DIGITAL COMMONS
@ UNIVERSITY OF SOUTH FLORIDA

University of South Florida
Digital Commons @ University of South Florida

USF Tampa Graduate Theses and Dissertations
USF Graduate Theses and Dissertations

7-18-2008

Factors Affecting Message Intelligibility of Cued Speech Transliterators

Katherine Pelley
University of South Florida

Follow this and additional works at: https://digitalcommons.usf.edu/etd

Part of the American Studies Commons

Scholar Commons Citation
Pelley, Katherine, "Factors Affecting Message Intelligibility of Cued Speech Transliterators" (2008). *USF Tampa Graduate Theses and Dissertations.*
https://digitalcommons.usf.edu/etd/446

This Thesis is brought to you for free and open access by the USF Graduate Theses and Dissertations at Digital Commons @ University of South Florida. It has been accepted for inclusion in USF Tampa Graduate Theses and Dissertations by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact digitalcommons@usf.edu.

Factors Affecting Message Intelligibility of Cued Speech Transliterators

by

Katherine Pelley

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science
Department of Communication Sciences and Disorders
College of Arts and Sciences
University of South Florida

Major Professor: Jean Krause, Ph.D.
Catherine Rogers, Ph.D.
Kelly Crain, Ph.D.
Patricia Blake-Rahter, Ph.D.

Date of Approval:
July 18, 2008

Keywords: Interpreter Skills, Deaf Education, Visual Communication, Lecture Reception, Accuracy, Lag Time, Experience, Rate

© Copyright 2008, Katherine Pelley

Acknowledgments

I would like to express my deepest gratitude to my committee chair, Professor Jean Krause for her persistent help and guidance on this thesis.

Thank you to each of the people whose work contributed to this research project: to Morgan Tessler for the completion of accuracy measurement guidelines manual and for accuracy measurement and analyses, to Kendall Tope for contributions to stimulus preparation and for conducting hearing subject experiment and analysis, to Danielle Milanese for volunteering her time and transliteration skills for practice items, to Dana Husaim and Jessica Lindsay for their contributions to accuracy measurements and analyses, to Jane Smart for her contribution to CST video clip editing and lag time measurements, and to Wendy Park for completion of lag time measurements.

Additionally, I would like to thank my committee members, Professor Catherine Rogers, Professor Kelly Crain, and Professor Patricia Blake-Rahter for their contributions to this project.

Table of Contents

List of Tables

List of Figures

Factors Affecting Message Intelligibility of Cued Speech Transliterators

Katherine Pelley

ABSTRACT

While a majority of deaf students mainstreamed in public schools rely on interpreters, little research has investigated interpreter skills and no research to date has focused on interpreter intelligibility (Kluwin & Stewart, 2001). This thesis is the second in a series of experiments designed to quantify the contribution of various factors affecting the intelligibility of interpreters (transliterators) who use English-based communication modes. In the first experiment, 12 Cued Speech transliterators were asked to transliterate an audio lecture. Two aspects of these transliterated performances were then analyzed: 1) accuracy, as measured as the percent-correct cues produced, and 2) lag time, the average delay between lecture and transliterated message. For this thesis, eight expert receivers of Cued Speech were presented with visual stimuli from the transliterated messages and asked to transcribe the stimuli. Intelligibility was measured as the percentage of words correctly received. Results show a positive nonlinear relationship exits between transliterator accuracy and message intelligibility. Intelligibility improved with accuracy at the same rate for both novice and veteran transliterators, but receiver task difficulty was less for stimuli produced by veterans than novices (as evidenced by a left shift in the psychometric function for veterans compared to novices). No large effects of lag time were found in the accuracy-intelligibility relationship, but an “optimal lag time” range was noted from 1 to 1.5 seconds, for which intelligibility scores were higher overall. Intelligibility scores were generally higher than accuracy, but not all transliterators followed the same accuracy-intelligibility pattern due to other sources of variability. Possible sources of transliterator variability included rate of cueing, visible speech clarity, facial expression, timing (to show syllable stress or word emphasis), cueing mechanics, and mouth-cue synchronization. Further research is needed to determine the impact these factors have on intelligibility so that future transliterator training and certification can focus on all factors necessary to ensure highly intelligible Cued Speech transliterators.

