January 2010

Sizing up the competition: Quantifying the influence of the mental lexicon on auditory, visual, and audiovisual spoken word recognition

Julia Feld
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Psychology

Dissertation Examination Committee:
Mitchell Sommers, Chair
Dennis Barbour
Joe Barcroft
Sandra Hale
Mark McDaniel
Joel Myerson
Nancy Tye-Murray

SIZING UP THE COMPETITION:

QUANTIFYING THE INFLUENCE OF THE MENTAL LEXICON ON

AUDITORY, VISUAL, AND AUDIOVISUAL SPOKEN WORD RECOGNITION

by

Julia Elise Feld

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2010

Saint Louis, Missouri

ABSTRACT OF THE DISSERTATION

Sizing up the Competition: Quantifying the Influence of the Mental Lexicon on Auditory, Visual, and Audiovisual Spoken Word Recognition

by

Julia Elise Feld

Doctor of Philosophy in Psychology

Washington University in St. Louis, 2010

Professor Mitchell Sommers, Chairperson

A central question in research on spoken word recognition is whether spoken words are recognized relationally, in the context of other words in the mental lexicon (McClelland & Elman, 1986; Norris, 1994; Luce & Pisoni, 1998). The current research evaluated metrics for measuring the influence of the mental lexicon on spoken word recognition in auditory-only (A-only), visual-only (V-only) and audiovisual (AV) conditions, and assessed the extent to which lexical properties influence recognition similarly across modality of input. Lexical competition (the extent to which perceptually similar words influence recognition of a stimulus word) was quantified using metrics that are well-established in the literature, as well as a novel statistical method for calculating perceptual confusability, based on the Phi-square statistic.

The Phi-square statistic proved an effective measure for assessing lexical competition and explained significant variance in A-only and V-only spoken word identification beyond that accounted for by traditional metrics. Because these values include the influence of all words in the lexicon (rather than only perceptually very similar words), it suggests that even perceptually distant words may receive some activation, and therefore provide competition, during spoken word recognition. Spoken word recognition in A-only, V-only, and AV was sensitive to modality-specific lexical competition and stimulus frequency. These findings extend the scope of activation-competition models of spoken word recognition and suggest that the perceptual and lexical properties underlying spoken word recognition are not unique to the A-only domain.

ACKNOWLEDGEMENTS

It is my pleasure to recognize the many people who have helped me through this process. This work would not have been possible without the generous guidance and encouragement of my advisor, Mitchell Sommers and collaborators Sandra Hale, Joel Myerson, Brent Spehar, and Nancy Tye-Murray. Thanks also to Sharon Chien for collecting and coding data and to Nathan Rose, Chad Rogers, and McKenzie Ballou for a wealth of technical and moral support. I have benefited greatly from the wisdom of Martha Storandt who taught me the importance of measuring continuous variables continuously and talked me out of giving up. I am deeply grateful to my parents Barry and Patricia Feld for nurturing and celebrating my intellectual growth, and to Jon Strand for keeping me so happy while I find my way.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Figures

Equations

Tables

GLOSSARY OF ACRONYMS

Acronym	Definition	Page defined
CVC	Consonant vowel consonant	32
ELP	English Lexicon Project	32
FWNPR	Frequency-weighted neighborhood probability rule . . .	9
FWNWP	Frequency weighted neighborhood word probability. . .	48
NAM	Neighborhood Activation Model	3
NCR	Neighborhood choice rule	6
NWP	Neighborhood word probability	8
SNR	Signal to noise ratio	34
SWP	Stimulus word probability	6

CHAPTER 1: INTRODUCTION & LITERATURE REVIEW

A long-standing question in research on spoken word recognition has been how humans are able to map stimulus information about speech onto meaningful lexical representations in memory. Given the enormity of the mental lexicon [minimum estimates suggest at least 40,000 words in the average adult lexicon (Aitchison, 2003)], discriminating between the appropriate lexical item and all other items in memory is a large and complex task. Remarkably, we manage to complete this task in an almost instantaneous and effortless fashion. A wealth of research has sought to describe the process by which stimulus information from the speech signal activates words in memory, how a specific word is selected from among the activated words, and how the properties of words influence this process.

The speed and accuracy of spoken word recognition seems only attainable if there is an efficient means of searching a highly organized lexicon. There is a growing consensus that the way that words are organized in memory influences our ability to recognize them (see Jusczyk & Luce, 2002, for a review). Although the majority of research on the influence of lexical properties on spoken word recognition has been done within the realm of auditory (A-only) speech, there is growing evidence that lipread, or visually (V-only) perceived speech is also influenced by lexical organization (Auer, 2002; Mattys, Bernstein, & Auer, 2002). Only two studies (Kaiser, Kirk, Lachs & Pisoni, 2003; Tye-Murray, Sommers, & Spehar, 2007a) have investigated how the lexical properties of words identified in an audiovisual (AV) setting influence recognition. Their results were in general accord with the findings in A-only and V-only spoken word recognition, namely that spoken word recognition is sensitive to the lexical properties of the stimulus input.

In the sections that follow, I will review pertinent research on A-only spoken word recognition, including how the structure of the mental lexicon and lexical properties like frequency of occurrence influence recognition accuracy. Then, I will review parallel findings in the V-only and AV spoken word recognition literature. Next, I will discuss the methods by which lexical competition has been quantified in past research and review possible limitations with these methods. Finally, I will discuss a novel method for quantifying lexical competition that may be applied to A-only, V-only, and AV spoken word recognition.

Auditory Spoken Word Recognition

Most current models of auditory spoken word recognition [Neighborhood Activation Model (Luce 1986; Luce & Pisoni, 1998); TRACE (McClelland & Elman, 1986), Shortlist (Norris, 1994)] propose that acoustic-phonetic input from a stimulus word activates a set of perceptually similar lexical candidates in memory, and that these lexical candidates compete for recognition. These *activation-competition* models propose that the amount of activation a lexical item receives depends on the degree of perceptual similarity between the input and the memory representation (Luce & Pisoni, 1998; Marslen-Wilson, 1995). The models further propose that perceptually similar words (called competitors) receive some activation from the stimulus input and provide competition to the stimulus. For example, following the presentation of stimulus word /cat/, the lexical representation for /kit/ receives more activation than the perceptually less similar representation for /bog/. Because each perceptually similar word provides competition for the stimulus word, words with more competitors should be more difficult to recognize than stimulus words with fewer competitors. Activation-competition models have been empirically supported by findings that the number of words that are perceptually similar to the stimulus word influences the speed and accuracy of word recognition: experiments using perceptual identification, lexical decision, and auditory naming tasks have all demonstrated that words with many competitors are recognized more slowly and less accurately than words with few competitors (Goldinger, Luce, & Pisoni, 1989; Luce & Pisoni, 1998; Vitevitch & Luce, 1998).

In addition to the amount of competition a word encounters, the frequency with which a stimulus word occurs in the language also influences spoken word recognition. Word frequency has been well established as a significant predictor of word identification accuracy: high-frequency words are identified more quickly and with greater accuracy than low-frequency words (Savin, 1963; Luce & Pisoni, 1998). Although there is not a clear consensus about the specific mechanism by which word frequency effects operate, most theorists concur that frequency serves to bias or weight activation levels to influence the identification decision (Luce & Pisoni, 1998; Marslen-Wilson, 1995; Morton, 1979).

To account for the roles of both lexical competitors and frequency of occurrence, Luce & Pisoni (1998) proposed a model of spoken word recognition called the Neighborhood Activation Model (NAM). The NAM posits that stimulus input from the speech signal activates a set of acoustic-phonetic patterns in memory. These acoustic-phonetic patterns then activate a set of word decision units that are tuned to particular patterns of input. These word decision units begin monitoring higher-level lexical information (such as word frequency) that is relevant to the words to which they correspond. Therefore, word units serve as an intermediary between acoustic-phonetic (bottom-up) and lexical (top-down) information and are responsible for monitoring both sources of information. In addition, word units are assumed to be interconnected so that they are able to influence activation levels of other word decision units as well as monitor the overall level of activity in the system of units. Word recognition occurs once the word decision unit for a given acoustic phonetic pattern surpasses a criterion relative to activation in the rest of the lexicon. Figure 1 shows a schematic of the architecture and dynamics of word recognition within the NAM.

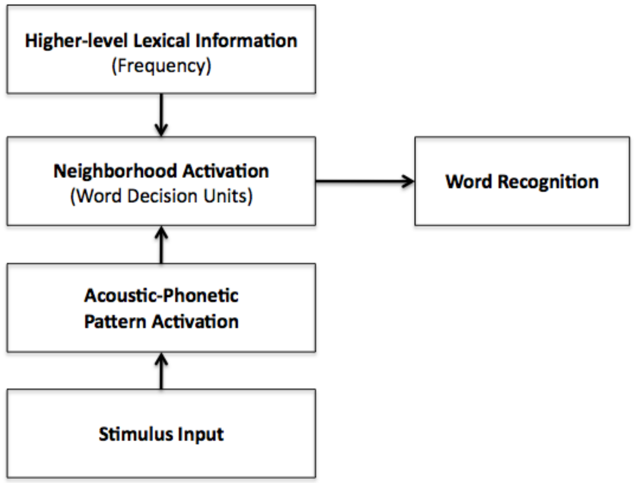


Figure 1. Flow chart of the Neighborhood Activation Model (Luce & Pisoni, 1998)

In order to explicitly quantify the influences of both perceptual and higher-order lexical information on spoken word recognition, the NAM includes a quantitative account of lexical competition that predicts the likelihood of choosing a stimulus word from among its competitors, based on Luce's Choice Rule (1959). Luce's Choice Rule states that the probability of selecting a given item *i* is equal to the probability of item *i* divided by the probability of *i* plus the sum of the probabilities of *j* other items. Luce and Pisoni (1998) applied this to the problem of predicting spoken word identification, assuming that the probability of identifying the stimulus word is equal to the probability of the stimulus word divided by the probability of the stimulus word plus the probabilities of identifying the competitor words. This, referred to as the Neighborhood Choice Rule (NCR), is mathematically expressed as:

$$p(ID) = \frac{p(S)}{p(S) + \sum_{j=1}^{n} p(N_j)} \qquad \textbf{[1]}$$

where p(ID) is the probability of correctly identifying the stimulus word, p(S) is the probability of the stimulus word and $p(N_j)$ is the probability of the *j*th neighbor.

In the numerator of the NCR, Luce & Pisoni (1998) included one perceptual and one lexical factor thought to influence the likelihood of accurately identifying a stimulus word. The perceptual factor is referred to as the Stimulus Word Probability (SWP) and quantifies the likelihood of identifying the acoustic-phonetic pattern of the stimulus word when it is presented. SWP may be thought of as a measure of intelligibility (Auer, 2002), because it measures the probability of recognition, independent of other factors such as lexical competition or frequency. Importantly, SWP is sensitive to only the bottom-up, perceptual support for a word and is independent from the amount of competition a word experiences. SWP is mathematically expressed as:

$$\mathrm{SWP} = \prod_{j=1}^{n} p(PS_i|PS_i) \quad \textbf{[2]}$$

where $(PS_i|PS_i)$ is the conditional probability of accurately identifying phoneme *i* in a forced-choice phoneme identification task. For example, the SWP of the word /bat/is p(b|b) * p(a|a) * p(t|t). The NAM assumes that the decision units of words containing easily identified segments receive more support from the acoustic-phonetic input of that word than words containing segments that are difficult to identify. For example, if the phoneme /tʃ/ is correctly identified 80% of the time and the phoneme /f/ is correctly identified only 40% of the time, the word /chat/ will have a higher SWP than the word /fat/. Thus, SWP reflects the likelihood of correctly identifying a stimulus word's segments, independently of its similarity to other words in the lexicon. In addition to the bottom-up SWP, the numerator of the NCR also contains a top-down, lexical component: stimulus word frequency. The NAM assumes that, initially, the word recognition system is driven completely by the stimulus input. However, once the acoustic-phonetic input has activated a given word decision unit (and many such units will be initially activated), frequency biases the incremental activation for those word decision unit as it accumulates activation.

Because word decision units also monitor the activity of other units in the system, the NAM also requires a metric for assessing the amount of support competitor words receive from the acoustic-phonetic input of a stimulus word, expressed in the denominator of the NCR. To calculate the perceptual similarity of a stimulus word and a competitor word, a measure called Neighborhood Word Probability (NWP) is employed. NWPs predict the likelihood that a stimulus word will be confused with a competitor word by comparing the perceptual similarity of the stimulus word's segments to its competitor's segments. These similarities are approximated using the probability that phonemes will be confused with one another based on results from a forced-choice identification task. To calculate the confusability of two words, the probabilities that a stimulus word's position-specific phonemes will be confused with its competitor's position-specific phonemes are multiplied. The NWP is mathematically expressed as:

$$\text{NWP} = \prod_{i=1}^{n} p(PN_i|PS_i) \qquad \textbf{[3]}$$

where PN_i is the *i*th phoneme of the neighbor and PS_i is the *i*th phoneme of the stimulus word. For example, the probability of responding /mad/ given that the actual stimulus presentation was /mat/ is given by mad|mat = p(m|m) * p(a|a) * p(d|t). This method provides a method for assessing the perceptual similarity of two words on a continuous scale. To incorporate higher-level information about the competitor words, along with the perceptual information from the NWP, each NWP value is weighted by the frequency with which that word occurs in the language.

Thus, to combine information about segmental intelligibility and frequency with competitor confusability and frequency, the NCR is altered to render:

$$p(\text{ID}) = \frac{\prod_{i=1}^{n} p(PS_i|PS_i) * \text{Freq}_s}{\left\{\left[\prod_{i=1}^{n} p(PS_i|PS_i)\right] * \text{Freq}_s\right\} + \sum_{j=1}^{nn}\left\{\left[\prod_{i=1}^{n} p(PN_{ij}|PS_i)\right] * \text{Freq}_{N_j}\right\}} \qquad \textbf{[4]}$$

referred to as the Frequency-Weighted Neighborhood Probability Rule (FWNPR). The Frequency-weighted SWP appears both in the numerator and the denominator. The denominator also contains the sum of the frequency weighted NWPs. The output of the FWNPR predicts performance on word identification, with correlations between FWNPR and A-only word identification accuracy ranging from $r = .23$ to $r = .47$ (Luce, 1986; Luce & Pisoni, 1998).

A coarser (but mathematically simpler) method for quantifying the amount of competition that a word encounters is to define competitors (called *neighbors* by the NAM) [1] as words that may be formed by the addition, deletion, or subtraction of one phoneme of the stimulus. For instance, neighbors of /cat/ include /cot/ (a substitution), /at/ (a deletion) and /cast/ (an addition). This procedure, sometimes referred to as the one-phoneme shortcut method (Dahan & Magnuson, 2006), is appealing because of its computational simplicity, and because it may be used when appropriate confusion matrices are not available or appropriate to calculate similarity continuously. For example, when stimulus materials are presented without background noise (as is often done during lexical decision tasks), using phoneme confusion matrices based on syllable identification tasks in masked noise is not an accurate index of perceptual similarity. Neighborhood density (number of neighbors) has been demonstrated to predict word identification accuracy: words with few neighbors are identified more quickly and accurately than those with many neighbors (Kaiser et al., 2003; Luce, 1986; Vitevitch & Luce, 1998).

[1] Some research uses the term *neighbor* to refer to any perceptually similar word, while some use it only for words obtained using the one-phoneme shortcut method. To avoid confusion, here the term *neighbor* refers to the latter. The term *competitor* is used when comparing a stimulus word to all words in the lexicon.

A common method for simultaneously including the influences of word frequency, competitor similarity, and competitor frequency on spoken word identification has been to group words into lexically *easy* and *hard* clusters (Luce, 1986). Lexically easy words are high in frequency and have few low-frequency neighbors, whereas lexically difficult words are low in frequency and have many high-frequency neighbors. Lexically easy words are identified more quickly and accurately than hard words (Kaiser et al., 2003; Luce, 1986; Vitevitch & Luce, 1998). These results have been demonstrated in many populations, including children (Eisenberg, Martinez, Holowecky, Pogorelsky, 2002), middle-aged (Dirks, Takayana, & Moshfegh, 2001) and older adults (Dirks et al., 2001; Sommers, 1996), pediatric cochlear implant users (Kirk, Pisoni, & Osberger, 1995; Kirk, Hay-McCutcheon, Holt, Gao, Qi, & Gerlain, 2007), and adults with cochlear implants (Kaiser et al., 2003). These findings strongly suggest that the lexical properties of a stimulus word, including its frequency and the perceptual similarity and frequency of its competitors, are important factors in predicting accuracy at A-only spoken word identification.

Visual Spoken Word Recognition

Although spoken word recognition is generally thought of as an auditory phenomenon, speech may also be perceived visually, through observation of the speaker's articulators, including the lips, tongue, teeth, and face. Lipreading has been recognized as a method for perceiving speech by hearing impaired populations for hundreds of years (Johnson, 1775), but more recent behavioral and neurophysiological evidence argues that normal-hearing people incorporate visual information about speech automatically and early in perception (McGurk & McDonald, 1976; see Woodhouse, Hickson, & Dodd, 2009 for a review). Although some compelling evidence exists that similar cortical substrates are employed for the perception of heard and seen speech (Calvert et al., 1997), A-only and V-only speech signals differ in that some physical information about the speech signal is not easily accessible to the lipreader. Certain movements that distinguish between phonemes are obscured within the mouth of a talker, such as whether the vocal folds are vibrating. As a result, phonemes that differ only with respect to voicing, such as /d/ and /t/, are often difficult to distinguish visually. Therefore, there is no one-to-one correspondence between phoneme and lip movement, rendering the visual speech signal more ambiguous than the auditory speech signal (Summerfield, 1992). Unlike the auditory signal, which may be transmitted in the dark or if the speaker is obscured from view, visual speech perception requires more controlled conditions: the speaker must be well lit and located at a distance and angle where the lipreader may easily see his or her movements (Jackson, 1988). Despite the difficulties associated with V-only speech, however, it is possible to attain high levels of accuracy at lipreading. Some lipreading tests show scores for the proportion of words accurately identified up to 89% (Hall, Fussell, & Summerfield, 2005).

There are large individual differences in lipreading ability, with at least one study obtaining a performance range of 7% to 89% accuracy across a relatively homogenous participant sample (Hall et al., 2005). The correlates of these large differences have been the focus of much research. Perhaps surprisingly, numerous studies have found that there is little relationship between lipreading and hearing ability for individuals who have postlingual hearing loss (Clouser, 1977; Farrimond, 1959; Lyxell & Rönnberg, 1989; Lyxell & Rönnberg, 1991; Owens & Blazek, 1985; Rönnberg, 1990; but see Bernstein, Demorest, and Tucker, 2000 for evidence of improved lipreading in congenitally deafened individuals). In some studies (Dancer, Krain, Thompson, Davis, & Glen, 1994; Johnson, Hicks, Goldberg, & Myslobodsky, 1988) females showed better lipreading performance than males, but these effects are usually small and often fail to reach significance (Aloufy, Lapidot, & Myslobodsky, 1996; Irwin, Whalen, & Fowler, 2006; Tye-Murray, Sommers, & Spehar, 2007b). In addition to the studies that have investigated the influence of demographic factors on lipreading abilities, numerous studies have sought to explain the large individual differences with cognitive predictors, often with limited success (see Jeffers & Barley, 1971; Woodhouse et al., 2009, for reviews). Several studies have reported that overall intelligence is a relatively poor predictor of lipreading ability (Elphick, 1996; see Jeffers & Barley, 1971 for a review), as are verbal reasoning abilities (Jeffers & Barley, 1971; Summerfield, 1991), vocabulary (Lyxell & Rönnberg, 1992; Simmons, 1959), and education level (Dancer et al., 1994). Working memory and speed of processing have most consistently been identified as predictors of lipreading performance (Feld & Sommers, 2009; Lidestam, Lyxell, & Andersson 1999; Lyxell & Holmberg, 2000).

Following work in the A-only domain showing that the speed and accuracy of word identification depends on the lexical properties of the stimulus (e.g., stimulus word frequency and neighborhood density), several studies have explored whether the lexical properties of a stimulus word influence recognition accuracy similarly in the V-only domain (Auer, 2002; Mattys et al., 2002; Tye-Murray et al., 2007a). Within the A-only domain, the amount of competition a stimulus word encounters depends in part on its acoustic-phonetic similarity to other words. Therefore, within the V-only domain, the amount of competition should depend upon the *visual* similarity of a target word and its competitor words.

Importantly, there is reason to expect that the extent to which two words are perceptually similar may differ depending on the modality of presentation. Because of the nature of the A-only and V-only signals, phonemic contrasts that are difficult to discriminate in one modality may be easily differentiated in the other (Iverson, Bernstein, & Auer, 1998). For example, acoustic cues to place of articulation for consonants are often difficult to perceive in noisy environments, but the shape of the mouth and articulatory movements that correspond to those consonants are often visually clear (Summerfield, 1992). Therefore, words that are perceptually similar in one modality may not be in the other. For example, aurally, /pin/ (which begins with a voiceless bilabial) is more likely to be confused with /tin/ (which begins with a voiceless alveolar) than with /bin/ (which begins with a voiced bilabial), because the contrast of voicing is not disrupted by noise, whereas the contrast of place of articulation is (Binnie, Montgomery, & Jackson, 1974). Visually, however, /pin/ is more likely to be confused with /bin/ than with /tin/, because the bilabial gesture is visibly apparent, but information about voicing is not available (Mattys et al., 2002). Therefore, to investigate lexical competition in V-only speech, it is necessary to quantify perceptual similarity between words using a method that takes into account the perceptual information available to a lipreader.

The perceptual similarity of phonemes is generally approximated by categorizing phonemes into visually similar groupings called visemes[2] (Fisher, 1968; Owens &

[2] The terms *viseme group* and *Phonemic Equivalence Class (PEC)* are synonymous, as are *homophene group*, *Lexical Equivalence Class (LEC)*, and *visual neighborhood*. Here, the terms *viseme* and *homophene* are used because of their established place in the literature.

