein, & Tucker, 1997; Jackson, 1988; Kricos & Lesner, 1982, 1985; Montgomery & Jackson, 1983). Typically, visually available phonetic information is characterized as extremely impoverished (Fisher, 1968; Kuhl & Meltzoff, 1988; Massaro, 1998; Owens & Blazek, 1985). Along with reduced phonetic distinctiveness, the patterns of similarity among speech segments are known to be different when viewing speech compared with listening to speech degraded by noise (Breeuwer & Plomp, 1985; Grant & Braida, 1991; Grant & Walden, 1996). For example, the consonants /p/, /t/, and /k/ are perceptually similar for auditory speech presented in noise, however these consonants remain visually distinct. In contrast, /b/, /p/, and /m/ are perceptually similar when viewing a talker, however 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 & Bernstein, 1997). The method involved four steps: (1) 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*). (2) Retranscription rules are applied to the words in a phonemically transcribed, computer-readable lexicon. (3) 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*). (4) 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 & Bernstein, 1997). Iverson, Bernstein, and Auer (1998) extended the analysis to compare mono- and multisyllabic words. Iverson et al., 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.

A prediction that derives from the manipulation of the number 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 et al, (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 (Auer & Bernstein, 2007; Bernstein et al., 2002; Bernstein, Auer, & Tucker, 2001), 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 amongst active of 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 & 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 (Balota & Chumbley, 1984; Forster, 1976; Gaskell & Marslen-Wilson, 2002; Howes, 1957; Luce & Pisoni, 1998; Marslen-Wilson, 1989; McClelland & Elman, 1986; Norris, 1994; Norris & McQueen, In Press; Savin, 1963).

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 & 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 twelve deaf participants visually identified isolated spoken words that had either sparse or dense neighborhoods. Words from sparse-neighborhood were identified considerably more accurately than dense-neighborhood words by both participant groups. The results were interpreted as evidence that competition based on perceptual 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, Sommers, & Spehar, 2007).

Mattys, Bernstein, and Auer (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 multi-syllabic 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 di-syllabic. Accuracy of mono- and di-syllabic 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. Auer (in press) replicated the Mattys et al. 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 Auer's 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, in press).

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 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. An alternative approach to estimating lexical knowledge is through the use of subjective estimates of word experience. We and others have proposed that subjective measures of lexical knowledge are well suited for use in the study of word recognition by clinical populations (Auer & Bernstein, 2008; Auer, Bernstein, & Tucker, 2000; Gernsbacher, 1984).

Auer and Bernstein (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 7-point scale that ranged from never seen, heard, or read the word to know the words 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 ($r = .90$). However, 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 familiarity with words was more similar to other deaf participants than to 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, 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 Juhasz, 2005; Morrison & Ellis, 1995). Typically, early learned words are recognized more quickly and easily than late learned words. Prior investigating whether the same relationship holds visual spoken word recognition, 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, speechreading, and/or the use of some type of an English-based manual sign system. To compare their experience with 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 (Dunn & Dunn, 1981) (deaf group, $r = .950$; hearing group, $r = .946$). The average age of acquisition ratings for stimulus items were also highly correlated across participant groups ($r = .971$). However, the two participant groups differed in when words were rated as learned, with the deaf participants rating words 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 they know 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 adaptation of three key elements associated with this theoretical framework is reviewed in relation to

speechreading and deafness. This approach has proved successful for developing a theoretical framework, however much work remains both in developing the framework and in investigating individual differences.

To understand spoken language, the perceiver must recognize words spoken in 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 & Dukes, 1977). The result of numerous studies investigating boundary cues has resulted in evidence suggesting that word segmentation is aided by multiple cues, including phonotactic cues (Jusczyk, Luce, & Charles-Luce, 1994), transitional probabilities (Saffran, Aslin, & Newport, 1996), and rhythmic cues (Cutler, 1990; Cutler & Norris, 1988).

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 & 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 & 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 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 its component perceptual and cognitive processes are responsible for the individual differences in speechreading.

Acknowledgments

Portions of the research reviewed here were supported by a grant from NIH/NIDCD DC004856.