Chapter One

Introduction

In 1975, Public Law 94-142, the Education of All Handicapped Children Act, now referred to as the Individuals with Disabilities Education Act (IDEA), was passed to prevent educational discrimination against children with disabilities (Marschark, Sapere, Convertino, & Seewagen, 2005). The passing of IDEA caused a dramatic increase in the number deaf and hard of hearing children being mainstreamed in public schools (Schick, Williams, & Bolster, 1999). While 80% of deaf students attended special schools in the 1950’s and 1960’s, today, 80% of deaf students attend mainstream public schools (Marschark, et al., 2005). Many of these deaf and hard of hearing children rely on interpreters to gain access to the classroom and communicate with those around them. Yet limited quantitative research presently exists to verify the efficacy of educational interpreters for the deaf. Educational interpreters are known to exhibit a wide range of skill levels (Schick, Williams, and Kupermintz, 2006), but Marschark and colleagues (2005) observe that “even in optimal learning conditions, we know very little about what students can learn through an interpreter,” (p.39). Similarly, Kluwin and Stewart (2001) emphasize that there is “no empirical evidence” to indicate how well deaf students understand their interpreters (p. 15). In order to gain insight into these issues, this study is the beginning of a larger project that aims to quantitatively determine the access interpreters are providing deaf students. In particular, this study focuses on interpreters who use a communication option known as Cued Speech (CS).

Communication Options

In order to evaluate the level of access provided by interpreters, one important factor to consider is the mode of communication used by the interpreter. Deaf students who rely on educational interpreters in the United States currently use various means of communication, including, but not limited to, American Sign Language, Manually Coded English, and Cued Speech.

American Sign Language (ASL). American Sign Language is a complete language, separate and independent from English, with its own vocabulary and syntax (Scheetz, 2001). It is composed of signs (hand-movement configurations), which convey general meaning, produced with non-manual signals (facial expressions and body postures) that function grammatically and/or refine the meaning of the signs in the sentence.

Manually Coded English (MCE). The term Manually Coded English refers to sign systems which integrate components of both ASL and English, generally using or modifying signs from ASL in conjunction with structural elements of English (such as English word order and/or spoken English mouth movements). A variety of MCE systems exist either by invention, with the aim of bridging the gap between the two languages (and improving literacy in deaf children), or as a natural result of the interaction between deaf signers and hearing English speakers (Scheetz, 2001). Each MCE system varies with regard to the degree of structure it has and how closely it follows English.

Signing Exact English (SEE II; Gustason, 1990) is an example of a very structured MCE system that incorporates English elements to a high degree. In SEE II, English word order is used, and initialized signs differentiate between English words which would be represented by a single sign in ASL. In this system, any English words that share at least two of the following characteristics are represented by the same sign: a common pronunciation, a common spelling, and/or a common meaning (Gustason, 1990). Under this system, the words “bat” (as in baseball) and “bat” (as in the animal) would share the same sign because they share a common spelling and pronunciation; however, the words “rain” and “reign” would be signed differently because they share only a common pronunciation. SEE II also incorporates invented markers, created specifically to show English morphology (word endings, tenses, and affixes), and additional invented signs (beyond those available in ASL), created to show English words which could not be represented by a single sign.

At the other end of the MCE spectrum is Conceptually Accurate Signed English (CASE), a less-structured system that is less tied to English word order and grammar. Because it resembles a Pidgin (natural mixing of two languages), CASE is sometimes referred to as Pidgin Signed English. In CASE, all signs and non-manual markers are borrowed from ASL for practicality and efficiency (i.e. no invented signs are used), but English word order and mouthing of English words is still used wherever possible to provide English structure (Winston, 1989). Although the main focus is to convey English, users have more flexibility in sign choice (Winston, 1989).