Blazek, 1985; Walden, Prosek, Montgomery, Scherr, & Jones, 1977). These viseme groups are intended to represent speech sounds that are visually very similar or indistinguishable. Although viseme groupings depend upon speaker idiosyncrasies, stimulus materials, and participant population, there are commonalities across groupings, and place of articulation is the strongest defining feature (Jackson, 1988). For instance, the bilabial phonemes /b/, /m/, and /p/ generally belong to the same viseme group. There are less consistent visematic groupings for vowels than for consonants because every vowel is produced with a distinct oral cavity shape, so none are truly identical (Jackson, 1988). A common criterion for defining viseme membership is to establish a cut-off point at which perceptually similar phonemes are grouped. Previously, a within-group identification rate of 75% (Walden et al., 1977) has been used. For example, /s/, /t/, and /z/ constitute a viseme group if 75% of presentations of /s/, t/, and /z/ result in a response of /s/, /t/, or /z/. Based on this criterion, consonants and vowels may each be categorized by 5-8 viseme groups (see Jackson, 1988, for a review). Table 1 contains viseme groups determined by Iverson et al. (1998) using the 75% within-group identification rate procedure.

Table 1.

Viseme Groupings, determined by Iverson et al. (1998)

Consonants: {b, m, p} {f, v} {θ, ð} {w} {r} {tʃ, dʒ, ʒ, ʃ, d} {t, s, z} {k, g, h, ŋ, j} {n} {l}
Vowels: {i, ɪ, eɪ, aɪ, ɛ, æ, ʌ} { ɝ, oʊ, ɔɪ, ʊ, u} {a} { aʊ}

From these viseme groups, clusters of visually similar words, called homophenes, may be derived (Mattys et al., 2002; Nitchie, 1930; Tye-Murray et al., 2007a). Words that differ only by phonemes within the same viseme groups are members of the same homophene group. For example, because /b/, /m/, and /p/ are members of the same viseme group, /bat/, /mat/, and /pat/ are homophenes. Although within-viseme group substitutions are possible for any word, not all substitutions result in lexically valid outcomes. For instance, /bog/ will have fewer homophenes than /bat/, because /mog/ and /pog/ are nonwords, and therefore don't serve as competitors.

If homophenes truly represent perceptually indistinguishable units, then using them as a measure of lexical competition is somewhat suspect. If homophenes cannot be distinguished based on the physical properties of the input alone, then density effects could be statistical artifacts. That is, selecting from among a small set of identical options will result in more correct answers simply by chance, than will selecting from among a large set of identical options. If homophenes are truly visually identical, then they are more similar to auditory homophones (/right/ and /write/) than they are to auditory neighbors. If, however, homophenes represent words that are perceptually similar, but not identical, to the stimulus word, they are more similar to auditory neighbors. Support for this second possibility has been demonstrated in that that lipreaders show sensitivity to within-homophene distinctions: Bernstein, Iverson, and Auer (1997) reported that lipreaders are able to distinguish between members of a homophene group (/bite/ and /mite/). This suggests that visemes and homophenes may underestimate the perceptual information available to a lipreader. Therefore, although the 75% within-group identification rate is a convenient method for grouping similar phonemes into visemes, viseme groupings should not be rigidly interpreted as perceptually *identical* sets of phonemes.

In parallel to neighborhood density effects in A-only, the size of a stimulus word's homophene group has been shown to influence V-only word recognition; i.e., stimulus words with few homophenes were identified more accurately than words with many (Auer, 2009; Mattys et al., 2002; Tye-Murray et al. 2007a). Also in accord with A-only findings, high-frequency stimulus words were accurately lipread more often than low-frequency words (Auer, 2009; Mattys et al., 2002). These findings, suggesting that V-only speech is sensitive to the some of the same lexical properties as is A-only speech, led Mattys et al. (2002) to propose the existence of a modality-independent spoken word recognition mechanism that is sensitive to the perceptual properties of the input.

Based on the findings that V-only speech appears to be sensitive to the lexical properties of the stimuli, Auer (2002) applied the NAM to the problem of predicting V-only spoken word identification. Although the NAM was designed to model A-only spoken word recognition, it is readily applicable to V-only perception. The only input required for calculating perceptual similarity in the NAM are confusion matrices displaying the frequency with which pairs of phonemes are (mis)identified as one another in a specific modality. From these matrices, SWPs and NWPs may be easily calculated to quantify the amount of stimulus-based support and competition for a particular stimulus word. This allows competition in V-only to be modeled using a continuous scale, and allows the influence of multiple lexical properties to be evaluated simultaneously.

Using an existing set of V-only phoneme confusions, Auer (2002) calculated V-only NWPs that represent the perceptual similarity (and by extension, amount of competition) of a stimulus word and an individual competitor word. Following the protocol of Luce & Pisoni, all the individual V-only NWPs were weighted by their frequency of occurrence and summed to quantify the total amount of competition exerted by all competitor words on the stimulus word. Auer found that this visually-based lexical density predicted word identification accuracy, in that words with less competition were identified more accurately than words with more competition. Auer also calculated V-only FWNPRs values to include information about the intelligibility of the stimulus word's segments, its frequency of occurrence, its perceptual similarity to other words in the lexicon, and the frequency of those words. Auer (2002) reported a correlation between FWNPR and V-only word identification accuracy of $r = .44$, ($p < .01$). This correlation is of a comparable size to the correlation between A-only accuracy and A-only FWNPR ($r = .23$ to $r = .47$; Luce & Pisoni, 1998). These findings provide evidence that both A-only and V-only spoken word recognition are sensitive to the lexical properties of the input.

Audiovisual Spoken Word Recognition

Although speech may be perceived unimodally, as in A-only and V-only conditions, a majority of speech perception occurs audiovisually, in situations where the perceiver may both see and hear the speaker. Seminal work by Sumby & Pollack (1954) showed that AV word recognition is more resistant to background noise than is recognition in the A-only domain. This was the first evidence that watching a speaker may serve to augment and enhance auditory perception. McGurk & McDonald (1976) argued that when both auditory and visual information about a speaker are available to a listener, integration of the two channels of information is obligatory and unconscious. In this study, participants were simultaneously presented with a visual signal of a speaker saying /ga/ while they heard /ba/. Participants overwhelmingly reported perceiving a /da/, which is a fusion of the features perceived visually and acoustically.[3] This finding has been replicated many times and holds in languages other than English (Sekiyama & Tohkura, 1991), when participants are explicitly instructed to ignore the visual or auditory signal (Massaro, 1987) when the speaker that is heard and the speaker that is seen are different genders (Green, Kuhl, Meltzoff, & Stevens, 1991), and has even been demonstrated in 5-month old infants (Rosenblum, Schmuckler, Johnson, 1997) using a habituation paradigm. Evidence that the combination of auditory and visual information is mandatory and occurs early in processing has led some researchers to argue that multimodal speech is the primary mode of speech perception: ". . .the operations, neurophysiology, information, and evolution of speech perception are based on primitives which are not tied to any single modality" (Rosenblum, 2004, pp. 51).

In natural speech settings, visual information about speech is usually congruent with the auditory information (rather than incongruent, as in the McGurk paradigm). In these cases, when acoustic information about speech is degraded, seeing a speaker as well as hearing them significantly increases intelligibility compared with listening alone (Erber, 1969; Grant, Walden, & Seitz, 1998; Sumby & Pollack, 1954). This improvement in performance is partially due to the fact that visual information about speech (such as

[3] Not all McGurk presentations result in the illusion of perceiving a third, fused phoneme. However, lip kinematics of participants repeating aloud the phoneme they perceived suggests that even when the participant reports only the visual or auditory channel, features of both signals contribute to the response (Gentilucci & Cattaneo, 2005).

identifying place of articulation) can complement auditory information, especially in noisy or reverberant settings (MacLeod & Summerfield, 1987), where information about place of articulation may be lost (Binnie et al., 1974). Thus, visual speech can compensate for reductions in auditory information by providing an alternative modality to obtain phonetic and temporal information about speech. Middleweerd and Plomp (1987) demonstrated that adding visual speech information to an auditory signal was, on average, equivalent to a 4.3 dB improvement in signal-to-noise (S/N) ratio. Based on a 7.4% per dB increase (MacLeod & Summerfield, 1990), this translates to about a 32% improvement in speech perception.

The process by which visual and auditory information about the speech signal are combined is not well understood. At least two groups of researchers (Grant & Seitz, 1998; Massaro & Cohen, 2000) concur that the process of extracting information from the A and V signals (referred to as cue extraction) is distinct from the process of integrating the information derived from A and V signals (referred to as integration). It is clear that there are individual differences in cue extraction ability (evidenced by individual differences in performance on V-only and A-only speech perception tasks), but disagreements persist about whether individuals differ in their ability to efficiently integrate speech information. Grant and collaborators (Grant, 2002; Grant & Seitz, 1998; Grant, Walden & Seitz, 1998) reported that individual participants differ with respect to their ability to efficiently integrate cues from A-only and V-only speech. In contrast, Massaro and Cohen (2000) found that a model assuming optimal or maximally efficient integration accurately predicts human performance, and therefore concluded that there is no evidence that individuals differ in their ability to integrate efficiently. These conflicting findings are difficult to reconcile, given the lack of consensus on how to quantify integration, an absence of correlations between different measures thereof (see Grant & Seitz, 1998 for a review), and observed differences within an individual for measures of integration as a function of unimodal performance [i.e. finding that integration performance differs as a function of the ability to extract unimodal information (Sommers, Spehar, & Tye-Murray, 2005)].

Given the robust findings linking lexical competition and recognition accuracy in A-only and V-only speech, it is somewhat surprising that little work has explored how the lexical properties of stimuli influence AV word recognition. Only two studies (Kaiser, et al., 2003; Tye-Murray, et al., 2007a) have investigated the influence of lexical properties on AV word recognition. Kaiser et al. (2003) presented participants (both normal hearing and cochlear implant users) with lexically *easy* and *hard* words to identify in A-only, V-only, and AV conditions. Lexically *easy* words were high frequency words with few, low frequency neighbors whereas lexically *hard* words were low frequency words with many high frequency neighbors. Importantly, Kaiser et al. computed perceptual similarity using exclusively A-only confusion data rather than calculating perceptual similarity separately for each modality. That is, the number of A-only neighbors (here quantified using the one-phoneme shortcut method) was used to quantify the amount of competition a word encounters in A-only, V-only, and AV domains. Because it has been well established that patterns of phoneme confusion differ significantly in visual and auditory conditions (Iverson et al., 1998; Summerfield, 1992), knowing the number of words that are aurally similar to a stimulus word should not necessarily be expected to predict the number of words to which it is visually similar, and therefore, the amount of competition it undergoes during recognition. Despite this, Kaiser et al. found that *easy* words were identified more accurately than *hard* words in A-only, V-only, and AV modalities. However, because three variables (frequency, neighborhood size, and neighborhood frequency) were used to create the *easy* and *hard* groupings, it is not clear the extent to which the different lexical properties influenced identification rates.

Another method for quantifying lexical competition in AV word recognition was devised by Tye-Murray et al. (2007a). In this study, the investigators proposed the existence of AV neighborhoods that incorporate the influence of both A-only and V-only neighborhoods. Unlike the Kaiser et al. (2003) study, this involved determining neighborhoods separately for A-only and V-only domains based on the phonetic characteristics of each modality. To determine the number of AV neighbors, the number of one-phoneme substitution neighbors (for A-only) and homophenes (in V-only) were determined for each stimulus word. Audiovisual neighbors were defined as the words that were present in the intersection of both modalities (see Figure 2).

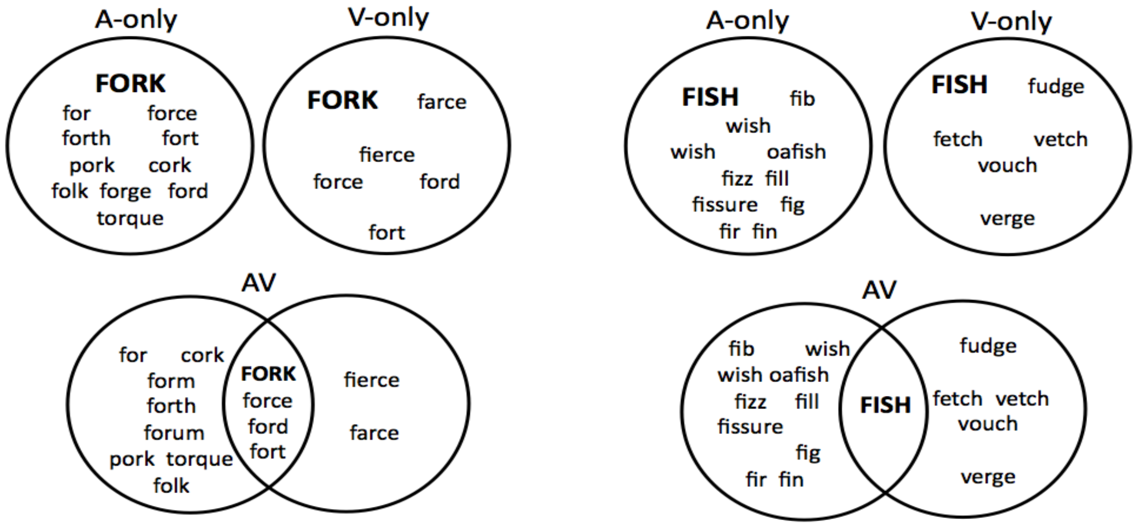


Figure 2. Schematic depicting the interaction of auditory and visual neighbors. The upper half shows the A-only and V-only neighborhoods for the stimulus word /fork/ and /fish/. Both words have a similar number of neighbors in A-only and V-only, but /fork/ has more words in the intersection of A-only and V-only. From Tye-Murray et al. (2007a).

Tye-Murray et al. (2007a) found that stimulus words with fewer words in the overlap of the A and V neighborhoods were identified more accurately than words with many words in the overlap. Due to stimulus constraints, it was not possible to investigate the effects of word frequency and neighborhood frequency on identification accuracy. This study provides initial support for the hypothesis that AV spoken word recognition, like that in A-only and V-only, depends upon the lexical properties of the stimuli.

Quantifying Lexical Competition in Spoken Word Recognition.

In the studies described above, researchers have employed various metrics to approximate the perceptual similarity of phonemes, and by extension, the amount of competition a word encounters. These metrics include the NWP and FWNPR values (Luce & Pisoni) and the one-phoneme short cut method for A-only, the homophene grouping method (Mattys et al., 2002; Tye-Murray et al. 2007a) and NWP method (Auer, 2002) for V-only, and the neighborhood intersection method for AV (Tye-Murray et al. 2007a). These metrics were intended to measure the same construct (form-based lexical similarity), but employed different methods to quantify it. Although each method has shown success at predicting performance in word identification, each has potential limitations.

Categorical Measures

One disadvantage to using viseme and homophene groups to quantify similarity between phonemes and words is that it reduces the range of perceptual similarity between items to binary. For instance, phonemes that share a viseme group are interpreted as perceptually identical, and phonemes across viseme groups are functionally completely distinct. By extension, words are interpreted as either identical (homophenes) or wholly dissimilar (not homophenes). The rigidity of this system has been questioned previously, ". . . although these [viseme] approaches to lipreading have proved informative, they probably remain a coarse approximation of the perceptual experience involved in lipreading" (Mattys et al., 2002, p. 676). Although viseme and homophene groupings certainly capture information about perceptual similarity, it may be that the experience of perceiving speech is better approximated through a more flexible, continuous system.

In the A-only one-phoneme shortcut method, neighbors are considered perceptually similar, but perceptually distinguishable from the target. Like the homophene system, however, the one-phoneme shortcut system also does not allow for differences in the perceptual similarity of neighbors to be assessed. Although some words may be more confusable (perceptually similar to the target) than others, all neighbors are treated as providing the same amount of competition to the stimulus word. For example, two words may have the same number of neighbors, but one may have many neighbors that differ by place of articulation (a feature easily lost in noise or reverberation), while the other may have a majority of neighbors that differ by voicing (a feature that is very resistant to interference). The one-phoneme shortcut method assumes that both words receive similar levels of competition, despite the fact that the phonetic characteristics of the second example word make it perceptually more distinctive.

Also described above is the practice of grouping words into *easy* and *hard* categories, based on the influences of several lexical properties. Although these categories predict spoken word recognition across many populations, there are two disadvantages to this approach. First, as described above, when a continuous measure (lexical difficulty) is treated as categorical, it results in information loss. Second, if only the *easy/hard* grouping is used, it is unclear to what extent each of the lexical properties (frequency, neighborhood density, neighborhood frequency) is driving the effects in identification accuracy.

Continuous Measures

Instead of quantifying perceptual similarity and competition categorically, continuous measures including the application of the NWP calculation method for both A-only (Luce & Pisoni, 1998) and V-only (Auer, 2002) domains have been employed. However, a potential limitation with these methods rests in their use of probability of confusion as an estimate of perceptual similarity. Although the likelihood that two phonemes will be confused seems a reasonable proxy for how perceptually similar they are, this has a limitation: response percentages depend upon the number of perceptually similar alternatives (Iverson et al. 1998). For example, the phonemes /f/ and /v/ look very similar on the face, and will, in general, be confused on roughly 50% of V-only trials. The phonemes / tʃ/, /dʒ/, /ʃ/ , and /ʒ/ are also visually very similar, so any of the four will be confused with another on roughly 25% of trials. In this case, the response percentages give the erroneous impression that /f/ and /v/ are twice as similar as, for instance, / tʃ/ and /dʒ/, despite the fact that, in both cases, they are nearly identical.

To overcome the confound of using probability of confusion as a similarity estimate, Iverson et al. (1998) introduced the Phi-square statistic to the speech perception literature. The Phi-square statistic, a normalized version of the chi-squared test, quantifies the similarity of two response distributions and is mathematically expressed as:

$$p(\mathrm{ID}) = 1 - \sqrt{\frac{\sum \frac{(x_i - E(x_i))^2}{E(x_i)} + \sum \frac{(y_i - E(y_i))^2}{E(y_i)}}{N}} \qquad \textbf{[5]}$$

where x_i and y_i are the frequencies with which phonemes x and y were identified as category i, E_{xi} and E_{yi} are the expected frequencies of response for x_i and y_i if the two phonemes are perceptually identical, and N is the total number of responses to phonemes x_i and y_i. The expected values (E_{xi} and E_{yi}) are determined by summing the frequency with which phoneme x was identified as category i and the frequency with which phoneme y was identified as category i, divided by two. The rationale for this method is that if phonemes x and y are perceptually identical, they should be identified as members of a given category with equal frequency. The Phi-square statistic reaches a value of one when the distributions of responses for two phonemes are identical (participants select each response alternative equally for both phonemes), and reaches a value of zero when the distributions have no overlap (that is, participants did not use any of the same response categories for the two stimulus words).[4] Because the statistic compares the response distributions across all categories, the magnitude of the output is independent of the number of similar alternatives.

[4] In Iverson (1998), Phi-square values were not subtracted from one. The change is made here for two reasons. First, if Phi-square values are not subtracted from one, the value of any phoneme, given itself is 0. This confounds analyses that involve the calculation of conditional probabilities (described below). A second reason for this transformation is ease of interpretation: it makes the scale of Phi-square values the same direction as probability of confusion (higher numbers represent greater similarity)

Figure 3 displays a graphical representation of response distributions to three consonant stimuli. The horizontal axis shows all possible response alternatives, and the vertical axis shows the frequency with which these responses occur to stimuli /b/, /m/, and /s/. For example, when the visual stimulus /b/ is presented, participants most frequently identify it as /b/ or /p/, and rarely identify it as /g/. From this graph, it is clear that the response distributions of /b/ and /m/ are much more similar to one another than they are to /s/. That is, participants show more similar patterns of responding to /b/ and /m/ than to /b/ and /s/ or /m/ and /s/. The similarity between /b/ and /m/ is quantified as $\Phi^2 = .87$, while the similarity of /b/ and /s/ and of /m/ and /s/ are $\Phi^2 = .03$, and $\Phi^2 = .01$, respectively.

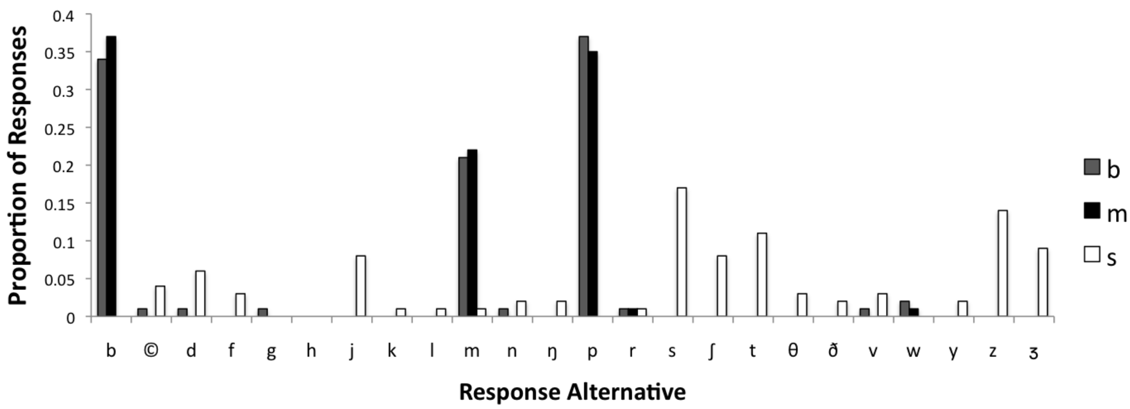


Figure 3. Graphical representation of responses to visually presented /b/, /m/, and /s/.