References

- Auer E. Spoken word recognition by eye. *Scandinavian Journal of Psychology*. (in press).
- Auer E, Reed R. Investigating lexical influences on the accuracy of speechreading words presented in isolation and in sentence context. *Journal of the Acoustical Society of America*. 2008; 124(4):2459.
- Auer ET Jr. The influence of the lexicon on speech read word recognition: contrasting segmental and lexical distinctiveness. *Psychonomic Bulletin Review*. 2002; 9(2):341–347. [PubMed: 12120798]
- Auer ET Jr, Bernstein LE. Speechreading and the structure of the lexicon: computationally modeling the effects of reduced phonetic distinctiveness on lexical uniqueness. *Journal of the Acoustical Society of America*. 1997; 102(6):3704–3710. [PubMed: 9407662]
- Auer ET Jr, Bernstein LE. Enhanced visual speech perception in individuals with early-onset hearing impairment. *Journal of Speech Language Hearing Research*. 2007; 50(5):423–435.

- Auer ET Jr, Bernstein LE. Estimating when and how words are acquired: a natural experiment on the development of the mental lexicon. *Journal of Speech Language Hearing Research*. 2008; 51(3): 750–758.
- Auer ET Jr, Bernstein LE, Tucker PE. Is subjective word familiarity a meter of ambient language? A natural experiment on effects of perceptual experience. *Memory and Cognition*. 2000; 28(5):789–797. [PubMed: 10983453]
- Auer, ET., Jr; Bernstein, LE.; Waldstein, RS.; Tucker, PE. Effects of phonetic variation and the structure of the lexicon on the uniqueness of words. September 26–27, 1997; Paper presented at the ESCA/ESCOP Workshop on Audio-Visual Speech Processing; Rhodes, Greece. 1997.
- Balota DA, Chumbley JI. Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*. 1984; 10(3):340–357. [PubMed: 6242411]
- Bernstein, LE.; Auer, ET, Jr. Word recognition in speechreading; NATO ASI Series F. In: Stork, D.; Hennecke, M., editors. *Speechreading by humans and machines*. Vol. 150. Berlin: Springer-Verlag; 1996.
- Bernstein LE, Auer ET Jr, Moore JK, Ponton CW, Don M, Singh M. Visual speech perception without primary auditory cortex activation. *NeuroReport*. 2002; 13(3):311–315. [PubMed: 11930129]
- Bernstein LE, Auer ET Jr, Tucker PE. Enhanced speechreading in deaf adults: can short-term training/practice close the gap for hearing adults? *Journal of Speech, Language, and Hearing Research*. 2001; 44(1):5–18.
- Bernstein LE, Demorest ME, Tucker PE. Speech perception without hearing. *Perception & Psychophysics*. 2000; 62(2):233–252. [PubMed: 10723205]
- Breeuwer M, Plomp R. Speechreading supplemented with formant-frequency information from voiced speech. *Journal of the Acoustical Society of America*. 1985; 77(1):314–317. [PubMed: 3973225]
- Cutler, A. Exploiting prosodic probabilities in speech segmentation. In: Altmann, GTM., editor. *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*. Cambridge, MA: MIT Press; 1990. p. 105–121.
- Cutler A, Norris D. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*. 1988; 14(1):113–121.
- Dunn, LM.; Dunn, LM. *Peabody Picture Vocabulary Test-Revised*. Circle Pines, MN: American Guidance Service; 1981.
- Fisher CG. Confusions among visually perceived consonants. *Journal of Speech and Hearing Research*. 1968; 11:796–804. [PubMed: 5719234]
- Forster, KI. Accessing the mental lexicon. In: Wales, RJ.; Walker, E., editors. *New approaches to language mechanisms*. Amsterdam: North-Holland; 1976. p. 257–284.
- Gaskell MG, Marslen-Wilson WD. Representation and competition in the perception of spoken words. *Cognitive Psychology*. 2002; 45(2):220–266. [PubMed: 12528902]
- Gernsbacher MA. Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*. 1984; 113(2):256–281. [PubMed: 6242753]
- Grant KW, Braid LD. Evaluating the articulation index for auditory-visual input. *Journal of the Acoustical Society of America*. 1991; 89(6):2952–2960. [PubMed: 1918633]
- Grant KW, Walden BE. Evaluating the articulation index for auditory-visual consonant recognition. *Journal of the Acoustical Society of America*. 1996; 100(4 Pt 1):2415–2424. [PubMed: 8865647]
- Harrington J, Johnstone A. The effects of equivalence classes on parsing phonemes into words in continuous speech recognition. *Computer Speech Language*. 1987; 2:273–288.
- Howes D. On the relation between the intelligibility and frequency of occurrence of English words. *Journal of the Acoustical Society of America*. 1957; 29(2):296–305.
- Jackson PL. The theoretical minimal unit for visual speech perception: Visemes and coarticulation. *Volta Review*. 1988; 90(5):99–115.
- Juhász BJ. Age-of-acquisition effects in word and picture identification. *Psychological Bulletin*. 2005; 131(5):684–712. [PubMed: 16187854]