Cued Speech. Cued Speech is a system for conveying a traditionally spoken language, such as English, in a fully visually accessible form (Cornett, 1967). It combines the mouth movements of speech (with or without voice) with a system of handshapes at specified placements around the mouth, designed to clarify the elements of speech that would be ambiguous through speechreading alone (“speechreading” is another term for “lip reading” but implies watching not only the lips, but all the visible articulatory structures of speech). Handshapes represent consonant sounds and placements represent vowel sounds. Each handshape-placement combination, or cue, is produced in synchrony with the mouth movements for the corresponding consonant-vowel combination. Each of the speech sounds (or phonemes) is thus made visually distinct through cues.

Cued Speech Research

Deaf individuals and/or their parents decide to use Cued Speech as a mode of communication for a variety of reasons. Because Cued Speech is a closed system, some may find it relatively easy to learn or find it appealing that once the system has been mastered, it can be used to express any utterance that can be spoken, including unfamiliar vocabulary and foreign languages. In other instances, parents become interested in CS when they learn that it has been shown to be an effective tool of communication: children who are experienced users of Cued Speech (i.e., those with at least four years of Cued Speech experience) have near-perfect reception of cued materials, with averages for 18 profoundly deaf children ranging from 95% to 96.6% accuracy in the reception of both high and low probability key words (Nicholls & Ling, 1982). Moreover, such benefits in the reception of English via Cued Speech over speechreading alone begin to appear in deaf and hard of hearing children within the first 1-2 years of using Cued Speech (Clark & Ling, 1976).

Not only does CS improve the efficacy of communication over speechreading alone, but a growing body of research indicates that its use also can facilitate language and literacy development in deaf children, especially when started early in life (Leybaert & Charlier, 2003). For example, Leybaert and Charlier (1996) reported that deaf children who used Cued Speech for communication had phonological skills similar to those of hearing children for tasks that involved rhyming, remembering, and spelling words. In the rhyming task, the following six groups of children were tested on their ability to detect rhymes using pictures (rather than written words, to avoid the influence of spelling): 12 hearing children (mean age 8;7 years), 16 children (mean age 10;1 years) who learned Cued Speech early, 18 children (mean age 12;7 years) who learned Cued Speech late, 12 children (mean age 13;3 years) relying on the oral method, 12 children (mean age 10;4 years) who learned sign language early, and 20 children (mean age 10;1) who learned sign language late. The children exposed to CS from an early age (three years old or younger) were able to detect pairs of rhyming words, whether or not the words were orthographically similar, with roughly the same level of accuracy (orthographically similar: 97.4%; orthographically different: 94.4%) as the hearing children (orthographically similar: 95.8%; orthographically different: 97.0%), outperforming all the other groups of deaf children. The pattern of performance between groups was similar for tasks related to remembering and spelling words. In addition, there is evidence that this pattern of performance may extend to reading skills as well.

While the average reading comprehension score for a deaf or hard-of-hearing eighteen year old is below the fourth grade reading level (Traxler, 2000), Wandel (1989) found that students with profound hearing loss using Cued Speech as their primary communication mode scored similarly to hearing students on the reading comprehension portion of the Stanford Achievement Test across all ages tested (7 to 16 years old).

Given these data, LaSasso and Metzger (1998) suggest that there are theoretical advantages to a communication philosophy that incorporates the use of both American Sign Language and Cued Speech, as the child enters school. They argue that the language of the child's home should be his/her first language (i.e. children with hearing parents should learn English as their first language through Cued Speech and children with deaf parents should learn ASL as their first language) in order to take full advantage of using parents as a language model. LaSasso and Metzger emphasize that this philosophy would allow educators to take advantage of the benefits of both ASL and Cued Speech when teaching different subjects in school (for example, using Cued Speech when working on written English). Although this philosophy has proven attractive to some parents who choose to incorporate Cued Speech as one of their child's communication options, few schools to date have adopted this new bilingual/bicultural philosophy using ASL and Cued Speech. Therefore, for most parents, choosing Cued Speech as a communication option will mean mainstreaming their child with an interpreter who uses Cued Speech, known as a Cued Speech transliterator.