Another advantage outlined in Iverson et al. (1998) to using the Phi-square statistic instead of probability of confusion values is that it negates the influence of response biases and asymmetries in the data set. For example, a participant in a visual-only phoneme identification task may select response /b/ at a disproportionate rate for a reason that is unrelated to signal information (e.g., their name begins with /b/). In this case, the probability of confusion will result in artificially deflated relationships between /m/ and /p/ (which are visually very similar to /b/). This occurs because when /m/ or /p/ are presented, the response bias of choosing /b/ reduces the frequency with which the other option is chosen. The problem of addressing response biases in phoneme confusion data has been raised before: "Response biases frequently observed in confusion matrices of this type may introduce significant sources of noise in predicting confusions among real words" (Luce & Pisoni, 1998, p. 10). The Phi-square statistic overcomes the problems associated with response biases because it compares overall response distributions without taking into account which response options are selected.

Iverson et al. (1998) used the Phi-square statistic to provide a mathematical basis for categorizing sounds into viseme groups while overcoming the problems associated with the probability of confusion metric. However, because the statistic renders a value quantifying the similarity of every pair of phonemes, it is also possible to use these values directly as a measure of perceptual similarity, rather than using them as a basis for discrete groupings. Using the Phi-square statistic in this manner overcomes both the limitations of a homophene based system and the confounds of using probability of confusion as a proxy for similarity. The Phi-square statistic also provides an elegant solution to the difficulties of comparing V-only homophene groups to A-only neighborhoods described above. Given that the only input necessary to derive Phi-square values is syllable confusion matrices, perceptual similarities may be readily calculated for any pair of phonemes in any modality. This allows the opportunity to directly compare processes of lexical competition in different perceptual modalities.

Overview of Experiments

Word and syllable identification data were collected in A-only, V-only, and AV modalities. The first objective of this research was to evaluate the efficacy of existing metrics at calculating perceptual similarity against a new metric, based on the Phi-square statistic. This will be the first time the Phi-square statistic has been used directly as a measure of perceptual similarity in any modality. It is hypothesized that metrics based on Phi-square will better predict spoken word identification accuracy than those based on probability of confusion.

The second objective is to explore the processes underlying spoken word recognition by examining how lexical properties, including the frequency and intelligibility of the stimulus word and the confusability and frequency of its competitors influence recognition in A-only, V-only, and AV domains. No investigations thus far have compared the influence of lexical properties on spoken word identification in all three modalities. If spoken word recognition depends on similar processes of activation and competition, regardless of the modality of the stimuli, it is expected that the lexical properties of the stimuli would influence recognition similarly in all modalities. The methods by which these questions are tested are described below.

CHAPTER 2: METHODS

Participants

Seventy-two native English speakers with self-reported normal hearing and normal or corrected-to normal vision were recruited from Washington University's undergraduate participant pool. Participants (55 female) ranged in age from 18 to 22 (M = 19.1, SD = 1.07). Testing took approximately 3 hours, which was split into two 1.5 hour sessions. Participants were awarded course credit for their participation, and all procedures were approved by the Human Research Protection Office of Washington University in St. Louis.

Stimuli

The stimuli were recordings of words and syllables produced by six talkers (three male, three female). The stimuli consisted of 24 consonants (b, tʃ, d, f, g, h, dʒ, k, l, m, n, ŋ, p, r, s, ʃ, t, θ, ð, v, w, j, z, ʒ,), 14 vowels (i, ɪ, ɛ, eɪ, æ, a, aʊ, aɪ, ʌ, ɔɪ, oʊ, ʊ, u, ɝ), and 540 Consonant-Vowel-Consonant (CVC) words. To select the stimulus words, a corpus of all CVC words in English was compiled, using the English Lexicon Project (ELP: Balota et al., 2007). This list of 1590 words was pruned to include homophones only once.[5] The list was further trimmed to exclude proper nouns and taboo or profane words.[6] This resulted in 1306 possible stimulus words. From this list, iterations of three sets of 180 words were randomly selected until all three lists met the following criteria:

[5] For homophones, the lexical information of the most frequent member was used.

[6] Although these words were excluded from the list of possible stimulus words, they are included as potential competitors in the neighbor analyses.

lists were matched on Hal_{log} frequency ($p > .92$ for all contrasts), mean lexical decision reaction times on the ELP ($p > .41$ for all contrasts), orthographic length ($p > .35$ for all contrasts), and number of substitution-only phonological neighbors ($p > .29$ for all contrasts). Additionally, lists were checked against one another to ensure that they had equivalent numbers of each part of speech and similar representations of each phoneme in each position. The three lists were randomly assigned to be used for A, V, and AV identification tasks. These analyses were conducted to ensure that the lists were equivalent on critical measures that may influence the speed and accuracy with which they are processed, that they are representative of English CVCs in general, and that they contain a large range of values on all variables of potential interest.

Both syllable and word stimuli were recorded with a Cannon Elura 85 digital video camera connected to a Dell Precision PC and recorded at a 16-bit resolution and sampling rate of 48000. Digital capture and editing was done in Adobe Premiere Elements 1.0. Each talker sat in front of a neutral grey background and spoke the stimuli into the camera as they appeared on a teleprompter. The audio portions of the stimuli were equated for RMS amplitude using Adobe Audition. Auditory and visual information was recorded for all stimuli, but only the visual signal was presented for the V-only tasks, and only the auditory signal was presented for the A-only tasks.

For all A-only and AV identification tasks, background noise (six-talker babble) was set at 60dB SPL. Audio stimuli were presented through a Maico MA42 audiometer over two loudspeakers orientated +/- 45 degrees in front of the participant. Amplitude levels were checked daily to ensure calibration using a handheld sound meter (Quest Technologies Model 2004 Sound Meter). Pilot testing revealed that a consistent signal to noise ratio (SNR) was not appropriate for all tasks: a SNR of -12 resulted in floor-level performance on the A-only word and consonant identification tasks and ceiling-level performance on the AV vowel task. Accordingly, the amplitude at which the signal was presented was manipulated from SNR = -12 until performance was off floor and ceiling for all tasks. Within a modality, consonant and word stimuli were presented at the same SNR, and the vowel task signal was 8dB quieter. The levels and signal to noise ratios are listed in Table 2.

Table 2.

SNR at which stimuli were presented for each task

	Signal (dB SPL)	Babble (dB SPL)	SNR
A-only Words	56	60	-4
A-only Consonants	56	60	-4
A-only Vowels	48	60	-12
AV Words	52	60	-8
AV Consonants	52	60	-8
AV Vowels	44	60	-16

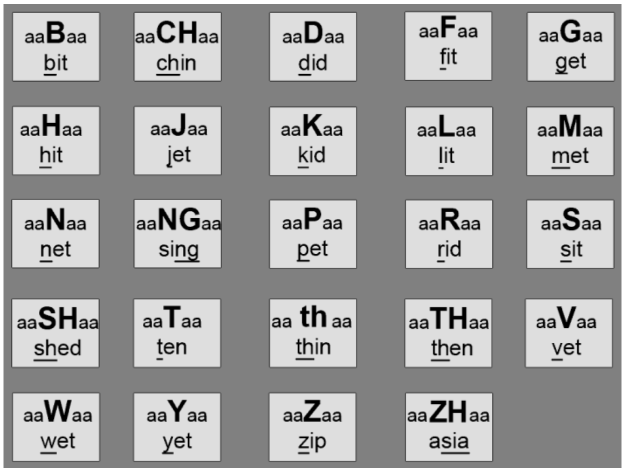
Procedures

Participants read an information sheet, gave verbal consent, and were seated in a sound-proof booth (IAC 120A) approximately 0.5m from a 17-inch Touchsystems monitor (ELO-170C) running Superlab presentation software (Version 4.0.7b, Cedrus Corporation, 2009). Participants were presented with short audio and video clips of consonants, vowels, and single syllable words in A-only, V-only, and AV conditions. They responded to the stimuli via touchscreen button presses or keyboard input. Order of completion of the nine tasks (A-only, V-only, and AV versions of consonant, vowel, and word identification) was randomly determined for each participant. Within each task, speakers were blocked, but across tasks, neither modality nor stimulus type were blocked.

Syllable Identification.

Participants were presented with a series of audio and video clips of a speaker producing a syllable, followed by a response screen listing each phoneme and an example word that contains it. Phonemic contexts (/hVd/ for vowels and /aCa/ for consonants) were selected to minimize co-articulation. Participants made their identification responses by touching the button with the appropriate phoneme (see Figure 4).

A.



B.

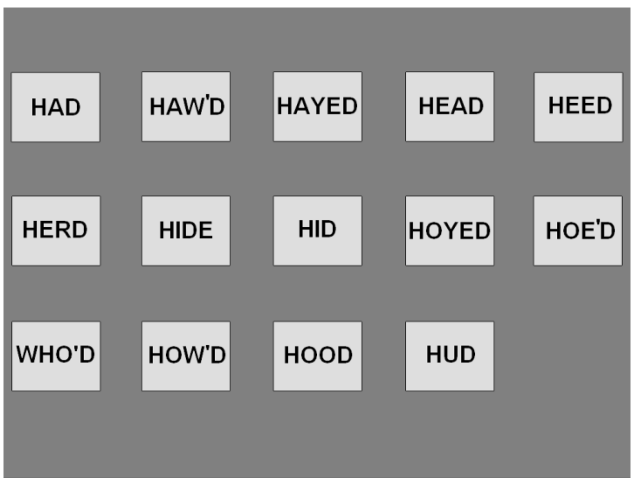


Figure 4. Response screen for A.) Consonant identification and B.) Vowel identification

There were two presentations of each phoneme, spoken by each talker, resulting in 288 consonant trials and 168 vowel trials per participant, per modality. Consonant and vowel tokens were identified in separate blocks and presented in a randomized order, blocked by speaker[7]. Before each participant performed the consonant or syllable task the first time, an experimenter spoke aloud each of the syllable sounds, and participants demonstrated familiarity with them by repeating them aloud in the presence of the experimenter. Participants completed practice trials that consisted of one presentation of each token by a different speaker than was used in the test trials.

Word Identification.

Participants were presented with 180 clips of speakers producing a CVC word, embedded in the carrier phrase "Say the word . . .". They identified the words by typing their responses on a keyboard, and were encouraged to guess when they were unsure. The 180 words consisted of six sets of 30 words, and each set was spoken by a different talker. The word sets were counterbalanced across six participant groups (N = 12 in each), so that each of the words was identified for every speaker. Different sets of words were used for each modality, so although each participant identified 540 words, each modality had a unique list of 180 words (see Appendix A for word lists). Participants completed 6 practice trials in each condition spoken by a different talker than the test trials.

[7] Evidence exists that there may be a processing cost associated with switching talkers in speech identification tasks. Accuracy rates in phoneme identification in conditions in which speakers are blocked are higher than in conditions where speaker order is randomized (Macchi, 1980). To avoid this, speakers were blocked and changes in speaker were preceded by a screen that read "Next Speaker."

CHAPTER 3: LEXICAL VARIABLES

In order to estimate neighborhood structure, a phonetically coded lexicon was obtained from the English Lexicon Project (ELP; Balota et al, 2007). This list, referred to here as the ELP master list, consists of 40,000 English words and contains information about lexical variables such as frequency. Using this list, values for stimulus word frequency, segmental intelligibility, competitor density, and competitor frequency were obtained for all stimulus words.

Stimulus Word Frequency

Two measures of stimulus word frequency were obtained from the English lexicon project. HAL_{Raw} refers to the Hyperspace Analogue to Language (HAL) raw frequency counts reported by Lund & Burgess (1996). These frequency counts are based on 131 million words gathered across 3,000 usenet newsgroups in 1995, and values represent the number of times a given word appeared in the corpus. HAL_{Log} are the log-transformed HAL_{Raw} values. Although it would be preferable to use frequency norms of spoken, rather than written speech, a large database of spoken word frequencies is not available. Despite this limitation, written word frequency has been shown to predict spoken word accuracy both in A-only (Luce & Pisoni, 1998; Savin, 1963) and V-only (Auer, 2009; Mattys et al., 2002).

Stimulus Word Intelligibility

SWP values were derived from the modality-appropriate confusion matrices for A-only, V-only, and AV. Probability SWPs were calculated using the conditional probability of accurately identifying the stimulus word's phonemes, following the procedure described in the introduction (as a reminder, this involves multiplying the probabilities of accurately identifying the phonemes of the stimulus word). To calculate Phi-square SWPs, a slightly different method is necessary because the Phi-square value of any phoneme, given itself, is meaningless. It is, in effect, comparing the response distribution for that phoneme to itself. Therefore, intelligibility values are calculated by averaging all Phi-square values comparing a target phoneme to all other phonemes[8]. For example, the Phi-square SWP of /b/ is the average of (Φ^2 d|b), (Φ^2 f|b), (Φ^2 . . .|b). This serves as a metric for intelligibility because a stimulus phoneme that has many phonemes to which it is similar will tend to have a higher average value than a phoneme that has few other similar phonemes. For example, in V-only presentation, /w/ is not similar to many other phonemes, so has an average Phi-square value of 0.94 whereas /y/ is similar to more phonemes, so has an average Phi-square value of 0.73. In the spirit of calculating a conditional probability, these values are multiplied to reach the Phi-square SWP of a stimulus word.

Competitor density: Established metrics of competition

Auditory-only.

Two measures of phonological neighborhood size were calculated for each stimulus word; Density B and Density A. Density B is the number of words that are only one phoneme addition, subtraction, or deletion away from the stimulus word. For example, for the stimulus word /cat/, Density B neighbors include /scat/, /at/, and /bat/.

[8] Here, the Phi-square values used are not subtracted from 1 as they are for the NWP analyses. This manipulation does not change the outcome of the results, but is done for ease of interpretation. In Probability SWPs, higher numbers indicate greater perceptual support for the stimulus. When Phi-square SWPs are not subtracted from 1, the direction of the relationship is the same; higher values indicate greater intelligibility.

Density A is the number of words that may be formed by a one-phoneme substitution only. For example, Density A neighbors of /cat/ are /bat/, /cot/, and /cap/. Density values were generated manually from the ELP master list using the following process. All stimulus words were phonetically coded and compared to the phonetically coded ELP master list. The number of words requiring only a one-phoneme substitution (Density A) was calculated, as was the number requiring a one-phoneme addition, subtraction, or substitution (Density B). Following the procedure used in Luce & Pisoni (1998), plurals and inflected forms of the stimulus were not included as neighbors.

Although values for Density A and B are available elsewhere (Balota et al, 2007; WUSTL Neighborhood Search), neighborhood size was calculated manually for two reasons. First, calculating neighborhoods from a single lexicon ensures that computations are consistent across A-only, V-only, and AV neighborhood calculations. If available values for A neighborhood size are based on a 20,000 word lexicon, and V neighborhoods are calculated using the 40,000 word lexicon, differences in A and V density effects could be due either to different processes of word recognition in different modalities, or differences in how density is calculated. Second, if there are differences in the predictive power of discrete measures of competition and continuous measures, but the measures were calculated using different lexicons, it is unclear to what extent the differences are due to measurement and not to the size and nature of the words being compared. Therefore, measures of competition for each modality were calculated using the same methods and same lexicon[9].

[9] It is worth noting that using this method produces neighborhood sizes that are very similar to those available elsewhere. The correlation between density measures obtained from ELP and those generated using the above process correlate at $r = .89$ ($p < .001$).

Competitor density was also calculated for each word using the probabilities from the A-only syllable identification confusion matrices following the procedure described in the Introduction. The phoneme confusion matrices were collapsed across speakers, and the input values for the NWPs were the probabilities of confusing any given pair of phonemes in an A-only condition. The sum of all NWPs served as a continuous measure for lexicon competition, and is referred to as A-only Probability ∑NWP (the word *probability* is included to differentiate these values from a parallel analysis using the Phi-square values, described below).

Visual-only.

Viseme groupings for consonants and vowels were determined based on the procedures of Walden et al. (1977) and Iverson (1998). First, V-only phoneme confusion matrices were submitted to a hierarchical cluster analysis. This procedure generated a tree structure that grouped phonemes by confusability. At the lowest level of the structure, each phoneme is in a unique class, and at each successive level, the most similar phoneme pair is joined until, at the highest level, all phonemes belong to a single class. Viseme groupings were defined operationally as the lowest level at which 70%[10] of responses are within viseme class. For example, when presented with /b/, /m/, and /p/, if 70% of responses are either are /b/, /m/, or /p/, they constitute a viseme group. Although viseme groupings have been published previously, (Iverson, 1998; Owens & Blazek, 1985; Walden et al., 1977)), speaker idiosyncrasies can result in differing

[10] Walden and Iverson previously defined a viseme cluster as one within which 75% of responses occur. However, this criterion has proved too rigid in some investigations (Owens & Blazek, 1985; Binnie, Jackson, & Montgomery, 1976), as it did in this one. One consonant cluster (t, d, s, z) showed 72% within viseme responses and one vowel cluster (ɑ, ʌ) showed 70% within viseme responses. Because these two clusters were otherwise unclassifiable, the criterion was relaxed to 70%.

patterns of phoneme confusion (Jackson, 1988). Because of this, a quantitative method for building talker-specific viseme groups has the advantage of more specifically measuring the speech patterns of a given set of talkers. The resulting viseme groups are presented in Table 3. Although viseme groups do vary across studies (presumably due to differences in speakers and materials), these groupings are similar to others in the literature (see Jackson et al., 1988 for other examples).

Table 3.

Viseme groupings derived from V-only phoneme confusions

Consonants:	{b, m, p} {f, r, v} {tʃ, d ʒ, ʃ, ʒ} {g, h, k, l, n, ŋ, y} {d, s, t, z} {θ, ð} {w}
Vowels:	{İ, ɪ} {e, ɛ, æ, Ĩ} {ɝ, ʊ} {ɑ, ʌ} {œ, o} {u} {Ă}

To build homophene groups, all stimulus words and all CVC words from the ELP master list were coded into viseme groups. Next, homophenes for each stimulus word were identified by selecting all CVC words in the ELP master list that had identical viseme strings. The size of the homophene group serves as a categorical measure of neighborhood density.

Competitor density was also calculated with a continuous metric, using the sum of all V-only NWPs. This (referred to as V-only probability $\sum$NWP) was calculated for each stimulus word, using the probabilities from the V-only syllable identification confusion matrices.

Audiovisual.

Following the procedure of Tye-Murray et al. (2007a), the number of words that appear in both the A-only neighborhood (using Density A) and the V-only homophene group for each stimulus word was determined. Because homophene groups include substitution-only neighbors, Density B (which includes words that may be formed from single phoneme additions and subtractions as neighbors) was not used. This "overlap" in A and V neighborhoods represents the number of words that are perceptually similar to the stimulus word in both A-only and V-only modalities. Although, depending on what visemes are understood to represent, this overlap may be characterized as words that are perceptually similar to the stimulus in A-only and perceptually similar or *indistinguishable* in V-only.

Competitor density: Phi-square metrics of competition

To calculate Phi-square measures of competition, a common method was used for A-only, V-only, and AV domains, but using the appropriate modality of phoneme confusions as the input. For each modality, responses for the vowel and consonant identification task were collapsed across speakers and participants, rendering 6 confusion matrices: one each for consonants and vowels in A-only, V-only, and AV. These matrices display the frequency with which each phoneme is identified as every other phoneme in that modality. To calculate Phi-square values for every phoneme pair, the raw frequency confusion matrices for modality-specific vowel and consonant identification were converted to Phi-square values using SPSS (SPSS for Windows, version 18.0), following the procedure described in Iverson (1998). This rendered 6 new matrices that contain Phi-square values rather than frequency of confusion (see Appendix B).

The perceptual similarity of a stimulus word and a competitor word was evaluated following the procedure of calculating NWPs, but using Phi-square values in place of probability of confusion. For instance, the Phi-square NWP of mad|bet = $\Phi^2(m|b)$ * $\Phi^2(a|e)$ * $\Phi^2(d|t)$. For each stimulus word, the NWPs of all other words in the ELP master list were summed to the quantify amount of competition exerted by all other words in the lexicon. To differentiate this metric from the Probability ∑NWPs described above, this value will be referred to as Phi-square ∑NWP. This process was done separately for A-only, V-only, and AV stimulus words. These three values (A-only, V-only, and AV Phi-square ∑NWP) represent the total amount of modality-specific competition a stimulus word encounters. These are theoretically similar to neighborhood density or homophene group size, but include the influence of all words in the lexicon, not just the perceptually similar ones.

Tye-Murray et al. (2007a) suggested that high levels of AV performance on word identification tasks might be due to the interaction between acoustic and visual lexical neighborhoods during ongoing speech recognition. Therefore, an additional value (called AV_{Pre} NWP) was calculated for the AV stimulus words that was designed to capture information about the similarity of the stimulus word and competitor words in A-only and V-only modalities simultaneously. This is, in theory, similar to the concept of using the overlap of A-only and V-only neighborhoods as a predictor of AV word recognition (Tye-Murray et al., 2007a) in that it includes information about stimulus-competitor similarity in A-only and V-only simultaneously.

AV_{Pre} NWPs were calculated by multiplying the A-only and V-only NWPs for each competitor word. Using this procedure, a stimulus word that has high NWP values in both A-only and V-only for a given competitor word will tend to have a high AV_{Pre} NWP value, whereas a competitor that is perceptually similar (has a high NWP) in only one modality will tend to be lower (but is still higher than a competitor word that isn't perceptually similar in A-only or V-only). For example, the stimulus word /pen/ is fairly confusable with competitor /pin/ in both A-only and V-only speech. Therefore, the AV_{Pre} NWP of /pin/ | /pen/ is relatively high. However, /pen/ is similar to competitor /hen/ only in A-only presentation, but is relatively distinct in V-only. Therefore, the AV_{Pre} NWP of /hen/ | /pen/ is somewhat lower. Simply put, competitor words that are perceptually similar to a stimulus word in both A-only and V-only provide more competition in AV than do competitors that are similar in only A-only or V-only. All AV_{Pre} NWPs were summed to reach AV_{Pre} ∑NWP, a predicted AV density based on the unimodal information that is theoretically parallel to the overlap of A-only and V-only neighborhoods (Tye-Murray et al, 2007a).