- Jusczyk PW, Luce PA, Charles-Luce J. Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*. 1994; 33:630–645.
- Kricos PB, Lesner SA. Differences in visual intelligibility across talkers. *Volta Review*. 1982; 84(4): 219–225.
- Kricos PB, Lesner SA. Effect of talker differences on the speechreading of hearing-impaired teenagers. *Volta Review*. 1985; 87(1):5–14.
- Kuhl, PK.; Meltzoff, AN. Speech as an intermodal object of perception. In: Yonas, A., editor. *Perceptual Development in Infancy*. Vol. 20. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1988. p. 235-266. Vol. *The Minnesota Symposia on Child Psychology*
- Lehiste, I. *Suprasegmentals*. Cambridge: MIT Press; 1970.
- Luce, PA.; McLennan, CT. Spoken word recognition: The challenge of variation. In: Pisoni, DB.; Remez, RE., editors. *The handbook of speech perception*. Malden: Blackwell; 2005. p. 591-609.
- Luce PA, Pisoni DB. Recognizing spoken words: the neighborhood activation model. *Ear and Hearing*. 1998; 19(1):1–36. [PubMed: 9504270]
- MacLeod A, Summerfield Q. Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*. 1987; 21(2):131–141. [PubMed: 3594015]
- Marslen-Wilson, WD. Access and integration: Projecting sound onto meaning. In: Marslen-Wilson, WD., editor. *Lexical access and representation*. Cambridge, MA: Bradford Books; 1989. p. 3-24.
- Massaro, DW. *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. Cambridge, MA: MIT Press; 1998.
- Mattys SL, Bernstein LE, Auer ET Jr. Stimulus-based lexical distinctiveness as a general word-recognition mechanism. *Perception & Psychophysics*. 2002; 64(4):667–679. [PubMed: 12132766]
- McClelland JL, Elman JL. The TRACE model of speech perception. *Cognitive Psychology*. 1986; 18(1):1–86. [PubMed: 3753912]
- Montgomery AA, Jackson PL. Physical characteristics of the lips underlying vowel lipreading performance. *Journal of the Acoustical Society of America*. 1983; 73(6):2134–2144. [PubMed: 6875099]
- Morrison CM, Ellis AW. Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental psychology: Learning, Memory, and Cognition*. 1995; 21(1): 116–133.
- Nakatani LH, Dukes KD. Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*. 1977; 62(3):714–719. [PubMed: 903512]
- Norris D. Shortlist: A connectionist model of continuous speech recognition. *Cognition*. 1994; 52:189–234.
- Norris D, McQueen JM. Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*. (In Press).
- Owens E, Blazek B. Visemes observed by hearing-impaired and normal hearing adult viewers. *Journal of Speech and Hearing Research*. 1985; 28:381–393. [PubMed: 4046579]
- Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. *Science*. 1996; 274(5294):1926–1928. [PubMed: 8943209]
- Savin HB. Word-frequency effect and errors in the perception of speech. *Journal of the Acoustical Society of America*. 1963; 35(2):200–206.
- Tye-Murray N, Sommers M, Spehar B. Auditory and visual lexical neighborhoods in audiovisual speech perception. *Trends in Amplification*. 2007; 11(4):233–241. [PubMed: 18003867]