Interpreting vs. Transliterating

Depending on the communication option used by the deaf child, the terminology pertaining to an interpreting professional can vary. *Interpreter* is the term typically used to describe a professional who facilitates communication between two different languages (for example, between English and the separate language of American Sign Language). When referring to interpreters who use Cued Speech, the term "transliterator" is typically used. A *transliterator* is someone who facilitates communication between different modes of the same language (for example, between spoken English and written English or between spoken English and visual English through Cued Speech).

The larger study (Krause, 2006), in which this thesis is a part, will focus on the efficacy of *transliterators* because even though many deaf adults prefer to use ASL interpreters, transliterators are much more commonly used in educational settings. According to a survey by Jones, Clark, and Soltz (1997), more than 95% of educational interpreters use an English-based sign system, rather than American Sign Language. Of the 222 educational interpreters employed in Kansas, Missouri, and Nebraska who responded to the survey, 55.7% reported using Pidgin Signed English (PSE) or Conceptually Accurate Signed English (CASE) in their jobs. Another 32.7% reported using Signing Exact English (SEE II), while less than 5% of interpreters surveyed reported using pure ASL interpreting as educational interpreters. Although the exact number of Cued Speech transliterators is unknown, Cued Speech is an attractive candidate for initial study because there is a definitive mapping from the spoken message to the cued message (and vice versa). This one-to-one correspondence of spoken English phonemes and cued phonemes allows for extremely straightforward assessment of the cues produced by a transliterator (relative to the original spoken message) and how those cues affect the message received by deaf students. Therefore, the present study will focus specifically on assessing these relationships for transliterators who use the communication option of Cued Speech.

Assessment of Interpreters and Transliterators

There is a great deal of research that needs to be done in the field of interpreting to ensure that deaf and hard of hearing students in the mainstream setting are receiving adequate access to classroom communication from the interpreters and transliterators on which they rely. This level of access depends primarily on *intelligibility*, or the percentage of the interpreted/transliterated message correctly received by the deaf receiver. Although research directly assessing the intelligibility of interpreters and transliterators is lacking, other methods of assessment have been used to evaluate the quality of interpreting services. Strong and Rudser (1985), for example, created one of the first objective rating scales to assess the accuracy of sign language interpreters. They assessed interpreter performances by analyzing each proposition (defined as a unit of text carrying a single semantic idea) interpreted, rather than attempting to analyze the interpreter's entire performance with a single rating. Although some measure of subjectivity was necessary in order to include considerations of cultural adjustment (a trickier, more subjective decision-making component of this assessment scale), this rating method proved highly reliable, with Pearson r correlation coefficients for pairs of scores ranging from .9749 to .9985 (Strong & Rudser, 1985). Moreover, this assessment method was also shown to be much more reliable than the shorter, more subjective measure for assessing interpreter skills that Strong & Rudser (1986) later developed, for which Pearson r correlation coefficients ranged from only .52 to .86.

More recently, an assessment tool that has gained widespread acceptance is the Educational Interpreter Performance Assessment, or EIPA (Schick, Williams, & Bolster, 1999). The EIPA is an evaluation tool designed specifically to assess and certify educational interpreters in a classroom setting. It evaluates the voice-to-sign and sign-to-voice skills of interpreters who use ASL, MCE, or Pidgin Signed English (Schick & Williams, 2001). Skills relating to grammar, prosody, sign vocabulary, fingerspelling, and other behaviors "critical to competent interpreting" are rated on a Likert scale from zero to five (zero being "no skills," five being "advanced") (Schick, Williams, & Kupermintz, 2006). Rating is completed by an evaluation team of three people, at least one of whom is proficient in the specific sign system used by the interpreter (Schick & Williams, 2001).

As a measurement tool, the EIPA has been shown to have good reliability, with correlations between rating teams ranging from 0.86 to 0.94 across the domains of evaluation. Coefficients of internal consistency of skills within each domain of the assessment are also high (ranging from 0.93 to 0.98), while interdomain correlations used to assess validity suggest that each domain taps a different aspect of an interpreter's performance. As further evidence of validity, 42 interpreters with RID certification averaged a score of 4.2 (SD = .06) on the EIPA (Schick et al., 2006). Therefore, individuals with RID certification can be expected to score in the Advanced range (4.0 or better) on the EIPA.