Competitor frequency

Established metrics of competitor frequency

Frequency-weighted neighborhood word probabilities (FWNWP) were calculated for each stimulus word to quantify both the similarity of a competitor word and the frequency with which it occurs. Again, following the procedure of NWPs, these FWNWPs were summed to reach a frequency-weighted predicted density (called Probability ∑FWNWP). This was done separately for each modality, using the appropriate probability of confusion matrices for that domain. Although ∑FWNWPs are not reported independently in Luce & Pisoni (1998), they appear in the FWNPR.

Phi-square metrics of competitor frequency

∑FWNWPs will also be calculated for A-only, V-only, and AV domains using the Phi-square matrices rather than the probability matrices. These will be referred to as Phi-square ∑FWNWPs.

CHAPTER 4: RESULTS

Prior to analysis, responses to the word identification task were hand-checked for homophones and obvious entry errors. For homophonous stimulus words (e.g., “peace”), all alternate spellings were counted as correct (“piece”). For entry errors, only responses that formed nonwords were corrected, and these nonwords were only corrected in the following circumstances: the response contained a superfluous punctuation mark (e.g., “teeth]”), the response word had a letter pair reversed in a way that did not form a real word (“cheif”), the response word had a doubled letter that did not form a real word (“thiss”), or the response was misspelled in a phonetically probable way (“cowel”). These corrections accounted for approximately 1.5% of responses. No other deviations from the stimulus word (plurals, inflected forms) were counted as correct. Percent accuracy for each stimulus item was calculated and served as the criterion variable for all analyses described below.

Item accuracy ranged between 0-86% correct for A-only words, from 0-77% correct for V-only, and from 0-90% correct for AV words. Because the words were not screened to be highly intelligible in a given modality and were presented at relatively low SNRs for both A and AV modalities, it is not surprising that some words were never accurately identified. Although these words were few in the A-only and AV identification tasks, they were more plentiful in the V-only. To ensure that cross-modality comparisons would not be affected by the greater number of 0% responses in V-only, all analyses reported below are on the remaining words that were identified accurately by at least one participant. To test whether removing the 0% accuracy words influenced the outcome of the results, all analyses were conducted both including these words and excluding them and the pattern of results remained consistent. Including the 0% accuracy words increases the amount of variance accounted for in some analyses, but not the significance level of the analyses. After excluding the words that were never identified accurately, analyses were conducted on 171 words for A-only (range: 1% - 86%, mean accuracy = 0.30, SD = 0.19), 149 words for V-only (range: 1% - 77%, mean accuracy = 0.13, SD = 0.14), and 178 words for AV (range: 1% - 90%, mean accuracy = 0.45, SD = 0.23).

Correlation and regression analyses were conducted to examine the influence of lexical variables on word identification accuracy. These results are described in two parts. First, measures of competitor density were compared to assess their effectiveness at predicting spoken word recognition. These measures include variables that are well established in the speech perception literature as well as metrics based on the Phi-square statistic. Although the traditional metrics and the Phi-square metrics are concerned with the influence of the same underlying properties, the measurement of these properties differs.

Second, the influence of lexical properties including stimulus word frequency, segmental intelligibility, competitor density, and competitor frequency on spoken word identification were assessed. The amount of variance accounted for by each predictor in each modality is compared, and the method for predicting AV density from A-only and V-only phoneme identification is evaluated.

Quantifying Competition

One goal of this research was to compare the efficacy of different measures of lexical competition at predicting accuracy in spoken word identification. No investigations thus far have included the Phi-square statistic as a measure of perceptual similarity, nor has a single investigation employed both a categorical measure of competition (such as homophene group size) and a continuous measure (such as ∑NWP). Therefore, the abilities of these methods to quantify competitor density and predict spoken word recognition have not been compared.

Several aspects of these analyses warrant clarification. For statistical reasons, it may seem obvious that a continuous measure of a continuous variable (perceptual similarity) will be a more accurate representation of that variable than is any categorical measure. Therefore, comparing the predictive power of categorical and continuous measures may seem trivial. However, there are theoretical reasons to argue that the continuous measures may fail to predict variance above and beyond the categorical measures. Because competitors that fall into Density A and B are likely to have higher NWPs than those that do not, it may be the case that $\sum$NWP of all competitor words does not explain any additional variance beyond the categorical measures. That is, knowing the number of perceptually very similar neighbors may be all the information necessary to quantify competition. Although the NAM predicts that words are recognized in the context of the rest of the lexicon, it is possible that the amount of competition provided by the closest neighbors could so overpower the influence of the more distant neighbors as to make them insignificant. For example, for stimulus word /bar/, words like /car/, /bat/, and /par/ are likely highly confusable, and therefore provide significant competition. If stimulus input is processed such that only near-competitors are evaluated in relation to it, then it would not be expected that including the perceptual similarity of a distant neighbor such as /hit/ would add predictive value to the metric. Theoretically, limiting competition to perceptually similar neighbors may be one way of achieving highly efficient lexical access despite a large lexicon. This may be one reason to expect that including the influence of the more distant neighbors may not improve predictive power of density measures. If, however, stimulus input is evaluated in the context of the entirety of the mental lexicon (as the NAM predicts), then knowing the perceptual similarity of distant neighbors such as /hit/ would improve the predictive power of the measure.

Although it might be argued that continuous measures of perceptual similarity will likely account for greater variance in spoken word recognition than categorical ones (or that the measures will be roughly equivalent), direct comparisons of these methods are absent from the literature for any modality. This work will clarify the extent to which computationally simple methods (categorical density) results in a loss of resolution for differentiating between the perceptual densities of words. It is possible that the improvement in the predictive power of using a continuous measure is small enough to justify the computationally simplicity of discrete methods. Descriptive statistics for the measures of competitor density are shown in Table 4.

Table 4.

Descriptive statistics for measures of competitor density

	Range	Mean	SD
A-only Density A	7.00 - 37.00	21.53	6.67
A-only Phi-square ∑NWP	4.39 - 30.42	15.00	5.62
A-only Probability ∑NWP	0.06 - 0.68	0.26	0.11
V-only Homophene Group Size	1.0 - 42.0	15.29	10.67
V-only Phi-square ∑NWP	1.52 - 43.81	17.64	10.33
V-only Probability ∑NWP	0.02 - 0.59	0.26	0.13
AV Overlap	1.0 - 10.0	4.11	2.17
AV Phi-square ∑NWP	0.74 - 6.61	2.64	1.20
AV Probability ∑NWP	0.01 - 0.55	0.20	0.13

A-only Competitor Density

Correlation analyses were conducted to determine the relationship between A-only word identification and Density A[11], A-only Probability ∑NWP, and A-only Phi-square ∑NWP. These correlations are shown in Table 5, below the diagonal. A-only word identification accuracy was significantly correlated with Density A and A-only Phi-square ∑NWP, but the correlation with A-only Probability ∑NWP was small and failed to reach significance. Given previous findings, it is surprising that A-only Probability ∑NWP failed to significantly predicted A-only word identification accuracy. However, the correlations may be influenced by the simultaneous influence of other lexical properties on identification accuracy. For example, words that are in sparse regions of the lexicon but are also very low in frequency will be recognized at lower rates than equivalently sparse words that are high in frequency. Therefore, the partial correlations between measures of competitor density and accuracy, controlling for stimulus word frequency, are also presented. After controlling for the frequency of the stimulus word, the magnitude of all correlations increased, and the correlations between A-only accuracy and all measures of density reached significance. The partial correlations, controlling for frequency are shown above the diagonal in Table 5. All measures of competitor density are negatively correlated with identification accuracy, indicating that words with less similar competitors are identified more accurately. Table 5 also reveals that all measures of competitor density are significantly correlated with one another.

[11] Although Density A is reported here, all analyses also were conducted using Density B, and showed very similar patterns of correlation and variation explained. Density A was used because it is a more similar to measures of categorical V-only density (homophene group size), and is the measure that is applied in calculating AV overlap.

Table 5.

Correlations between A-only accuracy and measures of lexical density. Values below the diagonal are Pearson correlation coefficients. Those above the diagonal are the partial correlations, controlling for stimulus word frequency.

	1.	2.	3.	4.
1. A-only Identification Accuracy	-	-.21*	-.16*	-.33**
2. Density A	-.16*	-	.64**	.31**
3. A-only Probability ∑NWP	-.12	.64**	-	.42**
4. A-only Phi-square ∑NWP	-.28**	.33**	.43**	-

*$p < .05$, **$p < .01$

In order to assess the amount of unique variance in A-only spoken word identification accounted for by each measure, four regressions were conducted. For these analyses, the criterion variable is the probability of stimulus word identification. Table 6A and 6B compare the amount of unique variance accounted for by Density A and A-only Phi-square ∑NWP. Analyses 6C and 6D compare A-only Probability ∑NWP and A-only Phi-square ∑NWP. Although these regressions do not include frequency, the patterns of results are unchanged whether frequency is included or not. The results of these regressions are displayed in Table 6[12].

[12] For all regressions shown, the Beta values are the weights of a particular variable when all variables are included in the regression (ie, at the final step)

Table 6.

Comparing the influence of Density A, A-only Probability ∑NWP, and A-only Phi-square ∑NWP on A-only word identification

Comparing Density A and Phi-square ∑NWP

6A: Density A precedes Phi-square ∑NWP

	β	R^2	ΔR^2
Step 1: Density A	-.08	.03	.03
Step 2: A-only Phi-square ∑NWP	-.25	.08	.05**

6B: Phi-square ∑NWP precedes Density A

	β	R^2	ΔR^2
Step 1: A-only Phi-square ∑NWP	-.25	.08	.08**
Step 2: Density A	-.08	.08	.00

Comparing Probability ∑NWP and Phi-square ∑NWP

6C: Probability ∑NWP precedes Phi-square ∑NWP

	β	R^2	ΔR^2
Step 1: A-only Probability ∑NWP	.01	.02	.02
Step 2: A-only Phi-square ∑NWP	-.27	.08	.06**

6D: Phi-square ∑NWP precedes Probability ∑NWP

	β	R^2	ΔR^2
Step 1: A-only Phi-square ∑NWP	-.27	.08	.08**
Step 2: A-only Probability ∑NWP	-.01	.08	.00

$^{**}p < .01$

Table 6 shows that A-only Phi-square ∑NWP accounts for significant variance above and beyond that accounted for by Density A or A-only Probability ∑NWP. However, the inverse is not true: After accounting for A-only Phi-square ∑NWP, both Density A and A-only Probability ∑NWP fail to explain additional variance in A-only spoken word identification. Taken together, these analyses demonstrate that A-only Phi-square ∑NWP accounts for more variance in spoken word identification accuracy than existing measures and captures unique aspects of the lexical competition process.

V-only Competitor Density

A parallel set of analyses was conducted for V-only word identification accuracy, using homophene group size, V-only Probability ∑NWP, and V-only Phi-square ∑NWP. Table 7 shows the correlations between V-only word identification accuracy and measures of density (below the diagonal), as well as the partial correlations controlling for stimulus word frequency (above the diagonal).

Table 7.

Correlations between V-only accuracy and measures of competitor density. Values below the diagonal are Pearson correlation coefficients. Those above the diagonal are the partial correlations, controlling for stimulus word frequency.

	1.	2.	3.	4.
1. V-only Identification Accuracy	-	-.41**	-.40**	-.57**
2. Homophene Group Size	-.35**	-	.60**	.83**
3. V-only Probability ∑NWP	-.35**	.61**	-	.59**
4. V-only Phi-square ∑NWP	-.48**	.84**	.61**	-

**p < .01

Homophene group size, V-only Probability ∑NWP, and V-only Phi-square ∑NWP were negatively correlated with V-only spoken word identification and were positively correlated with one another. Following the procedure of A-only analyses, a series of regressions was conducted to examine the amount of unique variance accounted for by each measure. These results are presented in Table 8.

Table 8.

Comparing the influence of Homophene group size, V-only Probability ∑NWP, and V-only Phi-square ∑NWP on V-only accuracy

Comparing Homophene group size and Phi-square ∑NWP

8A: Homophene group size precedes Phi-square ∑NWP

	β	R^2	ΔR^2
Step 1: Homophene Group Size	.16	.13	.13**
Step 2: V-only Phi-square ∑NWP	-.61	.24	.11**

8B: Phi-square ∑NWP precedes homophene group size

	β	R^2	ΔR^2
Step 1: V-only Phi-square ∑NWP	-.61	.23	.23**
Step 2: Homophene Group size	.16	.24	.01

Comparing Probability ∑NWP and Phi-square ∑NWP

8C: Probability ∑NWP precedes Phi-square ∑NWP

	β	R^2	ΔR^2
Step 1: V-only Probability ∑NWP	-.09	.12	.12**
Step 2: V-only Phi-square ∑NWP	-.43	.24	.12**

8D: Phi-square ∑NWP precedes Probability ∑NWP

	β	R^2	ΔR^2
Step 1: V-only Phi-square ∑NWP	-.43	.23	.23**
Step 2: V-only Probability ∑NWP	-.09	.24	.01

$^{**}p < .01$

Following the pattern of results in the A-only analyses, these regressions reveal that V-only Phi-square ∑NWP accounts for additional unique variance in V-only word identification after controlling for homophene group size or V-only Probability ∑NWP. Indeed, the variance accounted for by homophene group size and V-only Probability ∑NWP is redundant to that explained by V-only Phi-square ∑NWP.

AV Competitor Density

The relationship between measures of AV lexical density and AV word identification was evaluated next. Correlations between AV word identification accuracy and AV overlap size, AV Probability ∑NWP, AV_{Pre} Probability ∑NWP and AV_{Pre} Phi-square ∑NWP are presented in Table 9.

Table 9.

Correlations between AV accuracy and measures of competitor density. Values below the diagonal are Pearson correlation coefficients. Those above the diagonal are the partial correlations, controlling for stimulus word frequency.

	1.	2.	3.	4.	5.	6.
1. AV Accuracy	-	-.21**	-.17*	-.21**	-.19**	-.26**
2. AV Overlap size	-.13	-	.62**	.71**	.61**	.68**
3. AV Probability ∑NWP	-.12	.63**	-	.73**	.76**	.78**
4. AV_{PRE} Probability ∑NWP	-.15*	.72**	.73**	-	.70**	.75**
5. AV Phi-square ∑NWP	-.14*	.62**	.77**	.71**	-	.89**
6. AV_{PRE} Phi-square ∑NWP	-.19**	.69**	.79**	.76**	.90**	-

*$p < .05$, **$p < .01$

Again, to assess the amount of unique variance accounted for by each measure of competitor density, a series of regressions was done to compare the predictive power of AV overlap size, AV Probability ∑NWP , and AV Phi-square ∑NWP. The results of these regressions are presented in Table 10.

Table 10.

Comparing the influence of A and V neighborhood overlap, AV Probability ∑NWP, and AV Phi-square ∑NWP on AV identification.

Comparing A & V overlap and AV Phi-square ∑NWP

10A: A&V overlap precedes AV Phi-square ∑NWP	β	R^2	ΔR^2
Step 1: A & V Overlap	-.05	.02	.02
Step 2: AV Phi-square ∑NWP	-.12	.02	.02

10B: AV Phi-square ∑NWP precedes A&V overlap	β	R^2	ΔR^2
Step 1: AV Phi-square ∑NWP	-.12	.02	.02*
Step 2: A & V Overlap	-.05	.02	.00

Comparing AV Probability ∑NWP and AV Phi-square ∑NWP

10C: AV Probability ∑NWP precedes AV Phi-square ∑NWP	β	R^2	ΔR^2
Step 1: AV Probability ∑NWP	-.03	.01	.01
Step 2: AV Phi-square ∑NWP	-.11	.02	.01

10D: AV Phi-square ∑NWP precedes AV Probability ∑NWP	β	R^2	ΔR^2
Step 1: AV Phi-square ∑NWP	-.11	.02	.02*
Step 2: AV Probability ∑NWP	-.03	.00	.00

*$p < .05$

Given the more robust correlations between A-only and V-only accuracy and measures of competitor density, the relatively weaker relationship between AV accuracy and measures of AV density is somewhat unexpected. However, there may be a simple mathematical (rather than theoretical) explanation. Accuracy for AV syllable identification matrices was high, with average accuracy levels of 70% for the consonant identification task and 75% for the vowel task. This poses a problem for modeling the perceptual space surrounding phonemes. For example, if a given syllable is accurately identified 90% of the time, the similarity it bears to other phonemes will be represented by only a handful of incorrect responses. These may be too few observations to accurately map the similarity of pairs of phonemes. Because accuracy levels were lower in A-only and V-only syllable identification task, they may be a better representation of general phoneme similarity. When AV_{Pre} ∑NWPs (∑NWPs based on A-only and V-only confusions, see Chapter 3) was used instead, the magnitude of the correlations between lexical density and accuracy rose slightly. The correlations increase further when frequency is controlled (see Table 9). Table 9 also shows that AV_{Pre} ∑NWP measures are highly correlated with obtained AV ∑NWP values, ($r = .90$, $p < .001$) suggesting that the perceptual information in the AV signal may be well-represented by a multiplicative combination of the A-only and V-only signals.

Table 11 presents two regressions to compare the influence of AV_{Pre} Probability ∑NWP and AV_{Pre} Phi-square ∑NWP on predicting AV spoken word identification. Here, AV_{Pre} Phi-square ∑NWP accounts for a small, but significant, amount of variance in AV spoken word recognition beyond that accounted for by AV_{Pre} Probability

$\sum$NWP.[13]

[13] The pattern of results is the same whether frequency is included in the regression or not, but the magnitude of variance accounted for by the measures of density is greater when frequency is controlled.

Table 11.

Comparing the variance in AV word identification accuracy accounted for by Frequency, AV_{Pre} Probability ∑NWP, and AV_{Pre} Phi-square ∑NWP

AV_{Pre} Probability ∑NWP Precedes AV_{Pre} Phi-square ∑NWP			
	β	R^2	ΔR^2
Step 1. Hal_{LOG} Freq	.38	.11	.11**
Step 2. AV_{Pre} Probability ∑NWP	-.04	.15	.04**
Step 3. AV_{Pre} Phi-square ∑NWP	-.22	.17	.02*
AV_{Pre} Phi-square ∑NWP Precedes AV_{Pre} Probability ∑NWP			
	β	R^2	ΔR^2
Step 1. Hal_{LOG} Freq	.38	.11	.11**
Step 2. AV_{Pre} Phi-square ∑NWP	-.22	.17	.06**
Step 3. AV_{Pre} Probability ∑NWP	-.04	.17	.00

*$p < .05$, ** $p < .01$

Only one previous investigation (Tye-Murray et al, 2007a) has considered the simultaneous influence of A-only and V-only perceptual similarity on spoken word recognition in AV. This study made (and found support for) three hypotheses, based on the principles of the NAM and assuming that visual and acoustic information influence AV word recognition. The first of these hypotheses was that stimulus words with few items in the AV overlap would have a greater likelihood of being recognized. The current project found support for this hypothesis. AV_{Pre} Phi-square ∑NWP (a theoretical counterpart to AV overlap) accounted for a significant (albeit small) amount variance in AV word identification accuracy, as did AV overlap size when stimulus word frequency was controlled. The second hypothesis of Tye-Murray et al. (2007a) was that unimodal lexical density would predict AV word identification accuracy. Support was also found for this hypothesis. Here, A-only Phi-square ∑NWP and V-only Phi-square ∑NWP both account for a small but significant amount of unique variance in AV word identification accuracy (6% for A-only, 3% for V-only, $p < .01$ for both).

The third hypothesis was that A-only density should be predictive of A-only accuracy but not of V-only accuracy, and that V-only density should be predictive of V-only accuracy, but not A-only accuracy. An important assumption of the NAM is that competitor effects are the result of the perceptual similarity of the target to its competitors. Therefore, because density depends on perceptually defined similarity within a given modality, confusions from one modality should not be expected to correlate with accuracy in another modality (see also Auer, 2002).[14] That is, the amount

[14] Importantly, this rests on the assumption that perceptual similarity in the two modalities is not necessarily correlated. That is, words that have many similar competitors in A-only should not be expected to have many similar competitors in V-only. Indeed, A-only and V-only Phi-square ∑NWPs are not significantly correlated ($r = .08$, $p < .05$)

of competition a word encounters in the visual modality should not predict A-only recognition accuracy. Phi-square ∑NWP values reveal exactly this pattern: A-only Phi-square ∑NWP values of words do not predict identification accuracy in V-only (r = -.08, p = .33), nor do V-only ∑NWP values predict A-only identification accuracy (r = .11, p = .16).[15] Simply put, the perceptually density of a neighborhood in one modality does not predict identification accuracy in the other modality, supporting the NAM's prediction that the density effects depend on perceptually-derived similarity of the competitors.

Extent of competition

As discussed above, the NAM predicts that words are evaluated in the context of all other words in the lexicon. It would, therefore, predict that quantifying the perceptual similarity of a stimulus word to every other word in the lexicon would offer greater predictive power than quantifying its similarity to only the most perceptually similar words. In order to test this prediction, the Phi-square NWPs of the categorically defined neighbors (those in Density A for A-only, homophene group for V-only, and AV overlap for AV) were calculated. This accounts for different levels of perceptual similarity within a categorical cluster. If word recognition depends upon each stimulus word being evaluated in the context of every word in the lexicon, it would be expected that these measures (which include only relatively close competitors) would account for less variance in word identification accuracy than do Phi-square ∑NWPs, which include all

[15] Probability ∑NWP values show roughly the same pattern, though are somewhat harder to interpret because the correlation between A-only Probability ∑NWP and A-only identification accuracy failed to reach significance. However, like the Phi-square ∑NWPs, A-only Probability ∑NWP do not predict V-only Accuracy (r = .10, p = .13), nor do V-only Probability ∑NWP values predict A-only accuracy (r = -.12, p = .11)

words in the lexicon. Correlations between spoken word identification and the above measures of density are shown in Table 12.

Table 12.

Correlations between word identification accuracy and measures of density, controlling for frequency

	A-only	V-only	AV
Categorical Density	-.16*	-.41**	-.21*
Phi-square ∑NWP of categorical density	-.18*	-.47**	-.16*
Phi-square ∑NWP	-.33**	-.57**	-.26**

* p <.05, ** p < .01

Note: All measures of density are calculated separately for A-only, V-only, and AV. Categorical density is Density A for A-only, homophene group size for V-only, and A-only and V-only overlap size for AV. Phi-square ∑NWP of categorical density is the sum of the NWP values within each measure of categorical density. For AV, AV_{PRE} values are shown.

For all modalities, Phi-square ∑NWP accounts for a significant amount of variance beyond that accounted for by the Phi-square ∑NWP of categorical density. This suggests that, even when the similarity of categorical members is measured continuously, comparing a stimulus word to all words in the lexicon accounts for greater variance in word identification than does comparing it to its closest competitors.

Summary

The above analyses demonstrate that the amount of modality-specific competition a stimulus word encounters influences the likelihood that it will be recognized. These analyses also support the value of using Phi-square values as a measure of lexical competition, and support the theoretical construct of modality-specific auditory and visual competition. Metrics based on the Phi-square statistic are as effective (in the case of AV) or more effective (as in A-only, V-only, and AV_{Pre}) than similarity estimates derived from identification probability values. In addition, comparing a stimulus word to all other words in the lexicon accounts for greater variance in identification accuracy than does comparing it to only a subset of perceptually similar competitors.

Activation and competition across modalities

In addition to evaluating how well different metrics are able to model lexical density, this project is also concerned with assessing whether metrics of lexical properties have similar predictive abilities across modalities. Four lexical properties are examined first independently and then in concert: stimulus word frequency, segmental intelligibility, neighborhood density, and neighborhood frequency. Table 13 contains descriptive statistics for stimulus word frequency and intelligibility.

Table 13.

Descriptive statistics for stimulus frequency and stimulus word intelligibility (SWP)

	Range	Mean	SD
HAL_{Log} Frequency	2.56 - 15.48	8.63	2.19
HAL_{Raw} Frequency	13 - 5,262,331	85,792.05	424,491.52
A-only Probability SWP	0.01 - 0.44	0.17	0.10
A-only Phi-square SWP	0.33 - 0.58	0.46	0.05
V-only Probability SWP	0.01 - 0.18	0.05	0.04
V-only Phi-square SWP	0.39 - 0.68	0.51	0.06
AV Probability SWP	0.13 - 0.75	0.38	0.14
AV Phi-square SWP	0.64 - 0.78	0.70	0.03

Note: Because A-only, V-only, and AV lists were matched on frequency, the frequency values for all lists combined are displayed.

Frequency

Consistent with earlier findings, the frequency of the stimulus word predicted a significant amount of variance in spoken word identification accuracy both in A-only (Landauer & Streeter, 1973; Luce & Pisoni, 1998) and in V-only (Auer, 2009; Mattys et al., 2002). HAL_{Log} was a significant predictor of word identification accuracy in all domains; it accounted for 7% of the variance in word identification accuracy in A-only word recognition, 5% in V, and 11% in AV ($p < .001$ for all).

Interestingly, HAL_{Raw} values predicted stimulus word identification accuracy less effectively than did HAL_{Log} values. HAL_{Raw} accounted for no variance in A-only, 4% of the variance in V-only ($p < .01$), and 2% (ns) in AV. Although these values are still significant for V-only, they are lower in magnitude than the correlations with HAL_{Log} and the correlation disappears completely for A-only and AV. The difference between HAL_{Log} and HAL_{Raw} may explained by the distribution of frequency values. HAL_{Raw} values have a large range but a very skewed distribution, with 90% of responses falling below 100,000, and the tail stretching to 5 million. Because the NAM assumes that word decision units are initially activated exclusively by the perceptual input and only later weighted by frequency, it is reasonable to expect that words with high competitor density may be difficult to recognize, even if they are high frequency. Therefore, when a frequency distribution is extremely skewed and the high frequency outliers are able to exert strong influence on the correlation, high frequency, high density words may deflate the relationship between accuracy and frequency. HAL_{Log} values avoid this issue because they have a smaller, more normally distributed range.

Stimulus Word Intelligibility

Descriptive statistics for SWP values, derived from modality-appropriate confusion matrices are shown in Table 13. Table 14 shows the correlations between these SWP values and identification accuracy for each modality.

Table 14.

Correlations between measures of intelligibility and word identification accuracy

	A-only	V-only	AV
Probability SWP	.17**	.45**	.13**
			.19** (AV_{Pre})
Phi-square SWP	.19**	.52**	.05
			.32** (AV_{Pre})

** *p* < .01

A series of regressions revealed that neither probability SWPs nor Phi-square SWPs account for significantly greater unique variance than the other. Indeed, Phi-square SWPs and Probability SWPs are highly correlated within each modality (r = .85 for A-only, r = .61 for V-only, and r = .84 for AV, all p < .001). This suggests that using the average Phi-square value for a given phoneme captures similar information about intelligibility as does using the probability that a phoneme will be accurately identified as itself.

Despite mulitcollinearity between Probability SWP and Phi-square SWP, the correlation between AV accuracy and Phi-square SWP was small and failed to reach significance. Given the correlations between SWP and accuracy in A-only and V-only, this finding is somewhat surprising. This may, again, be due to very high accuracy levels in AV phoneme confusion matrices. To attempt to overcome this, an AV_{Pre} SWP was calculated by multiplying the A-only SWP and V-only SWP of a word, both for Probability SWPs and Phi-square SWPs. Using this method, words that are perceptually intelligible in both A-only and V-only will have higher values than words that are perceptually intelligible in only one (or neither) domain. These values showed stronger correlations with AV word recognition accuracy (see Table 14).

Frequency Weighted Competitor Density

In the NAM, frequency effects are assumed to both weight the word decision unit tuned to the stimulus word, and also weight the activation of decision units tuned to competitor words. To examine how the frequency of the competitor words (as well as the perceptual similarity thereof) influence recognition, FWNWPs were calculated. For each stimulus word, the NWP of every competitor was weighted by its frequency of occurrence. Then, all frequency-weighted NWPs were summed. These ∑FWNWPs were calculated for all stimulus words, using both Phi-square values and probability values. The fundamental assumption of this analysis is that high frequency competitors should provide more competition for a given stimulus word than low frequency competitors, controlling for perceptual similarity. If a given competitor word is both high frequency and perceptually very similar to a stimulus word, the FWNWP will be inflated to reflect a high level of competition. Therefore, words whose close competitors tend to be high frequency will have higher overall ∑FWNWP values than words whose close competitors are infrequent.

The initial frequency weighting analyses employed HAL_{Log} values, as were used in the stimulus word frequency analyses. However, these values caused an unexpected statistical confound. When ∑FWNWPs were calculated using the HAL_{Log} values, it resulted in a correlation of $r = 0.999$ between the ∑FWNWP and the unweighted ∑NWP of a stimulus word. This indicates that weighting each of the competitor word by its frequency did not change the rank order of the unweighted ∑NWP. An examination of the frequency distributions of competitor words reveals the cause of this confound (see Figure 5).

A **B**

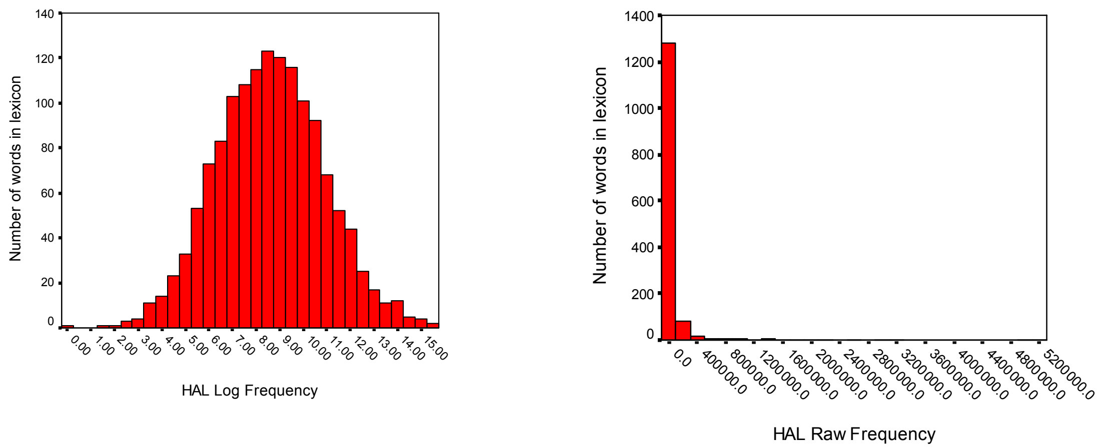


Figure 5. Distribution of frequency values of all competitor words in ELP Master Lexicon. The histogram on the left shows HAL_{Log} values, and on the right, HAL_{Raw} values. Some, but very few, HAL_{Raw} values extend to as high as 5,000,000.

The HAL_{Log} values of the competitor words follow the pattern of normal distribution, whereas the HAL_{Raw} values are very skewed with most values being less than 100,000 and only a handful stretching up to 5,000,000 or more. Because the HAL_{Log} are normally distributed, most competitors are clustered around the mean, with few in the tails. Even if the most similar neighbor is a high frequency one, the difference between the highest frequency value and lowest frequency value isn't large enough to significantly influence the distribution. Indeed, the result of weighting all ∑NWPs by the appropriate frequency was, in effect, to simply multiply the unweighted ∑NWPs by the average frequency of the lexicon, yielding an almost perfect correlation.

To overcome this confound, the HAL_{Raw} frequencies of competitor words were used in the weighting. HAL_{Raw} frequency values have a much larger range, so the large values of high frequency words are able to exert enough influence on NWP values to alter the ∑FWNWP. Using HAL_{Raw} values makes it possible to distinguish between stimulus words that are, on average, more perceptually similar to high frequency words from those that are, on average, more similar to low frequency words. Correlations between ∑FWNWP and word identification accuracy are shown in Table 15.

Table 15.

Correlations between word identification accuracy and Probability and Phi-square FWNWPs in A-only, V-only and AV domains.

	A-only	V-only	AV
Probability FWNWPs	-.12	-.15	-.16*
Phi-square FWNWPs	-.24**	-.37**	-.09

* $p < .05$, ** $p < .01$

After controlling for Probability ∑FWNWPs, Phi-square ∑FWNWPs accounted for additional unique variance in identification accuracy in A-only and V-only tasks. In the AV domain, the correlation between Phi-square ∑FWNWP and word identification accuracy failed to reach significance.

Unexpectedly, for all modalities, ∑FWNWPs accounted for no additional variance above that accounted for by the unweighted ∑NWPs. In fact, for A-only and V-only, the unweighted ∑NWPs accounted for significantly more unique variance in predicting accuracy than did their frequency-weighted counterparts. This finding is surprising given the importance placed on both the similarity and the frequency of the competitor words by the NAM. However, neither Luce & Pisoni (1998) nor Auer (2002) reported ∑NWPs and ∑FWNWPs separately, so it is unclear whether this finding is novel. It is possible that the lack of competitor frequency effects is due to the frequency measures used. HAL frequency counts (Lund & Burgess, 1996) were obtained from written, rather than spoken speech. The argument against competitor frequency effects would be strengthened by a similar demonstration, but using improved measures of word frequency (both more recent and derived from spoken, rather than written speech).

Frequency-Weighted Neighborhood Probability Rule

To simultaneously include the influence of multiple lexical properties on identification accuracy, FWNPRs were calculated for each modality. This followed the procedure described in the introduction. Because the ∑FWNWP showed smaller correlations with accuracy than did the ∑NWPs, the unweighted ∑NWPs were used in the denominator. Both Probability values and Phi-square values were used, and the correlations of each with accuracy are shown in Table 16.

Table 16.

Correlations between FWNPR and identification accuracy in A-only, V-only and AV

	A	V	AV
Probability FWNPR	.32**	.55**	.23**
			.31** (AV_{Pre})
Phi-square FWNPR	.34**	.66**	.30**
			.39** (AV_{Pre})

** $p < .01$

In all three domains, Phi-square FWNPRs accounted for significant additional variance beyond that explained by probability FWNPRs. In the A-only condition, the Phi-square FWNPR accounted for a small but significant additional 2% ($p < .05$) of the variance in identification accuracy. For V-only, it was an additional 14% percent ($p < .001$) and for AV, 4% ($p < .01$). Using AV_{Pre} values improved the predictive power of both AV Probability FWNPR and AV Phi-square FWNPRs. AV_{Pre} Phi-square FWNPR accounted for an additional 7% ($p < .01$) of variance in AV word recognition after accounting for AV_{Pre} Probability FWNPR.

The correlation between A-only Probability FWNPR and accuracy is very similar to what has been found previously and for A-only [$r = .23$ to $r = .47$ (Luce & Pisoni, 1998)]. The correlation between V-only FWNPR and accuracy is stronger than has been reported previously [$r = .44$ to $r = .48$ (Auer, 2002)]. This may be the result of a methodological difference between the two investigations. Auer (2002) used an existing set of phoneme confusions that were obtained from different talkers than the word identification task, and were identified by a different population group (hearing-impaired vs. normal-hearing). Because phoneme confusions differ as a function of speaker idiosyncrasies, it may be Auer (2002)'s FWNPR values are lower than they would be given speaker consistency across tasks.

Individual contributions of lexical properties across modalities

Although the FWNPR assesses the simultaneous influence of multiple lexical properties, further analyses are required to assess the independent influence of each property. To determine the amount of variance accounted for by each lexical property (and how patterns of variance explained differ across modalities), a series of regressions was conducted. Stepwise regressions revealed that frequency and Phi-square ∑NWP were the strongest predictors of identification accuracy. For V-only (but not A-only or AV), Phi-square SWP accounted for significant variance beyond frequency and Phi-square ∑NWP. A hierarchical regression, with order specified based on the stepwise analysis is presented in Table 17. This does not imply that Phi-square SWP does not predict word identification accuracy in A-only and AV, however. Indeed, forcing Phi-square SWP in the second step (after frequency) accounts for significant variance in identification accuracy, and entering Phi-square ∑NWP as the third step accounts for additional variance beyond that in both A-only and AV modalities.

Table 17.

Comparing the influence of lexical variables on spoken word identification in A-only, V-only, and AV domains

	A-only			V-only			AV		
	β	R^2	ΔR^2	β	R^2	ΔR^2	β	R^2	ΔR^2
Step 1: HAL_{log} Freq	.30	.07	.07**	.37	.05	.05**	.36	.11	.11**
Step 2: Phi-square ∑NWP	-.33	.16	.11**	-.25	.36	.31**	-.22	.15	.04**
Step 3: Phi-square SWP	-.03	.16	.00	.19	.40	.04**	-.05	.15	.00

**p < .01

Summary

These results reveal that spoken word recognition is sensitive to similar lexical properties whether the input is auditory, visual, or both, although some differences between modalities do exist. In all three domains, higher frequency stimulus words and those with less similar competitors tend to be identified more accurately than low frequency words, or those with highly similar competitors. For V-only (but not A-only or AV), stimulus intelligibility predicted significant additional variance in word identification accuracy, after controlling for stimulus word frequency and competitor density. The implications of these findings are discussed below.

CHAPTER 5: DISCUSSION

The current research was designed to evaluate methods of quantifying perceptual similarity and lexical competition in A-only, V-only, and AV domains, and to assess the extent to which lexical properties influence spoken word recognition similarly across input modality. The findings, along with clinical applications and possible limitations of the work, are discussed below.

Quantifying Competition

It was expected that continuous measures of lexical competition would be more successful at predicting spoken word recognition than would those that measure perceptual density categorically. This was seen as likely for several reasons. The NAM (Luce & Pisoni, 1998) proposes that activation of competitor words occurs as a function of their perceptual similarity to the stimulus word. Competitors that are very perceptually similar to the stimulus word receive more activation than those that are less similar. Metrics that calculate perceptual density categorically (like one-phoneme shortcut neighbors or homophenes) do not distinguish between the perceptually similarity of competitors. That is, competitor status is treated as a categorical, rather than a continuous variable. Treating a continuous variable (like perceptual similarity) as categorical results in information loss, so a continuous measurement is likely to be more predictive. The continuous/categorical comparison was also reported here due to the prevalence of categorical measures in the literature. Although there are statistical reasons to expect that continuous measures would have greater predictive power than categorical measures, a direct comparison is, thus far, absent from the literature.

The main measurement question posed here was whether the measures of lexical density that are based on the Phi-square statistic would have greater predictive power for spoken word identification than would measures of lexical density based on confusion probabilities. Although this is the first investigation to use the Phi-square statistic as a continuous measure of perceptual similarity, it might be expected to be a more predictive measure than a parallel analysis that uses confusion probabilities, based on several confounds associated with confusion probability methods (described in the Introduction; see Iverson et al., 1998). Overcoming these confounds may result in a more valid measure of perceptual similarity, and therefore, lexical density.

The results of the current investigation support these predictions and demonstrate that quantifying lexical competition using metrics based on the Phi-square statistic are as effective or more effective at predicting spoken word identification than are metrics that have been used previously. Indeed, Phi-square density accounted for additional variance in spoken word recognition beyond that accounted for by discrete metrics or those based on confusion probability values in A-only and V-only spoken word recognition. In AV, the amount of variance accounted for by all measures of lexical density was relatively small. When a predicted AV density was calculated (using the conditional probabilities of unimodal confusion), the measures based on Phi-square values accounted for additional variance beyond confusion probability measures.

Activation and competition across modalities

Because the Phi-square statistic enables lexical competition to be assessed on the same scale in A-only, V-only, and AV speech, the influence of lexical properties on recognition in each modality may be directly compared. It was expected that A-only, V-only, and AV spoken word recognition would be similarly influenced by lexical properties. This prediction stems from some similar findings in the three domains with respect to the influence of competitor density (Mattys et al., 2002; Auer et al., 2002; Tye-Murray et al., 2007) and frequency (Auer et al., 2002; Mattys et al., 2002) on spoken word identification. However, this marked the first investigation in which lexical properties were measured using the same method for A-only, V-only, and AV speech.

Frequency of the target word accounted for a significant proportion of the variance in A-only, V-only, and AV spoken word recognition, and the amount of variance accounted for was relatively similar across modalities. Competitor density also accounted for a significant amount of unique variance in all modalities, beyond that accounted for by target word frequency. For V-only, but not the other modalities, segmental intelligibility accounted for further additional variance in word identification accuracy.

Although the patterns of variance in spoken word recognition accounted for by the lexical properties are reasonably consistent across modalities, several differences warrant further discussion. First, the total predictive power of lexical properties at accounting for variance in spoken word identification differs by modality. In total, lexical properties account for 16% of the variance in A-only, 40% in V-only, and 15% in AV. It is possible that the stronger correlations observed between V-only word identification and measures of perceptual similarity are stronger than the A-only and AV counterparts because in the V-only condition, both vowel and consonant tasks were presented in identical conditions (see *Possible limitations in quantifying competition* section for more details).

Another way in which patterns of explained variance differ across modalities is that V-only word recognition is more influenced by the segmental intelligibility of the stimulus (SWP), after controlling for competitor density. Given that there is not a one-to-one correspondence between phoneme and mouth gesture in V-only speech, it may be that intelligibility plays a larger role in identifying words than in A-only, where each phoneme, at least in an ideal listening situation, is unique. These findings are counter to those reported by Auer (2002). In that study, segmental intelligibility didn't account for additional variance in spoken word identification after accounting for competitor density. Auer found this finding surprising and attributed it to the fact that the phoneme identification scores were obtained from a different population of participants (hearing-impaired vs. normal-hearing) identifying sounds from different speakers than were used in the word identification task. Therefore, he suggested that the phoneme confusions used might not accurately capture the segmental intelligibility of the words. He suggested that other ongoing studies in their lab that avoided these confounds (as the current investigation did) did find an effect of segmental intelligibility after controlling for density.

Theoretical Implications and Future Directions

The NAM was originally designed specifically to describe A-only spoken word recognition. However, the finding that V-only and AV spoken word recognition are also sensitive to lexical properties suggests that the scope of the NAM may be extended to include other modalities of speech perception. These results suggest that similar processes of activation and competition occur for spoken word recognition, whether the signal is seen, heard, or seen and heard. It is especially interesting that measures of density within a unimodal condition (e.g., A-only Phi-square density) do not predict spoken word recognition in another unimodal condition (e.g., V-only spoken word identification accuracy). This suggests that lexical density is not an inherent property of a word, but rather, depends upon the nature of the perceptual signal through which it is perceived. It should, therefore, not be surprising that AV word recognition is sensitive to the properties of both the A-only and V-only channels (Tye-Murray et al., 2007a).

Disagreements persist about whether individuals differ in their ability to efficiently integrate auditory and visual information in AV spoken word recognition (see Grant, 2002 and Massaro & Cohen, 2000). Although this research was not designed to resolve this debate, it supports that AV recognition does depend, in part, upon the perceptual and lexical properties of the stimuli. Some methods of assessing integration performance include calculating predicted AV accuracy from A-only and V-only error rates, such that errors are expected in AV only when errors occur in both A-only and V-only identification (Blamey, Cowan, Alcantara, Whitford, Clark, 1989). For example, if a participant misidentifies a word in both V-only and A-only, that word is expected to be misidentified in AV. However, these values are binary (stimuli are either identified or misidentified), so they fail to describe what perceptual information the perceiver extracts. If place of articulation for a critical phoneme is accurately identified in V-only (eg., /bad/ is identified as /pad/) and voicing is accurately identified in A-only (eg., /bad/ is identified as /dad/), then the probability of identification in AV seems much higher than if the information extracted is redundant in the two modalities. It may be that participants differ in the features they are able to extract in a given modality, suggesting that general measures of perceptual competition (eg., A-only neighborhood size) may not accurately describe the perceptual experience of a given individual. Therefore, when quantifying integration, it may be necessary to consider the perceptual properties of the words, which may depend upon the participants identifying them (see Grant, 2002 and *Clinical Applications*, below).