Because of its high validity and reliability, the EIPA has become an important research tool in assessing the quality of interpreters in educational settings. For example, results of EIPA testing have provided important research data which showed that out of the 2,091 sign language interpreters assessed on the EIPA from 2002 through 2004, only 38% were able to meet the minimum proficiency level of 3.5 required by most states (Schick, et al., 2006). Data reported by EIPA testing has also been instrumental in identifying skill areas that need improvement (e.g. sign-to-voice skills of interpreters working with younger children). Furthermore, future EIPA results could be used to monitor changes in the quality of educational interpreters over time. As interpreter training programs are modified to better address these skill areas (and others yet to be identified), the EIPA can thus serve as an evidence-based mechanism for assessing the efficacy of interpreter training efforts.

While such research provides highly valuable assessment information regarding aspects of interpreter performance, the quantitative relationship between those aspects (e.g. accuracy) and intelligibility remains unknown. Determining this quantitative relationship is important for ensuring that quality control standards for interpreters and transliterators are appropriate. As Schick et al. (2006) point out, even though a majority of educational interpreters are unable to meet the minimum standards of 3.5, it is still unclear as to whether or not this minimum level is high enough to ensure access to basic classroom content (Schick et al., 2006).

Even if the 3.5 standard is adequate, it would be difficult to analyze how the individual skills assessed in the EIPA affect intelligibility. Although each of the subskills (i.e. “appropriate eye contact/movement,” “developed a sense of the whole message” and “stress/emphasis of important words or phrases”) scored is averaged equally to obtain an overall EIPA score, it is unknown whether or not each of the subskills represents aspects of equal importance for receiver intelligibility. Also, while the Likert scale is well-chosen for the purposes of the EIPA, it does not provide sufficient resolution for determining how any particular skill or subskill (e.g., “stress/emphasis of important words or phrases”) affects an interpreter’s intelligibility. Because any particular rating on a scale of 0 to 5 will represent some variation in ability, the differences in ability between two interpreters who share the same score (for example, 3.5) on a given skill or subskill could still be great enough to affect intelligibility substantially. Therefore, the skill must be measured using a method that produces more resolution so that more levels of skills can be represented. With such a measure, the relationship between the skill score and the intelligibility of the message can then be analyzed empirically.

Accuracy

One skill that is very likely to affect intelligibility and can be measured with sufficient resolution for such an analysis is *accuracy*, or the percentage of the message correctly produced by the interpreter/transliterator. Although there is no known quantitative research regarding the accuracy of most types of interpreters, there is information available regarding the accuracy of some Cued Speech transliterators (CSTs). Pelley, Husaim, Tessler, Lindsay, & Krause (2006) analyzed the cue sequence produced by each of six transliterators relative to a target cue sequence (i.e., the correct sequence based on a phonetic transcription of the spoken message) in order to derive percent-correct scores representing each transliterator's accuracy for the two manual components of Cued Speech: handshape and placement. Of six CSTs employed in the educational setting, the average accuracy was 49% (among CSTs of different experience levels averaged over three different rates of presentation), with 33% of the target cues omitted, and 18% produced in error (i.e. substitutions, or incorrect cues). Insertions of cues accounted for an extraneous 6% beyond the expected target cues. At the phrase/sentence level, a wide range of accuracy scores resulted, ranging from near 0% to near 100%.

Such accuracy measurements, however, do not measure directly how accessible each phrase/sentence would be to a deaf receiver. The relationship between transliterator accuracy and intelligibility is currently unknown, and it must be empirically measured because the relationship is not likely to be perfectly linear. It cannot be assumed that transmitting 75% of the message faithfully renders it 75% intelligible. The only way to truly determine whether or not a particular accuracy level (50%, for example) is sufficient to provide an intelligible transliterated rendition of the original message is to determine the psychometric function that relates transliterator accuracy to intelligibility.