Although the NAM assumes that words are recognized in the context of all other words in the lexicon, this research serves as the first test of whether evaluating a stimulus word in relation to all other words in the lexicon accounts for greater variance in identification accuracy than does comparing a word to perceptually similar words only. The results support the NAM's predictions: Phi-square ∑NWP (which includes the influence of all words in the lexicon) accounts for more variance in spoken word identification accuracy than does a summed density of perceptually similar competitors (e.g., those in Density A). This supports the proposal that the stimulus input may activate a large set of candidate items in memory, and that the resulting competition of these units influences the likelihood of word recognition. Had both measures accounted for similar amounts of explained variance, it would have suggested that a threshold of perceptual similarity would need to be reached in order for a competitor's word decision unit to provide competition for the stimulus word's decision unit.

The finding that frequency weighted competitor density predicted word identification less effectively than competitor density alone was unexpected. The NAM assumes that word decision units (both those of the stimulus word and of all competitor words) are sensitive to higher-level lexical information like word frequency. It also assumes that word decision units are interconnected such that they can monitor the level of activity of other units in the system. Therefore, competitors that are both perceptually similar and occur frequently should provide more competition than words that are perceptually similar and occur infrequently. This hypothesis was not supported in the current study. Indeed, for all three modalities, competitor density predicted spoken word identification more effectively than did frequency weighted competitor density (but see above discussion on potential implications of using frequency norms based on written, rather than spoken, language).

Because the NAM proposes that word decision units are initially activated based solely on their perceptual similarity to the input, and only later influenced by higher level lexical information like frequency, it could be argued that only the frequencies of perceptually similar competitors should be evaluated. Based on the model, the word decision unit that corresponds to a highly frequent but perceptually dissimilar competitor will initially receive very little or no activation, so frequency information about that competitor may not influence recognition of the stimulus word at all. Weighting perceptual similarity by competitor frequency should theoretically address this issue (words that are perceptually dissimilar, when multiplied by frequency, still render small values). However, it is possible that, given the large number of these competitors, the combined influence may have a significant influence on the FWNWP outcome. Given this possibility, a follow-up analysis was conducted in which only the most perceptually similar competitors (both the 10 and the 50 most similar competitors)[16] were weighted by their frequency. This analysis renders a value that first evaluates perceptual similarity, and then, given high enough similarity, weights by frequency. However, this analysis

[16] Although these specific values were selected somewhat arbitrarily, they were chosen to reflect the influence of the most similar cluster of competitors (10) as well as a larger cluster of less similar words, but still potentially confusable words (50).

also failed to account for significant variance beyond that accounted for by unweighted competitor density in any modality.

This finding cannot be easily interpreted in the framework for the NAM. Although frequency-weighting neighbor similarity has been used previously (Luce & Pisoni, 1998; Auer, 2002), these studies failed to report the correlation between competitor similarity and identification accuracy as well as the correlation between frequency-weighted competitor similarity and identification accuracy separately. Instead, only frequency-weighted competitor similarity was reported. Therefore, it is unclear whether the current results are novel. It is also be possible that frequency effects take time to develop. In that case, responses made quickly may be less susceptible to frequency effects. To address this in future studies, measures of reaction time (in addition to identification accuracy scores) could be collected. If frequency effects do take time to develop, words that are responded to quickly would be expected to show less sensitivity to word frequency than would slowly responded to words.

An alternative to the mechanisms of competition described by the NAM is offered by Exemplar models (see Nosofsky, 1988). Proponents of these models maintain that classification and recognition judgments are based on similarity comparisons with stored exemplars. A key distinction between models of this type and Activation-Competition models is that Activation-Competition models include a mechanism for lateral inhibition or competition between words. Instead, exemplar models propose that differences in recognition performance are due to varying degrees of correspondence between the input and the exemplar. Evidence against this position comes from the finding that semantic context can influence recognition accuracy: the addition of semantic context to word recognition tests increases performance, particularly for words with high-frequency neighbors (Sommers & Danielson, 1999). Exemplar models, which stress the degree of perceptual match between the input and the lexical representation, have difficulty accounting for these findings. The NAM, however, can explain these results by positing that semantic information increases activation levels on word decision units that are consistent with the context without increasing activation on word decision units that are inconsistent with context.

Potential Limitations in Quantifying Competition

These results support the use of measures of lexical density based on the Phi-square statistic and provide additional evidence that the NAM successfully models human performance in A-only, V-only, and AV word recognition. Given that the NAM proposes that the amount of competition a word provides depends on its perceptual similarity to the stimulus word, it is not surprising that overcoming confounds in measuring similarity will increase the predictive power of the measures. There are several other issues pertaining to quantifying perceptual similarity that, if properly addressed, may increase the predictive power further.

In the current study, syllable identification tasks consisted of identifying individual phonemes, embedded in a consistent phonemic context. However, it is possible that the patterns of confusion observed could vary if the phonemic context were different, due to coarticulation. In that case, the single phoneme identifications may not entirely represent the confusability of phoneme strings that appear in real words. As an extreme example, it may be that a given pair of vowels is very confusable when they follow a stop consonant, but if they follow a fricative consonant, they are more clearly distinguishable. The current system of confusions does not allow for such nuanced distinctions.

Another limitation of using single phoneme identifications as a metric for similarity is that it is not possible to assess the similarity of individual phonemes and phoneme clusters. Some investigations (Auer, 2009) have included consonant clusters in viseme groups with single consonants. If two perceptually similar consonants occur in succession, it could be that a CCVC word could be perceptually similar to, and therefore provide strong competition for, a CVC stimulus word. For example, in a visual presentation, /stop/ and /top/ may be easily confused. Because the current investigation only includes CVC competitors (i.e., /top/ is compared to /tip/ and /hop/, but not /stop/ or /trip/), it may over or underestimate the average competitor density of a stimulus word, based on whether it has perceptually similar words that are not CVCs.

This analysis reveals another area that should be addressed in future investigations: how can the perceptual similarity of a stimulus word and a competitor word of different lengths (i.e., those that differ by the addition or subtraction of a phoneme) be quantified? For example, to assess the similarity of the stimulus word /top/ to /step/, if the words are lined up at the vowels, the /t/ of the stimulus word is aligned with the /t/ of the competitor word. However, the /s/ of the competitor words does not align with any phoneme of the stimulus word. In this case, the probability of the phoneme /s/ is, conceptually, the probability of identifying /s/ when no phoneme was presented. Luce (1986) resolved this issue by including a "null" phoneme in the syllable identification tasks. In some phoneme identification trials, no phoneme was presented, but participants were still forced to make a decision about what they heard. Participants also had the option of making the response that no phoneme had been presented. This enabled calculating conditional probabilities of identifying a specific phoneme when none was identified [e.g., $p(\text{s}|\text{ø})$] or the probability of failing to detect that a given phoneme had been presented [e.g., $p(\text{ø}|\text{s})$]. Using this method, Luce could calculate perceptual similarity for competitors that were longer or shorter than the stimulus words.

The method of including a null response works well for the A-only domain, when phonemes are masked and the background noise is perceptually similar to the signal. However, it is difficult to translate to the V-only modality where task difficulty stems not from similarity between signal and noise, but from an underspecified signal. The detection of a mouth movement is very salient, even if the identification of that mouth movement is difficult. Therefore, it seems extremely unlikely that a participant would ever fail to notice a speaker opening their mouth (choose the null response) or identify an unmoving face as a speaker producing a specific phoneme. Therefore, another method seems necessary for calculating the perceptual similarity of two words of differing length.

Finally, another possible limitation of the methods used here lies in the fact that different SNRs were used for the vowel identification task and the consonant and word tasks. Based on pilot testing, an appropriate SNR for consonant and word identification resulted in ceiling level performance for vowel identification. If patterns of syllable confusions differ by SNR, it is possible that the confusions obtained from the vowel identification task at one SNR do not ideally reflect the perceptual experience of perceiving vowels in words at another SNR (but see Miller & Nicely, 1955 for evidence that feature confusions are relatively stable across SNR). Given that these issues remain and likely add statistical noise to the data, the amount of variance accounted for by measures of perceptual density become more impressive.

Clinical Applications

Measuring lexical density using Phi-square values has several interesting clinical applications. First, understanding the manner by which lexical units are activated and compete with one another may have implications for how spoken word recognition may trained. Lipreading is notoriously hard to train (Tye-Murray, 2008) and methods of instruction have met with only limited success. Established methods for measuring lexical density could inform these training programs by revealing the types of words that are perceptually most distinct, and therefore, a potential starting place for training regimens. They could also inform which phonemic distinctions (e.g., /t/ vs /d/) would be most helpful at differentiating between most words, and therefore, worth the most aggressive instruction.

Additionally, because the only input necessary to calculate lexical densities is phoneme confusion matrices, it is possible to map the density of lexical space for specific population group. For example, the perceptual information about speech that cochlear implant users have access to is very different than that received by normal-hearing individuals. Therefore, it would not be surprising to find that patterns of confusion, and therefore lexical densities, may differ between population groups. If this is the case, spoken word identification could likely be best predicted by patterns of confusability generated by individuals of the same population.

Population-specific measures of lexical density may also inform research on word recognition in older adults. Evidence exists that older adults may be especially impaired at recognizing words from regions of the lexicon that are perceptually dense [as quantified by the one-phoneme shortcut method (Sommers, 1996)]. However, it is possible that older adults are less able to make distinctions between phonemic contrasts than are younger adults, due to either sensory deficits (i.e., presbycusis) or age-related cognitive changes (see Pichora-Fuller, 2003). Therefore, for any given stimulus word, there may be more words that serve as potential competitors for older adults than for younger adults, assuming that older adults will show impaired phoneme discrimination, compared with younger adults. It may be that the observed interaction between age and lexical difficulty is due to age-related changes in lexical density instead of (or, more likely, in addition to) age related cognitive changes (e.g., inhibitory deficits).

Conclusions

The present study evaluated methods of quantifying the perceptual similarity of speech sounds and explored the processes of activation and competition in A-only, V-only and AV spoken word recognition. Overall, the results suggest that measures of perceptual similarity based on the Phi-square statistic are effective at modeling the structure of the lexicon and predicting spoken word identification. In addition, the results demonstrate some similarities in the processes underlying spoken word recognition across modalities. Namely, stimulus word frequency and competitor density account for significant unique variance in spoken word recognition in each modality. These findings support and extend the scope of Activation-Competition models like the NAM and suggest that the processes underlying spoken word recognition are not specific to the A-only domain.

REFERENCES

Aloufy, S., Lapidot, M., & Myslobodsky, M. (1996). Differences in susceptibility to the "blending illusion" among native Hebrew and English speakers. *Brain and Language*, *53*, 51-57.

Auer, E. (2002). The influence of the lexicon on speech read word recognition: Contrasting segmental and lexical distinctiveness. *Psychonomic Bulletin & Review, 9*, 341-347.

Aitchison, J. (2003) *Words in the mind: An introduction to the mental lexicon.* London: Wiley-Blackwell.

Auer, E. (2009). Spoken word recognition by eye. *Scandinavian Journal of Psychology, 50*, 419-425.

Balota, D., Yap, M., Cortese, M., Hutchison, K., Kessler, B., Loftis, B., Neely, J., Nelson, D., Simpson, G., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39,* 445-459.

Bernstein, L.E., Iverson, P., & Auer, E.T, Jr. (1997). Elucidating the complex relationships between phonetic perception and word recognition in audiovisual speech perception. In C. Benoit & R. Campbell (Eds.) *Proceedings of the ESCA/ESCOP workshop on audio-visual speech processing* (pp.21-24). Rhodes, Greece.

Bernstein, L. E., Demorest, M. E., & Tucker, P. E. (2000). Speech perception without hearing. *Perception & Psychophysics*, *62*, 233-252.

Binnie, C., Jackson, P., Montgomery, A. (1976). Visual intelligibility of consonants: A lipreading screening test with implications for aural rehabilitation. *Journal of*

Speech and Hearing Disorders, 41, 530-539.

Binnie, C., Montgomery, A., Jackson, P. (1974). Auditory and visual contributions to the perception of consonants. *Journal of Speech and Hearing Research, 17,* 619-634.

Blamey, P., Cowan, R., Alcantara, J., Whitford, L., Clark, G. (1989). Speech perception using combinations of auditory, visual, and tactile information. *Journal of Rehabilitation Research and Development, 26*, 15-24.

Calvert, G., Bullmore, E., Brammer, M., Campbell, R., Williams, S., McGuire, P., Woodruff, P., Iversen, S., & David, A. (1997). Activation of auditory cortex during silent lipreading. *Science, 276*, 593-596.

Clouser, R. A. (1977). Relative phoneme visibility and lipreading performance. *Volta Review*, *79*, 27-34.

Dancer, J., Krain, M., Thompson, C., Davis, P., & Glen, J. (1994). A cross-sectional investigation of speechreading in adults: Effects of age, gender, practice, and education. *Volta Review, 96,* 31-40.

Dahan, D., & Magnuson, J. (2006). Spoken Word Recognition. In M. Traxler & M. Gernsbacher (Eds), *Handbook of Psycholinguistics, 2nd Edition* (pp. 249-283). Boston: Elsevier Academic Press.

Dirks, D., Takayana, S., & Moshfegh, A. (2001). Effects of lexical factors on word recognition among normal-hearing and hearing-impaired listeners. *Journal of the American Academy of Audiology, 12*, 233-244.

Eisenberg, L., Martinez, A., Holowecky, S., & Pogorelsky, S. (2002). Recognition of lexically controlled words and sentences by children with normal hearing and children with cochlear implants. *Ear & Hearing, 23*, 450-462.

Elphick, R. (1996). Issues in comparing the speechreading abilities of hearing-impaired and hearing 15 to 16 year-old pupils. *British Journal of Educational Psychology*. *66*, 357-365.

Erber, N. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech & Hearing Research*, *12*, 423-425.

Farrimond, T. (1959). Age differences in the ability to use visual cues in auditory communication. *Language and Speech*, *2*, 179-192.

Feld, J., & Sommers, M. (2009). Lipreading, processing speed, and working memory in younger and older adults. *Journal of Speech, Language, and Hearing Research, 52,* 1555-1565

Fisher, C. (1968) Confusions among visually perceived consonants. *Journal of Speech and Hearing Research, 11*, 796-804.

Gentilucci, M., & Cattaneo, L. (2005). Automatic audiovisual integration in speech perception. *Experimental Brain Research, 167*, 66-75.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language, 28,* 501–518.

Grant, K., & Seitz, P. (1998). Measures of auditory-visual integration in nonsense syllables and sentences. *Journal of the Acoustic Society of America, 104*, *2438-*

Grant, K. (2002). Measures of auditory-visual integration for speech understanding: A theoretical perspective. *Journal of the Acoustical Society of America, 112*, 30-33.

Grant, K., Walden, B., & Seitz, F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *Journal of the Acoustical Society of America*. *103*, 2677-2690.

Green, K., Kuhl, P., Meltzoff, A. & Stevens, E. (1991) Integrating speech information across talkers, gender, and sensory modalities: female faces and male voices in the McGurk effect. *Perception and Psychophysics, 50,* 524.

Hall, D., Fussell, C., & Summerfield, Q. (2005). Reading fluent speech from talking faces: Typical brain networks and individual differences. *Journal of Cognitive Neuroscience, 17*, 939-953.

Irwin, J. R., Whalen, D. H., & Fowler, C. A. (2006). A sex difference in visual influence on heard speech. *Perception & Psychophysics*. *68*, 582-592.

Iverson, P., Bernstein, L., & Auer, E. (1998) Modeling the interaction of phonemic intelligibility and lexical structure in audiovisual word recognition. *Speech Communication, 26*, 45-63.

Jackson, P. (1988). The theoretical minimal unit for visual speech perception: Visemes and coarticulation. *The Volta Review, 90*, 99-115.

Jeffers, J. & Barley, M. (1971). *Speechreading (lipreading).* Springfield, Illinois: Thomas.

Johnson, S. (1775). *A Journey to the Western Islands of Scotland*. London: Scholar.

Johnson, F. M., Hicks, L. H., Goldberg, T., & Myslobodsky, M. S. (1988). Sex differences in lipreading. *Psychonomic Bulletin & Review*, *26*, 106-108.

Jusczyk, P.W., Luce, P.A. (2002). Speech perception and spoken word recognition: Past and present. *Ear & Hearing, 23*, 2-40.

Kaiser, A., Kirk, K., Lachs, L., & Pisoni, D. (2003). Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *Journal of Speech, Language, and Hearing Research*, *46*, 390-404.

Kirk, K., Pisoni, D., & Osberger, M. (1995). Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear & Hearing, 16*, 470-481.

Kirk, K., Hay-McCutcheon, M., Holt, R., Gao, S., Qi, R., & Gerlain, B. (2007). Audiovisual spoken word recognition by children with cochlear implants. *Audiological Medicine, 5,* 250-261.

Landauer, T., & Streeter, K. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior, 12*, 119-131

Lidestam, B., Lyxell, B., & Andersson, G. (1999). Speech-reading: Cognitive predictors and displayed emotion. *Journal of Scandinavian Audiology*, *28*, 211-217.

Luce, R. (1959). *Individual Choice Behavior*. New York: Wiley

Luce, P. (1986). *Neighborhoods of words in the mental lexicon*. Doctoral Dissertation. Technical Report No. 6, Indiana University.

Luce, P., & Pisoni, D. (1998). Recognizing spoken words: The Neighborhood Activation Model. *Ear & Hearing, 19,* 1-36.

Lund, K. & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers, 28,* 203-208.

Lyxell, B., & Holmberg, I. (2000). Visual speechreading and cognitive performance in hearing-impaired and normal hearing children (11-14 years). *British Journal of*

Educational Psychology, *70*, 505-518.

Lyxell, B., & Rönnberg, J. (1989). Information-processing skill and speech-reading. *British Journal of Audiology, 23,* 339-347.

Lyxell, B., & Rönnberg, J. (1991). Word discrimination and chronological age related to sentence-based speech-reading skill. *British Journal of Audiology, 25,* 3-10.

Lyxell, B., & Rönnberg, J. (1992). The relationship between verbal ability and sentence-based speechreading. *Scandinavian Audiology*, *21*, 67-72.

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

Macchi, M. (1980). Identification of vowels spoken in isolation versus vowels spoken in consonantal context. *Journal of the Acoustical Society of America, 68*, 1636-1642.

Marslen-Wilson, W. (1995). Activation, competition, and frequency in lexical access. In G. Altman, (Ed.) *Cognitive Models of Speech Processing* (pp. 149-172). Boston: MIT Press.

Massaro, D. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, New Jersey: Lawrence Earlbaum Assoc.

Massaro, D., & Cohen, M. (2000). Tests of auditory-visual integration efficiency within the framework of the fuzzy logical model of perception. *Journal of the Acoustic Society of America, 108,* 784–789.

Mattys, S., Bernstein, L., & Auer, E. (2002). Stimulus-based lexical distinctiveness as a general word-recognition mechanism. *Perception & Psychophysics, 64*, 667-679.

McClelland, J., & Elman, J. (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.

Middelweerd, M., & Plomp, R. (1987). The effect of speechreading on the speech-reception threshold of sentences in noise. *Journal of the Acoustical Society of America*, *82*, 2145-2147.

Miller, G., & Nicely, P. (1955). An analysis of perceptual confusions among some English Consonants. *The Journal of the Acoustical Society of America, 27*, 338-352.

Morton, J. (1979). Word recognition. In Morton, J. & Marshall, J.C. (Eds.). *Psycholinguistics Volume 2 – Structures and Processes*. London: Paul Elek.

Nitchie, E. (1930). *Lip reading principles and practise.* New York: Fredrick A. Stokes.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.

Nosofsky, R. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 700-708.

Owens, E., & Blazek, B. (1985). Visemes observed by hearing-impaired and normal-hearing adult viewers. *Journal of Speech & Hearing Research*, *28*, 381-393.

Pichora-Fuller, M. (2003). Processing speed and timing in aging adults: Psychoacoustics, speech perception, and comprehension. *International Journal of Audiology, 42,* S59-S67.

Rönnberg, J. (1990). Cognitive and communicative function: The effects of chronological age and “handicap age.” *European Journal of Cognitive Psychology. 2,* 253-273.

Rosenblum, L. (2004). The primacy of multimodal speech perception. In Pisoni, David B. & Robert E. Remez (Eds). *The Handbook of Speech Perception*. Blackwell Publishing.

Rosenblum, L., Schmuckler, M., & Johnson, J. (1997). The McGurk effect in infants. *Perception and Psychophysics, 59*, 347-357.

Savin, H. (1963). Word-frequency effect and errors in the perception of speech. *Journal of the Acoustical Society of America, 35,* 200-206.

Sekiyama, K.& Tohkura, Y. (1991). Japanese subjects hearing Japanese syllables of high auditory intelligibility. *Journal of the Acoustical Society of America*, 90, 1797–1805.

Simmons, A. A. (1959). Factors related to lipreading. *Journal of Speech & Hearing Research*, *2*, 340-352.

Sommers, M (1996). The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition. *Psychology and Aging, 11*, 333-341.

Sommers, M., & Danielson, S. (1999). Inhibitory processes and spoken word recognition in younger and older adults: The interaction of lexical competition and semantic context. *Psychology and Aging, 14*, 458-472.

Sommers M., Spehar B, Tye-Murray N. (2005) The effects of signal-to-noise ratio on auditory-visual integration: integration and encoding are not independent. *Journal of the Acoustical Society of America*, *117,* 2574.

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

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, *26*, 212-215.

Summerfield, Q. (1991). Visual perception of phonetic gestures. In I.G. Mattingly & M. Studdert-Kennedy (Ed.) *Modularity and the motor theory of speech perception*. (pp. 117-137). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions: Biological Sciences*, *335,* 71-78.

Tye-Murray, N. (2008). *Foundations of aural rehabilitation: Children, adults, and their family members.* Delmar Cengage Learning: Clifton Park, NY

Tye-Murray, N., Sommers, M., & Spehar, B. (2007a). Auditory and visual lexical neighborhoods in audiovisual speech perception. *Trends in Amplification 11*, 233-241.

Tye-Murray, N., Sommers, M., & Spehar, B. (2007b). The effects of age and gender on lipreading abilities. *Journal of the American Academy of Audiology, 18*, 883-892.

Vitevitch, M., & Luce, P. (1998). When words compete: Levels of processing in spoken word perception. *Psychological Science*, 9, 325–329.

Walden, B., Prosek, R., Montgomery, A., Scherr, C., & Jones, C. (1977). Effects of training on the visual recognition of consonants. *Journal of Speech and Hearing Research, 20,* 130-145.

Washington University in St. Louis Speech and Hearing Lab Neighborhood Database as a web-based implementation of the 20, 000-word Hoosier Mental Lexicon http://128.252.27.56/ neighborhood/Home.asp.

Woodhouse, L., Hickson, L., & Dodd, B. (2009). Review of visual speech perception by hearing and hearing-impaired people: clinical implications. *International Journal of Communication Disorders, 44*, 253-270.

APPENDIX A: STIMULUS LISTS

A-only list		V-only List		AV List	
bad	lurk	bag	knit	bade	mace
balm	main	balk	knot	badge	mad
bar	mat	ban	known	bail	mar
bead	meal	bash	lab	bait	mass
bees	mime	beak	lame	base	match
bide	mole	beat	leap	beam	mike
bing	mooch	bet	leer	beck	mill
bird	mud	bib	lick	bed	moon
boat	muff	big	long	beige	move
bob	mum	birch	lose	bell	mug
boon	near	bole	maim	bike	mutt
botch	node	booth	man	bit	myth
both	noise	bull	mead	biz	name
boys	noun	bun	meat	boil	nap
burrs	page	cad	merge	boot	nib
buys	peace	calf	mice	buck	niche
cage	peal	calm	midge	bud	nick
can	peck	cap	miff	bus	nil
caught	peg	case	mirth	cab	pack
cease	pile	chalk	mock	came	pearl
chase	pine	chap	mode	cane	peas
cheek	pit	check	mom	catch	peep
chill	pole	cheer	mood	cheap	pen
chock	poof	chief	moor	chive	perk
choke	poor	chip	moss	comb	pick
chug	pun	chore	mouse	cone	pill
chum	purse	chute	mouth	core	pith
coach	race	cob	net	cowl	poke
code	rain	cog	newt	cut	pope
coin	ram	cook	nip	dash	pull
cup	rang	cop	null	dead	rag
curse	rate	cows	pail	dear	rave
deal	ream	cuff	pall	deep	rig
dean	rear	cull	peeve	dell	ring
debt	rife	curb	perm	ditch	room
died	riff	curl	pet	dock	rouge
dip	roar	date	pies	dodge	rough
done	role	daub	pin	dose	royal

dowel	rule	deck	pouch	dove	rug
dude	rune	den	psalm	down	said
dug	ruse	did	puff	dual	sake
fad	rut	dies	pup	face	sass
fade	sage	dig	rail	fed	seep
fair	sail	dill	rev	feel	seethe
fall	sate	din	rice	fine	serve
fang	says	dive	rim	fish	shake
fell	scene	doll	roam	foal	shell
fib	search	dot	robe	fudge	shin
fief	shack	doze	roof	gage	shod
foes	share	duck	rook	gave	shop
fog	shed	fab	root	gem	showed
fun	she'd	fain	rope	get	shun
gal	shirt	femme	rouse	gneiss	sick
gauze	should	fill	sat	goat	sign
gawk	sill	foam	seal	goes	sip
gear	size	folk	shame	gosh	sob
gene	soothe	for	shone	guide	soil
gin	soup	fuss	shout	heap	soon
gnash	sour	gab	shove	hers	soot
gong	serge	gall	sin	him	sop
hail	teach	gang	sing	hiss	sued
half	that	gape	sirs	hitch	tang
heave	thatch	girth	suit	hook	tech
heck	they'll	gob	tab	hung	teeth
hem	thick	good	tack	jade	terse
highs	thud	Goth	tail	jeer	their
hub	tic	gull	tan	jib	thin
hutch	tide	gush	ten	jock	this
jab	tile	gym	than	join	tomb
jaws	toes	hall	they'd	josh	tome
joke	tooth	hat	thought	keel	ton
kiss	tote	hawk	thyme	kit	tongue
kneel	touch	heal	tire	knave	veer
knees	towel	hedge	toil	knoll	veg
knife	tune	hid	tout	lad	vies
lac	use	hill	voice	late	wack
lace	vague	hip	waive	learn	wade
laid	vied	hole	ways	ledge	was
league	wain	home	wick	lies	wash

lean	weed	hull	wire	lip	white
leash	whig	Hun	wood	loaf	who've
leave	win	hush	wooed	loam	wile
lied	wipe	jeep	wool	loathe	womb
load	with	jewel	worm	loose	writhe
lobe	woke	jig	worse	loss	yak
look	wrath	joss	wren	lot	yawn
loud	writ	juke	yet	louse	yin
luck	yap	ken	yoke	lurch	yore
lug	yowl	keyed	yore	lure	youth
lull	zone	knight	zip	liar	zap

APPENDIX B: PHONEME CONFUSION MATRICES

A-only % accuracy consonant confusion matrix

Response Choice (columns); Target Phoneme (rows)

	b	tʃ	d	f	g	h	dʒ	k	l	m	n	ŋ	p	r	s	ʃ	t	θ	ð	v	w	y	z	ʒ
b	.73	.01	.00	.02	.00	.00	.00	.00	.00	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.19	.00	.00	.00	.00
tʃ	.00	.84	.00	.00	.00	.00	.04	.01	.00	.00	.00	.00	.00	.00	.00	.05	.02	.01	.00	.00	.00	.00	.00	.02
d	.01	.00	.71	.00	.09	.00	.02	.00	.00	.00	.00	.00	.00	.01	.00	.00	.02	.01	.02	.00	.00	.00	.08	.00
f	.23	.00	.00	.41	.00	.02	.00	.00	.03	.02	.02	.01	.05	.00	.05	.01	.00	.05	.01	.06	.00	.00	.00	.00
g	.00	.00	.04	.00	.81	.00	.00	.01	.02	.00	.01	.02	.00	.00	.00	.00	.00	.01	.00	.00	.00	.02	.03	.01
h	.01	.00	.00	.04	.00	.81	.00	.04	.00	.00	.00	.00	.06	.00	.00	.00	.01	.01	.00	.01	.00	.00	.00	.00
dʒ	.00	.02	.13	.00	.04	.00	.47	.00	.01	.00	.00	.00	.00	.00	.00	.01	.01	.01	.02	.01	.00	.09	.01	.16
k	.00	.00	.00	.01	.00	.01	.00	.80	.00	.00	.00	.00	.04	.00	.00	.00	.10	.00	.00	.00	.00	.00	.00	.00
l	.10	.00	.00	.00	.00	.00	.00	.00	.62	.07	.00	.01	.00	.02	.00	.00	.00	.00	.00	.04	.11	.02	.00	.00
m	.18	.00	.00	.02	.00	.00	.00	.00	.06	.57	.04	.01	.01	.00	.00	.00	.00	.01	.00	.07	.00	.00	.01	.00
n	.00	.00	.05	.00	.01	.01	.00	.02	.00	.00	.73	.05	.00	.00	.00	.00	.08	.00	.00	.00	.00	.01	.01	.00
ŋ	.01	.00	.00	.00	.03	.00	.00	.00	.08	.12	.27	.40	.00	.01	.00	.00	.01	.01	.02	.00	.00	.01	.01	.00
p	.02	.00	.00	.02	.00	.08	.00	.07	.00	.00	.00	.00	.73	.00	.00	.00	.05	.00	.00	.00	.00	.00	.00	.00
r	.02	.00	.01	.01	.03	.01	.00	.01	.17	.03	.01	.01	.00	.52	.00	.00	.00	.00	.01	.13	.04	.00	.00	.00
s	.07	.00	.01	.05	.03	.01	.00	.02	.00	.00	.00	.00	.01	.02	.29	.02	.01	.09	.07	.05	.00	.00	.20	.02
ʃ	.00	.07	.01	.01	.01	.00	.05	.00	.00	.00	.00	.00	.00	.00	.02	.60	.01	.01	.01	.00	.00	.00	.01	.17
t	.05	.01	.02	.07	.00	.04	.00	.06	.00	.00	.01	.02	.12	.00	.00	.00	.50	.04	.03	.01	.00	.00	.00	.00
θ	.13	.01	.02	.12	.02	.04	.01	.01	.02	.01	.02	.01	.02	.01	.09	.00	.03	.23	.09	.06	.00	.00	.03	.01
ð	.09	.00	.30	.01	.09	.00	.00	.01	.01	.00	.02	.00	.00	.02	.00	.00	.01	.06	.14	.10	.00	.00	.12	.00
v	.18	.00	.02	.03	.01	.00	.01	.00	.10	.01	.02	.01	.01	.03	.00	.00	.00	.02	.02	.48	.01	.00	.02	.02
w	.01	.00	.00	.00	.00	.00	.00	.00	.10	.01	.00	.00	.00	.01	.00	.00	.00	.00	.00	.02	.83	.00	.00	.00
y	.00	.00	.06	.00	.13	.00	.03	.00	.13	.00	.01	.02	.00	.05	.00	.00	.01	.00	.01	.00	.09	.42	.01	.02
z	.04	.00	.12	.00	.04	.00	.02	.00	.04	.00	.01	.01	.00	.01	.01	.00	.00	.07	.16	.04	.01	.04	.36	.02
ʒ	.00	.02	.02	.00	.01	.00	.20	.00	.00	.00	.00	.01	.00	.00	.00	.02	.00	.00	.01	.00	.00	.06	.03	.60

A-only Phi-square values consonant confusion matrix

Response Choice

Target Phoneme	b	tʃ	d	f	g	h	dʒ	k	l	m	n	ŋ	p	r	s	ʃ	t	θ	ð	v	w	y	z	ʒ
b	1	.03	.04	.32	.02	.05	.04	.03	.15	.27	.02	.04	.05	.14	.16	.04	.11	.24	.19	.41	.05	.03	.11	.03
tʃ	.03	1	.05	.04	.03	.03	.10	.05	.03	.03	.04	.03	.04	.02	.06	.19	.06	.06	.04	.05	.01	.05	.05	.11
d	.04	.05	1	.07	.19	.04	.24	.05	.04	.06	.11	.09	.04	.07	.18	.09	.10	.16	.50	.11	.02	.18	.32	.10
f	.32	.04	.07	1	.06	.13	.06	.07	.18	.28	.07	.13	.14	.15	.30	.08	.26	.49	.21	.32	.07	.08	.17	.05
g	.02	.03	.19	.06	1	.03	.13	.04	.06	.06	.10	.12	.03	.10	.11	.06	.07	.12	.21	.10	.04	.25	.18	.08
h	.05	.03	.04	.13	.03	1	.03	.11	.04	.06	.05	.04	.22	.05	.10	.04	.19	.15	.05	.06	.02	.03	.04	.02
dʒ	.04	.10	.24	.06	.13	.03	1	.04	.06	.05	.10	.09	.03	.07	.12	.22	.09	.14	.22	.12	.04	.27	.24	.48
k	.03	.05	.05	.07	.04	.11	.04	1	.02	.04	.09	.04	.18	.04	.09	.05	.24	.10	.05	.05	.01	.04	.04	.03
l	.15	.03	.04	.18	.06	.04	.06	.02	1	.26	.04	.19	.04	.31	.10	.03	.08	.16	.12	.26	.27	.25	.14	.04
m	.27	.03	.06	.28	.06	.06	.05	.04	.26	1	.08	.24	.06	.19	.15	.04	.13	.24	.18	.30	.09	.09	.15	.04
n	.02	.04	.11	.07	.10	.05	.10	.09	.04	.08	1	.35	.07	.06	.08	.06	.16	.13	.12	.08	.02	.11	.12	.07
ŋ	.04	.03	.09	.13	.12	.04	.09	.04	.19	.24	.35	1	.04	.18	.10	.05	.09	.16	.13	.16	.09	.17	.15	.07
p	.05	.04	.04	.14	.03	.22	.03	.18	.04	.06	.07	.04	1	.05	.10	.03	.30	.14	.06	.07	.02	.03	.04	.02
r	.14	.02	.07	.15	.10	.05	.07	.04	.31	.19	.06	.18	.05	1	.14	.04	.09	.18	.18	.32	.16	.24	.16	.04
s	.16	.06	.18	.30	.11	.10	.12	.09	.10	.15	.08	.10	.10	.14	1	.13	.21	.52	.38	.26	.04	.11	.44	.11
ʃ	.04	.19	.09	.08	.06	.04	.22	.05	.03	.04	.06	.05	.03	.04	.13	1	.07	.12	.09	.09	.01	.09	.11	.29
t	.11	.06	.10	.26	.07	.19	.09	.24	.08	.13	.16	.09	.30	.09	.21	.07	1	.31	.17	.16	.03	.08	.14	.06
θ	.24	.06	.16	.49	.12	.15	.14	.10	.16	.24	.13	.16	.14	.18	.52	.12	.31	1	.39	.36	.07	.14	.35	.10
ð	.19	.04	.50	.21	.21	.05	.22	.05	.12	.18	.12	.13	.06	.18	.38	.09	.17	.39	1	.32	.05	.20	.57	.10
v	.41	.05	.11	.32	.10	.06	.12	.05	.26	.30	.08	.16	.07	.32	.26	.09	.16	.36	.32	1	.12	.18	.26	.09
w	.05	.01	.02	.07	.04	.02	.04	.01	.27	.09	.02	.09	.02	.16	.04	.01	.03	.07	.05	.12	1	.18	.08	.02
y	.03	.05	.18	.08	.25	.03	.27	.04	.25	.09	.11	.17	.03	.24	.11	.09	.08	.14	.20	.18	.18	1	.24	.16
z	.11	.05	.32	.17	.18	.04	.24	.04	.14	.15	.12	.15	.04	.16	.44	.11	.14	.35	.57	.26	.08	.24	1	.16
ʒ	.03	.11	.10	.05	.08	.02	.48	.03	.04	.04	.07	.07	.02	.04	.11	.29	.06	.10	.10	.09	.02	.16	.16	1

V-only % accuracy consonant confusion matrix

Target Phoneme \ Response Choice	b	tʃ	d	f	g	h	dʒ	k	l	m	n	ŋ	p	r	s	ʃ	t	θ	ð	v	w	y	z	ʒ
b	.45	.00	.00	.00	.00	.00	.00	.00	.00	.24	.00	.00	.28	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
tʃ	.00	.20	.00	.00	.00	.00	.29	.00	.00	.00	.00	.00	.00	.00	.00	.15	.01	.00	.00	.00	.00	.01	.00	.32
d	.00	.01	.14	.01	.02	.00	.05	.02	.01	.00	.04	.01	.00	.01	.07	.02	.33	.01	.01	.01	.00	.05	.15	.03
f	.00	.00	.00	.64	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.00	.00	.00	.00	.00	.34	.00	.00	.00	.00
g	.00	.00	.01	.00	.21	.04	.00	.18	.09	.00	.07	.09	.00	.01	.00	.00	.02	.00	.00	.00	.00	.25	.00	.00
h	.01	.00	.00	.00	.08	.60	.00	.09	.03	.01	.01	.06	.01	.01	.00	.00	.01	.00	.00	.00	.00	.07	.00	.00
dʒ	.00	.14	.02	.00	.01	.00	.28	.00	.00	.00	.00	.00	.00	.00	.02	.16	.03	.00	.00	.00	.00	.00	.03	.29
k	.00	.00	.01	.00	.20	.06	.01	.18	.11	.00	.07	.07	.01	.01	.00	.00	.01	.00	.00	.00	.00	.23	.01	.00
l	.00	.00	.07	.00	.02	.00	.00	.02	.41	.00	.12	.02	.00	.00	.02	.00	.07	.05	.05	.00	.00	.08	.03	.00
m	.46	.00	.00	.00	.00	.00	.00	.00	.00	.30	.00	.00	.23	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
n	.00	.00	.07	.00	.06	.00	.00	.05	.22	.00	.20	.05	.00	.01	.01	.00	.07	.02	.02	.00	.00	.17	.02	.00
ŋ	.00	.00	.00	.00	.19	.11	.00	.14	.19	.00	.07	.09	.00	.01	.00	.00	.01	.00	.00	.00	.00	.18	.00	.00
p	.47	.00	.00	.00	.00	.00	.00	.00	.00	.21	.00	.00	.29	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
r	.01	.01	.01	.18	.02	.00	.02	.01	.01	.00	.01	.02	.01	.28	.00	.01	.01	.01	.00	.23	.11	.01	.01	.01
s	.00	.02	.08	.06	.01	.00	.03	.01	.00	.00	.02	.01	.00	.00	.17	.01	.19	.01	.01	.04	.00	.02	.28	.03
ʃ	.00	.15	.00	.00	.00	.00	.27	.00	.00	.00	.00	.00	.00	.00	.02	.19	.01	.00	.00	.00	.00	.00	.02	.32
t	.00	.01	.15	.00	.01	.00	.02	.01	.01	.00	.03	.02	.01	.01	.09	.02	.30	.02	.01	.01	.00	.04	.21	.03
θ	.00	.00	.01	.00	.00	.00	.00	.00	.09	.00	.00	.00	.00	.00	.00	.00	.01	.43	.44	.00	.00	.00	.00	.00
ð	.00	.00	.00	.00	.00	.00	.00	.00	.07	.00	.00	.00	.00	.00	.00	.00	.01	.42	.48	.00	.00	.00	.00	.00
v	.00	.00	.00	.51	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.00	.00	.01	.01	.01	.44	.00	.00	.00	.00
w	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.00	.00	.00	.00	.00	.00	.96	.01	.00	.00
y	.00	.00	.07	.00	.09	.00	.01	.08	.09	.00	.11	.04	.01	.01	.02	.00	.10	.01	.02	.00	.00	.28	.05	.01
z	.00	.01	.12	.02	.01	.00	.02	.02	.01	.00	.02	.01	.00	.00	.14	.01	.23	.02	.02	.02	.00	.03	.26	.03
ʒ	.00	.16	.01	.00	.00	.00	.25	.00	.00	.00	.00	.00	.00	.00	.02	.17	.03	.00	.00	.00	.00	.00	.01	.33

V -only Phi-square values consonant confusion matrix

Response Choice (columns); Target Phoneme (rows)

	b	tʃ	d	f	g	h	dʒ	k	l	m	n	ŋ	p	r	s	ʃ	t	θ	ð	v	w	y	z	ʒ
b	1	.02	.03	.01	.03	.05	.02	.03	.03	.87	.02	.03	.90	.05	.03	.02	.03	.01	.01	.02	.02	.04	.03	.02
tʃ	.02	1	.15	.01	.04	.03	.79	.05	.05	.01	.05	.03	.02	.08	.13	.85	.12	.02	.02	.03	.01	.06	.11	.85
d	.03	.15	1	.03	.20	.13	.22	.20	.34	.02	.35	.16	.03	.16	.65	.17	.84	.07	.06	.06	.02	.42	.75	.19
f	.01	.01	.03	1	.02	.02	.01	.02	.02	.00	.01	.02	.01	.36	.10	.01	.02	.01	.01	.82	.01	.02	.05	.01
g	.03	.04	.20	.02	1	.36	.05	.90	.35	.02	.51	.77	.03	.13	.13	.04	.17	.08	.07	.04	.02	.61	.15	.05
h	.05	.03	.13	.02	.36	1	.04	.39	.17	.03	.23	.43	.04	.12	.09	.03	.11	.04	.04	.03	.03	.26	.10	.03
dʒ	.02	.79	.22	.01	.05	.04	1	.06	.09	.01	.08	.04	.02	.09	.20	.86	.19	.03	.02	.03	.01	.11	.19	.89
k	.03	.05	.20	.02	.90	.39	.06	1	.36	.02	.52	.82	.03	.13	.13	.05	.17	.08	.07	.04	.02	.60	.15	.05
l	.03	.05	.34	.02	.35	.17	.09	.36	1	.02	.70	.39	.03	.13	.26	.06	.32	.22	.19	.05	.02	.52	.31	.08
m	.87	.01	.02	.00	.02	.03	.01	.02	.02	1	.01	.01	.84	.03	.01	.01	.02	.01	.01	.01	.02	.02	.02	.01
n	.02	.05	.35	.01	.51	.23	.08	.52	.70	.01	1	.51	.02	.13	.24	.06	.31	.15	.12	.04	.02	.71	.28	.07
ŋ	.03	.03	.16	.02	.77	.43	.04	.82	.39	.01	.51	1	.03	.12	.10	.03	.14	.09	.07	.03	.02	.52	.12	.03
p	.90	.02	.03	.01	.03	.04	.02	.03	.03	.84	.02	.03	1	.05	.02	.02	.03	.01	.02	.02	.02	.03	.03	.02
r	.05	.08	.16	.36	.13	.12	.09	.13	.13	.03	.13	.12	.05	1	.20	.08	.14	.05	.04	.41	.14	.15	.16	.09
s	.03	.13	.65	.10	.13	.09	.20	.13	.26	.01	.24	.10	.02	.20	1	.15	.70	.05	.04	.13	.02	.31	.81	.17
ʃ	.02	.85	.17	.01	.04	.03	.86	.05	.06	.01	.06	.03	.02	.08	.15	1	.14	.02	.02	.03	.01	.07	.14	.88
t	.03	.12	.84	.02	.17	.11	.19	.17	.32	.02	.31	.14	.03	.14	.70	.14	1	.07	.06	.05	.02	.39	.82	.16
θ	.01	.02	.07	.01	.08	.04	.03	.08	.22	.01	.15	.09	.01	.05	.05	.02	.07	1	.91	.03	.01	.11	.08	.03
ð	.01	.02	.06	.01	.07	.04	.02	.07	.19	.01	.12	.07	.02	.04	.04	.02	.06	.91	1	.03	.01	.09	.07	.02
v	.02	.03	.06	.82	.04	.03	.03	.04	.05	.01	.04	.03	.02	.41	.13	.03	.05	.03	.03	1	.02	.05	.08	.03
w	.02	.01	.02	.01	.02	.03	.01	.02	.02	.02	.02	.02	.02	.14	.02	.01	.02	.01	.01	.02	1	.03	.02	.01
y	.04	.06	.42	.02	.61	.26	.11	.60	.52	.02	.71	.52	.03	.15	.31	.07	.39	.11	.09	.05	.03	1	.35	.09
z	.03	.11	.75	.05	.15	.10	.19	.15	.31	.02	.28	.12	.03	.16	.81	.14	.82	.08	.07	.08	.02	.35	1	.16
ʒ	.02	.85	.19	.01	.05	.03	.89	.05	.08	.01	.07	.03	.02	.09	.17	.88	.16	.03	.02	.03	.01	.09	.16	1