Psychometric Functions

A psychometric function is a graphic plot of data points which relates the physical characteristics of a stimulus with an associated psychological percept. The psychological percept (output), which is some aspect of participant performance, is plotted as a function of changes in the physical characteristics of a stimulus (input), such as sound pressure level (in dB). Psychometric functions are usually sigmoidal shaped functions, with two properties that characterize the relationship between the input and output variables: slope and left-right shift. The slope is calculated on the linear portion of the function, which typically occurs between 20% and 80% of the maximum value of the output variable (Wilson & Strouse, 1999). The slope demonstrates how rapidly the psychological percept (plotted along the y-axis) changes with increases in the physical characteristics of the stimulus (plotted along the x-axis). It is highly influenced by inter- and intra-subject variability, with higher variability yielding a flatter slope and lower variability yielding a steeper slope. The left-right shift of a psychometric function is determined by the difficulty of the task (Wilson & Strouse, 1999), with easier tasks yielding psychometric plots at lower x-axis values (farther left) than harder tasks.

Given these properties, psychometric functions are a useful tool to characterize and display the effect of various factors on intelligibility. Many such psychometric functions have been documented previously for various factors influencing speech reception (e.g. Wilson & Strouse, 1999), but no psychometric functions have been obtained for Cued Speech or any other visual communication mode. Psychometric functions for each of the visual communication modes are thus necessary to demonstrate the role of the variables involved in their reception. As research is already available regarding the accuracy of Cued Speech messages produced by transliterators, the goal of this thesis is to determine a psychometric function for Cued Speech transliteration in order to characterize the relationship between the accuracy of the transliterated message (the independent variable plotted along the x-axis) and the message intelligibility (the dependent variable plotted along the y-axis).

In determining the accuracy-intelligibility psychometric function for Cued Speech transliterators, other factors that influence this relationship can also be explored. A number of factors are already known to influence the shapes of psychometric functions (Wilson & Strouse, 1999) including stimulus material (i.e. words vs. sentences), presentation level, type of grading (i.e. key word versus all word), population, and response mode (i.e. writing English responses vs. identifying by pointing). For Cued Speech transliterators, it is expected that additional factors, such as CST experience, lag time, and rate of presentation, will also have similar effects on the shape of the accuracy-intelligibility psychometric function.

Experience. Many factors contribute to transliterator intelligibility and may influence the accuracy-intelligibility psychometric function. One such factor is transliterator experience. Veteran CSTs are generally more accurate than novices, and are therefore likely to be more intelligible on average. In a study of six CSTs who were asked to transliterate materials at three different presentation rates, three of four veteran transliterators averaged 60% correct cues compared with only 45% correct cues found for the two novices (Pelley et al., 2006). As a result, the overall relationship between intelligibility and transliterator experience is expected to be similar to the relationship found by Pelley et al. (2006) between accuracy and experience. As experience increases, intelligibility is expected to be higher, with some exceptions due to differences in individual skill level.

Of more relevance for this study, however, is that Pelley, et al. (2006) found differences in the error patterns between veterans and novices; veterans produced substitution errors most frequently (25%) and fewer omission errors (15%), while novices' errors followed the opposite pattern (50% omissions and only 3% substitutions). Therefore, it is likely that *even when accuracy is controlled*, novices and veterans may vary in intelligibility due to effects of error type. In analyzing accuracy differences between the experience groups, Pelley et al. observed that veterans cued a large majority of the message and cued faster than novices, but often "hypocued," losing form and precision to cope with increasing speed. This technique produced more substitutions but allowed veterans to retain more cues. They omitted mostly cues within words and short sequences. Novices, on the other hand, tended to cue either slowly and highly accurately, with correct form, or to entirely omit large chunks of the message. Thus, when accuracy is controlled, it is likely that veterans will generally have higher intelligibility than novices, in spite of losing form and precision, because more of the message will be at least partially transmitted, allowing the receiver to fill in gaps and correct errors more easily than if a large portion of the message is simply missing (as is the case with the novices). However, novices could be more intelligible than veterans if the veterans' errors are undecipherable to the deaf receiver, which is more likely to be the case as the number of errors increases (i.e. as accuracy levels decrease). The accuracy percentage at which a veteran (with faster cueing ability) becomes so inaccurate that his/her intelligibility is poorer than a novice (who is unable to cue rapidly and therefore omits large portions of the message) is unknown. This study should demonstrate whether omissions or substitutions are more detrimental to intelligibility, by comparing the left-right shift (accuracy level for which intelligibility decreases dramatically) of the psychometric functions for transliterators of different experience levels.