AV % accuracy consonant confusion matrix

Target Phoneme	Response Choice																							
	b	tʃ	d	f	g	h	dʒ	k	l	m	n	ŋ	p	r	s	ʃ	t	θ	ð	v	w	y	z	ʒ
b	.83	.01	.00	.00	.00	.00	.00	.00	.00	.13	.00	.00	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
tʃ	.00	.66	.00	.00	.00	.00	.06	.00	.00	.00	.00	.00	.00	.00	.00	.12	.03	.00	.00	.00	.00	.00	.00	.12
d	.00	.01	.48	.01	.03	.00	.07	.00	.00	.00	.01	.00	.00	.00	.02	.01	.00	.00	.00	.03	.00	.02	.21	.09
f	.00	.00	.00	.88	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.00	.00	.01	.00	.08	.00	.00	.00	.00
g	.00	.00	.04	.00	.79	.00	.00	.02	.01	.00	.02	.05	.00	.00	.00	.00	.00	.00	.00	.00	.00	.05	.00	.00
h	.00	.00	.00	.00	.01	.87	.00	.10	.00	.00	.00	.00	.00	.00	.00	.00	.01	.00	.00	.00	.00	.00	.00	.00
dʒ	.00	.02	.02	.00	.01	.00	.60	.00	.00	.00	.00	.00	.00	.00	.00	.02	.00	.00	.00	.00	.00	.00	.00	.32
k	.00	.01	.00	.00	.01	.04	.00	.87	.01	.00	.01	.01	.01	.00	.00	.00	.02	.01	.00	.00	.00	.00	.00	.00
l	.00	.00	.00	.00	.00	.00	.00	.00	.80	.01	.06	.02	.00	.00	.01	.00	.00	.01	.02	.01	.00	.01	.04	.00
m	.23	.00	.00	.00	.00	.00	.00	.00	.00	.74	.00	.00	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
n	.00	.00	.04	.00	.02	.01	.00	.00	.03	.00	.73	.10	.00	.00	.00	.00	.01	.01	.01	.00	.00	.04	.00	.00
ŋ	.00	.00	.01	.00	.30	.04	.00	.05	.05	.00	.14	.35	.00	.00	.00	.00	.00	.00	.00	.00	.00	.05	.00	.00
p	.07	.00	.00	.00	.00	.00	.00	.01	.00	.02	.00	.00	.88	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
r	.00	.00	.00	.00	.00	.00	.00	.00	.03	.01	.00	.00	.00	.71	.00	.00	.00	.00	.00	.13	.10	.00	.00	.00
s	.00	.00	.05	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.44	.02	.18	.02	.01	.03	.00	.00	.21	.01
ʃ	.00	.09	.00	.00	.00	.00	.02	.00	.00	.00	.00	.00	.00	.00	.00	.78	.01	.00	.00	.00	.00	.00	.00	.09
t	.00	.01	.01	.00	.00	.01	.00	.06	.01	.00	.00	.00	.01	.00	.01	.00	.86	.01	.00	.00	.00	.00	.00	.00
θ	.00	.00	.00	.01	.00	.00	.00	.00	.01	.00	.00	.00	.00	.00	.01	.00	.16	.58	.20	.00	.00	.00	.01	.00
ð	.00	.00	.01	.00	.01	.00	.00	.00	.09	.00	.00	.00	.00	.00	.00	.00	.01	.29	.58	.01	.00	.00	.00	.00
v	.00	.00	.01	.07	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.00	.00	.00	.05	.11	.72	.00	.00	.01	.00
w	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.00	.00	.00	.00	.00	.00	.97	.00	.00	.00
y	.00	.00	.03	.00	.08	.00	.02	.00	.00	.00	.02	.03	.00	.00	.00	.00	.01	.01	.02	.00	.00	.76	.01	.00
z	.00	.00	.11	.00	.01	.00	.05	.00	.03	.00	.00	.00	.00	.00	.03	.00	.05	.02	.04	.02	.00	.02	.49	.10
ʒ	.00	.07	.00	.00	.00	.00	.18	.00	.00	.00	.00	.00	.00	.00	.00	.04	.00	.00	.00	.00	.00	.01	.01	.68

AV Phi-square consonant confusion matrix

Target Phoneme \ Response Choice	b	tʃ	d	f	g	h	dʒ	k	l	m	n	ŋ	p	r	s	ʃ	t	θ	ð	v	w	y	z	ʒ
b	1	.03	.04	.32	.02	.05	.04	.03	.15	.27	.02	.04	.05	.14	.16	.04	.11	.24	.19	.41	.05	.03	.11	.03
tʃ	.03	1	.05	.04	.03	.03	.10	.05	.03	.03	.04	.03	.04	.02	.06	.19	.06	.06	.04	.05	.01	.05	.05	.11
d	.04	.05	1	.07	.19	.04	.24	.05	.04	.06	.11	.09	.04	.07	.18	.09	.10	.16	.50	.11	.02	.18	.32	.10
f	.32	.04	.07	1	.06	.13	.06	.07	.18	.28	.07	.13	.14	.15	.30	.08	.26	.49	.21	.32	.07	.08	.17	.05
g	.02	.03	.19	.06	1	.03	.13	.04	.06	.06	.10	.12	.03	.10	.11	.06	.07	.12	.21	.10	.04	.25	.18	.08
h	.05	.03	.04	.13	.03	1	.03	.11	.04	.06	.05	.04	.22	.05	.10	.04	.19	.15	.05	.06	.02	.03	.04	.02
dʒ	.04	.10	.24	.06	.13	.03	1	.04	.06	.05	.10	.09	.03	.07	.12	.22	.09	.14	.22	.12	.04	.27	.24	.48
k	.03	.05	.05	.07	.04	.11	.04	1	.02	.04	.09	.04	.18	.04	.09	.05	.24	.10	.05	.05	.01	.04	.04	.03
l	.15	.03	.04	.18	.06	.04	.06	.02	1	.26	.04	.19	.04	.31	.10	.03	.08	.16	.12	.26	.27	.25	.14	.04
m	.27	.03	.06	.28	.06	.06	.05	.04	.26	1	.08	.24	.06	.19	.15	.04	.13	.24	.18	.30	.09	.09	.15	.04
n	.02	.04	.11	.07	.10	.05	.10	.09	.04	.08	1	.35	.07	.06	.08	.06	.16	.13	.12	.08	.02	.11	.12	.07
ŋ	.04	.03	.09	.13	.12	.04	.09	.04	.19	.24	.35	1	.04	.18	.10	.05	.09	.16	.13	.16	.09	.17	.15	.07
p	.05	.04	.04	.14	.03	.22	.03	.18	.04	.06	.07	.04	1	.05	.10	.03	.30	.14	.06	.07	.02	.03	.04	.02
r	.14	.02	.07	.15	.10	.05	.07	.04	.31	.19	.06	.18	.05	1	.14	.04	.09	.18	.18	.32	.16	.24	.16	.04
s	.16	.06	.18	.30	.11	.10	.12	.09	.10	.15	.08	.10	.10	.14	1	.13	.21	.52	.38	.26	.04	.11	.44	.11
ʃ	.04	.19	.09	.08	.06	.04	.22	.05	.03	.04	.06	.05	.03	.04	.13	1	.07	.12	.09	.09	.01	.09	.11	.29
t	.11	.06	.10	.26	.07	.19	.09	.24	.08	.13	.16	.09	.30	.09	.21	.07	1	.31	.17	.16	.03	.08	.14	.06
θ	.24	.06	.16	.49	.12	.15	.14	.10	.16	.24	.13	.16	.14	.18	.52	.12	.31	1	.39	.36	.07	.14	.35	.10
ð	.19	.04	.50	.21	.21	.05	.22	.05	.12	.18	.12	.13	.06	.18	.38	.09	.17	.39	1	.32	.05	.20	.57	.10
v	.41	.05	.11	.32	.10	.06	.12	.05	.26	.30	.08	.16	.07	.32	.26	.09	.16	.36	.32	1	.12	.18	.26	.09
w	.05	.01	.02	.07	.04	.02	.04	.01	.27	.09	.02	.09	.02	.16	.04	.01	.03	.07	.05	.12	1	.18	.08	.02
y	.03	.05	.18	.08	.25	.03	.27	.04	.25	.09	.11	.17	.03	.24	.11	.09	.08	.14	.20	.18	.18	1	.24	.16
z	.11	.05	.32	.17	.18	.04	.24	.04	.14	.15	.12	.15	.04	.16	.44	.11	.14	.35	.57	.26	.08	.24	1	.16
ʒ	.03	.11	.10	.05	.08	.02	.48	.03	.04	.04	.07	.07	.02	.04	.11	.29	.06	.10	.10	.09	.02	.16	.16	1

A -only % accuracy vowel confusion matrix

Target Phoneme	Response Choice: æ	ɑ	e	ɛ	i	ɝ	Ĩ	ɪ	œ	o	u	Ă	ʊ	ʌ
æ	0.68	0.01	0.02	0.06	0.01	0.01	0.01	0.04	0.01	0.02	0.01	0.10	0.01	0.01
ɑ	0.06	0.62	0.01	0.03	0.06	0.01	0.03	0.05	0.04	0.02	0.01	0.02	0.01	0.03
e	0.05	0.03	0.46	0.05	0.09	0.01	0.01	0.15	0.03	0.04	0.02	0.01	0.03	0.02
ɛ	0.06	0.01	0.03	0.66	0.06	0.02	0.01	0.07	0.01	0.01	0.02	0.01	0.01	0.01
i	0.03	0.02	0.21	0.05	0.20	0.03	0.02	0.13	0.04	0.04	0.07	0.07	0.04	0.03
ɝ	0.03	0.04	0.02	0.02	0.03	0.52	0.01	0.04	0.04	0.09	0.03	0.02	0.07	0.04
Ĩ	0.06	0.02	0.01	0.02	0.02	0.01	0.56	0.05	0.02	0.01	0.02	0.17	0.02	0.01
ɪ	0.02	0.01	0.05	0.03	0.09	0.06	0.03	0.50	0.03	0.05	0.02	0.04	0.02	0.04
œ	0.04	0.01	0.06	0.03	0.09	0.02	0.06	0.07	0.47	0.03	0.02	0.03	0.03	0.03
o	0.04	0.06	0.02	0.02	0.03	0.03	0.03	0.07	0.04	0.49	0.03	0.03	0.06	0.06
u	0.03	0.03	0.07	0.05	0.13	0.05	0.03	0.19	0.04	0.09	0.15	0.02	0.06	0.07
Ă	0.04	0.01	0.00	0.01	0.01	0.01	0.04	0.04	0.02	0.01	0.01	0.79	0.01	0.01
ʊ	0.03	0.09	0.02	0.02	0.04	0.11	0.03	0.07	0.09	0.14	0.06	0.03	0.22	0.05
ʌ	0.08	0.10	0.01	0.05	0.03	0.06	0.03	0.05	0.03	0.10	0.04	0.06	0.04	0.32

A -only Phi-square value vowel confusion matrix

Response Choice

Target Phoneme	æ	ɑ	e	ɛ	i	ɝ	Ĩ	ɪ	œ	o	u	Ă	ʊ	ʌ
æ	1.00	0.21	0.22	0.24	0.25	0.19	0.26	0.22	0.21	0.21	0.23	0.23	0.21	0.29
ɑ	0.21	1.00	0.27	0.22	0.31	0.26	0.23	0.27	0.28	0.29	0.32	0.17	0.35	0.38
e	0.22	0.27	1.00	0.30	0.64	0.27	0.21	0.44	0.36	0.31	0.49	0.15	0.34	0.31
ɛ	0.24	0.22	0.30	1.00	0.33	0.21	0.19	0.29	0.26	0.22	0.33	0.14	0.24	0.26
i	0.25	0.31	0.64	0.33	1.00	0.35	0.30	0.55	0.46	0.38	0.69	0.23	0.46	0.41
ɝ	0.19	0.26	0.27	0.21	0.35	1.00	0.21	0.34	0.30	0.39	0.42	0.16	0.55	0.42
Ĩ	0.26	0.23	0.21	0.19	0.30	0.21	1.00	0.26	0.28	0.24	0.28	0.34	0.26	0.30
ɪ	0.22	0.27	0.44	0.29	0.55	0.34	0.26	1.00	0.39	0.37	0.62	0.19	0.42	0.38
œ	0.21	0.28	0.36	0.26	0.46	0.30	0.28	0.39	1.00	0.34	0.46	0.19	0.43	0.35
o	0.21	0.29	0.31	0.22	0.38	0.39	0.24	0.37	0.34	1.00	0.48	0.18	0.56	0.49
u	0.23	0.32	0.49	0.33	0.69	0.42	0.28	0.62	0.46	0.48	1.00	0.19	0.58	0.50
Ă	0.23	0.17	0.15	0.14	0.23	0.16	0.34	0.19	0.19	0.18	0.19	1.00	0.19	0.23
ʊ	0.21	0.35	0.34	0.24	0.46	0.55	0.26	0.42	0.43	0.56	0.58	0.19	1.00	0.54
ʌ	0.29	0.38	0.31	0.26	0.41	0.42	0.30	0.38	0.35	0.49	0.50	0.23	0.54	1.00

V -only % accuracy vowel confusion matrix

Response Choice

Target Phoneme	æ	ɑ	e	ɛ	İ	ɝ	Ĩ	ɪ	œ	o	u	Ă	ʊ	ʌ
æ	0.57	0.02	0.12	0.15	0.01	0.00	0.05	0.06	0.00	0.00	0.00	0.02	0.00	0.00
ɑ	0.02	0.58	0.01	0.00	0.00	0.07	0.00	0.02	0.05	0.04	0.00	0.01	0.08	0.10
e	0.26	0.01	0.38	0.23	0.01	0.00	0.03	0.06	0.00	0.00	0.00	0.00	0.00	0.00
ɛ	0.34	0.01	0.12	0.31	0.02	0.00	0.05	0.13	0.00	0.00	0.00	0.00	0.00	0.01
İ	0.02	0.00	0.04	0.03	0.56	0.00	0.01	0.29	0.01	0.00	0.00	0.01	0.00	0.00
ɝ	0.02	0.01	0.00	0.00	0.00	0.58	0.00	0.02	0.02	0.02	0.06	0.00	0.25	0.03
Ĩ	0.33	0.04	0.05	0.10	0.01	0.00	0.40	0.05	0.00	0.00	0.00	0.01	0.00	0.02
ɪ	0.04	0.00	0.05	0.10	0.33	0.00	0.01	0.45	0.00	0.00	0.00	0.00	0.00	0.01
œ	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.64	0.11	0.17	0.00	0.03	0.00
o	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.08	0.69	0.13	0.03	0.03	0.00
u	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.04	0.04	0.82	0.01	0.05	0.00
Ă	0.03	0.00	0.01	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.01	0.91	0.00	0.00
ʊ	0.02	0.01	0.00	0.01	0.02	0.23	0.00	0.05	0.04	0.01	0.11	0.01	0.43	0.06
ʌ	0.05	0.27	0.02	0.03	0.03	0.01	0.06	0.09	0.00	0.01	0.00	0.01	0.03	0.39

V -only Phi-square value vowel confusion matrix

Target Phoneme / Response Choice

	æ	ɑ	e	ɛ	İ	ɝ	Ĩ	ɪ	œ	o	u	Ă	ʊ	ʌ
æ	1.00	0.09	0.62	0.70	0.17	0.05	0.55	0.23	0.05	0.05	0.04	0.09	0.08	0.20
ɑ	0.09	1.00	0.07	0.09	0.08	0.24	0.12	0.07	0.15	0.14	0.12	0.06	0.28	0.47
e	0.62	0.07	1.00	0.69	0.19	0.05	0.43	0.26	0.05	0.04	0.03	0.07	0.08	0.19
ɛ	0.70	0.09	0.69	1.00	0.24	0.06	0.52	0.34	0.05	0.04	0.04	0.07	0.10	0.23
İ	0.17	0.08	0.19	0.24	1.00	0.06	0.16	0.72	0.05	0.06	0.04	0.06	0.12	0.20
ɝ	0.05	0.24	0.05	0.06	0.06	1.00	0.06	0.05	0.14	0.14	0.17	0.04	0.62	0.12
Ĩ	0.55	0.12	0.43	0.52	0.16	0.06	1.00	0.21	0.06	0.05	0.04	0.08	0.09	0.26
ɪ	0.23	0.07	0.26	0.34	0.72	0.05	0.21	1.00	0.04	0.04	0.03	0.06	0.11	0.21
œ	0.05	0.15	0.05	0.05	0.05	0.14	0.06	0.04	1.00	0.33	0.29	0.05	0.20	0.08
o	0.05	0.14	0.04	0.04	0.06	0.14	0.05	0.04	0.33	1.00	0.25	0.07	0.18	0.07
u	0.04	0.12	0.03	0.04	0.04	0.17	0.04	0.03	0.29	0.25	1.00	0.05	0.24	0.07
Ă	0.09	0.06	0.07	0.07	0.06	0.04	0.08	0.06	0.05	0.07	0.05	1.00	0.06	0.07
ʊ	0.08	0.28	0.08	0.10	0.12	0.62	0.09	0.11	0.20	0.18	0.24	0.06	1.00	0.20
ʌ	0.20	0.47	0.19	0.23	0.20	0.12	0.26	0.21	0.08	0.07	0.07	0.07	0.20	1.00

AV % accuracy vowel confusion matrix

Target Phoneme	Response Choice: æ	ɑ	e	ɛ	i	ɝ	Ĩ	ɪ	œ	o	u	Ă	ʊ	ʌ
æ	0.87	0.00	0.01	0.06	0.00	0.00	0.03	0.02	0.00	0.00	0.00	0.00	0.00	0.00
ɑ	0.01	0.94	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01
e	0.07	0.01	0.68	0.10	0.07	0.00	0.01	0.07	0.00	0.00	0.00	0.00	0.00	0.00
ɛ	0.12	0.00	0.08	0.74	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00
i	0.01	0.00	0.01	0.01	0.84	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00
ɝ	0.02	0.00	0.00	0.00	0.00	0.65	0.00	0.02	0.00	0.01	0.01	0.00	0.25	0.03
Ĩ	0.15	0.01	0.01	0.01	0.00	0.00	0.78	0.03	0.00	0.00	0.00	0.00	0.00	0.01
ɪ	0.03	0.01	0.15	0.05	0.21	0.00	0.01	0.52	0.00	0.00	0.00	0.00	0.00	0.01
œ	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.02	0.80	0.05	0.03	0.00	0.05	0.01
o	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.87	0.05	0.01	0.01	0.00
u	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.04	0.06	0.75	0.00	0.10	0.01
Ă	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.93	0.00	0.00
ʊ	0.01	0.00	0.00	0.00	0.01	0.10	0.00	0.03	0.05	0.01	0.02	0.00	0.73	0.03
ʌ	0.05	0.18	0.02	0.07	0.01	0.00	0.11	0.05	0.00	0.00	0.00	0.00	0.02	0.46

AV Phi-square value vowel confusion matrix

Response Choice (columns); Target Phoneme (rows)

	æ	ɑ	e	ɛ	i	ɝ	Ĩ	ɪ	œ	o	u	Ă	ʊ	ʌ
æ	1.00	0.04	0.15	0.22	0.04	0.04	0.21	0.11	0.05	0.02	0.03	0.05	0.03	0.15
ɑ	0.04	1.00	0.05	0.04	0.03	0.03	0.04	0.05	0.04	0.03	0.03	0.03	0.03	0.22
e	0.15	0.05	1.00	0.27	0.15	0.04	0.12	0.37	0.06	0.03	0.04	0.04	0.05	0.17
ɛ	0.22	0.04	0.27	1.00	0.07	0.04	0.12	0.19	0.05	0.03	0.03	0.04	0.04	0.17
i	0.04	0.03	0.15	0.07	1.00	0.03	0.05	0.37	0.04	0.02	0.03	0.02	0.05	0.09
ɝ	0.04	0.03	0.04	0.04	0.03	1.00	0.04	0.05	0.09	0.06	0.13	0.04	0.41	0.09
Ĩ	0.21	0.04	0.12	0.12	0.05	0.04	1.00	0.11	0.05	0.03	0.04	0.05	0.04	0.21
ɪ	0.11	0.05	0.37	0.19	0.37	0.05	0.11	1.00	0.07	0.03	0.05	0.04	0.07	0.18
œ	0.05	0.04	0.06	0.05	0.04	0.09	0.05	0.07	1.00	0.12	0.16	0.04	0.16	0.08
o	0.02	0.03	0.03	0.03	0.02	0.06	0.03	0.03	0.12	1.00	0.16	0.04	0.08	0.04
u	0.03	0.03	0.04	0.03	0.03	0.13	0.04	0.05	0.16	0.16	1.00	0.03	0.19	0.06
Ă	0.05	0.03	0.04	0.04	0.02	0.04	0.05	0.04	0.04	0.04	0.03	1.00	0.03	0.05
ʊ	0.03	0.03	0.05	0.04	0.05	0.41	0.04	0.07	0.16	0.08	0.19	0.03	1.00	0.10
ʌ	1.00	0.04	0.15	0.22	0.04	0.04	0.21	0.11	0.05	0.02	0.03	0.05	0.03	0.15