Lag time. Another factor that may influence the accuracy-intelligibility psychometric function is lag time. Lag time is the amount of time between the original source message and the interpreter’s production of that message. Although the relationship between interpreter/transliterator intelligibility and lag time has yet to be established in research, there is some data regarding the relationship between *accuracy* and lag time. Cokely (1986) studied four ASL interpreters (with certifications from Registry of Interpreters for the Deaf) from a national conference and found that the average onset lag times for the two interpreters with higher accuracy was 4 seconds (ranging from 1 to 6 seconds), while two poorer performing interpreters had shorter lag times, averaging 2 seconds (ranging from 1 to 5 seconds). Specifically, the interpreters with longer lag times produced a greater number of sentences, and a greater number of correct sentences, than the interpreters with shorter lag times (Cokely, 1986). Cokely concluded that as lag time increases, the accuracy of ASL interpreters also increases, with an expected ceiling due to working memory (1986). In addition, he hypothesized that interpreter lag time is largely a function of the structural differences between the source language and the target language, that the more different the structures of languages are, the longer the lag time is expected to be. Thus, CS transliterators could be expected to require shorter lag times than ASL interpreters, as a result of the relative similarity between spoken English and cued English (through the phonemic level), in contrast to ASL being an entirely different language than spoken English.

While preliminary evidence does suggest shorter lag times for CS transliterators, it also suggests an inverse relationship between accuracy and lag time for these individuals. Pelley (2006) reported average lag times for one transliterator increased from 1.11 seconds for materials that were produced with 71% accuracy, to 1.23 seconds and 1.36 seconds for materials at 59% and 49% accuracy, respectively. However, the decreases in accuracy were also associated with increases in presentation rate, and it is not yet known which factor (lag time or presentation rate) is primarily responsible for the observed changes in accuracy. Whether the same inverse relationship between accuracy and lag time holds when presentation rate is controlled is not yet known. As a result, the overall relationship between intelligibility and lag time for CS transliterators is difficult to predict. Of more interest for this study, however, is whether differences in lag time are found to affect intelligibility *even when accuracy is controlled*.

Rate of presentation. A transliterator's rate of cueing is yet another factor that could affect intelligibility and the accuracy-intelligibility psychometric function. While no research has been conducted as of yet regarding the effect of cueing rate on the intelligibility of Cued Speech, multiple studies have evaluated the effect of rate on a person's ability to perceive linguistic stimuli in various other communication modes, including American Sign Language, the Rochester Method (or fingerspelling), and spoken English. Regardless of the communication mode, data suggest a production bottleneck for visually communicated sentences. That is, sentences can be received correctly at faster rates than they can be physically produced. Fisher, Delhourne, and Reed (1999), for example, increased the rates of playback of videotaped ASL signs and signed sentences and evaluated the percent correct scores of 14 native ASL viewers. Breakdowns in the ability to process the ASL stimuli did not occur until time compressions of 2.5 to 3 times the normal rate were made. This finding parallels the findings of Beasley, Bratt, and Rintelmann (1980), who researched the auditory reception of time-compressed speech (in sentences) and found near-perfect intelligibility for sentences at time compression factors up to 2.5, with intelligibility decreasing to 82% at a compression factor of 3.3 (the only compression factor tested above 2.5). Given that the average words per minute of sign language and speech are approximately equivalent (although speech is much more rapid, signing requires fewer units; Bellugi & Fisher, 1972), these findings suggest an upper limit to language processing that is independent of language and modality.

Not every communication mode can be used effectively at these rates, however. While the normal speech rate is 4 to 5 syllables per second, fingerspelling is four times slower at 0.5 to 2 syllables per second (Reed, Delhourne, Durlach, & Fisher, 1990). Even so, fingerspelling exhibits a similar production bottleneck. Reed et al. investigated the effects of time compression of the playback of videotaped fingerspelling and found that 6 deaf participants (ages 63 to 87) were able to receive substantial amounts of linguistic information at two to three times the normal rates of fingerspelling (again, rates in excess of what is physically possible to produce). While the average production rate of Cued Speech transliterators is unknown, the physical limitations required for the production of Cued Speech at a rapid rate are likely similar to that of fingerspelling, since both systems consist of individual characters produced by one hand. Cued Speech, however, has fewer individual characters required per word because only one cue is required for each CV phoneme pair (i.e., roughly one cue per syllable), as opposed to the one character per written grapheme required in fingerspelling. Therefore, the average rate of production of Cued Speech is expected to be faster than fingerspelling and similar to, but slower than, speech.

Regardless of its average rate of production, it seems likely that Cued Speech also exhibits a production bottleneck. That is, cued materials are likely to retain intelligibility at up to 2 to 3 times the average rate of production – *when materials are cued with 100% accuracy*. When the cued message contains errors, however, it is unknown whether it can be processed over such a wide range of rates. Increases in rate leave less time to complete the perceptual task of receiving and processing the interpreted/transliterated message: the less time available, the less opportunity for recognizing and correcting errors in production. As a result, it is possible that as rate of presentation increases, the intelligibility of Cued Speech transliterator performances of a given accuracy may decrease.

Other factors. In addition to experience and lag time, a variety of other factors may influence the accuracy-intelligibility psychometric function. Specifically, two interpreters with the same accuracy can have differences in the manner in which they portray a message, making the message more or less clear to the deaf consumer (hence, affecting intelligibility, but not accuracy). These factors include, but are not limited to, speechreadability (how well the transliterator visibly articulates every word, even when omitting cues or hypo-cueing), prosody (how well the transliterator pauses, portrays stress patterns, etc.), and transliterator error patterns. Different error patterns that are likely to affect intelligibility, but not accuracy, include:

1) *whole word omissions versus intra-word omissions*: If part or most of the word is cued, there is more information available to the deaf receiver than if the entire word is missing.
2) *type of word in error*: If some transliterators, especially more experienced transliterators, put greater emphasis on cueing important content words (i.e. nouns and verbs such as “research”) and less emphasis on function words (i.e. “the” and “to”), the psychometric relationship between intelligibility and accuracy of *key words only* would exhibit a right shift of values along the x-axis, in comparison to the psychometric relationship between intelligibility and overall accuracy.
3) *placement versus handshape errors*: It is possible that some placement errors may be less detrimental to intelligibility than handshape errors. For example, when cueing rapidly, veteran transliterators sometimes fail to achieve correct placement but still move toward the placement. In this case, the cue would be classified as a substitution error for placement, but some placement information would still be available to the deaf receiver.

Statement of the problem

This thesis focuses on how the intelligibility of Cued Speech transliterators, as measured by the percentage of transliterated words in meaningful sentences correctly received by deaf consumers, varies with the accuracy of the message produced by the transliterator. It is hypothesized that intelligibility will have a positive relationship with transliterator accuracy and that the psychometric function characterizing this relationship will be nonlinear. The shape of the accuracy-intelligibility psychometric function is expected to be affected by experience, with veterans producing psychometric functions that are less steep (more variable) and farther to the left compared to novices. As described earlier, psychometric functions shift to the left with easier tasks, and receiving transliterated messages from veteran transliterators is expected to be an easier task overall (for a given accuracy level) than receiving messages from novice transliterators. The effect of lag time on the accuracy-intelligibility psychometric function is expected to be minimal because lag time is thought to be directly and inversely correlated with accuracy. Even when experience and lag time are controlled, it is expected that variability due to other factors will affect the psychometric function. For example, it is hypothesized that in some cases two transliterators with the same accuracy will have different intelligibility scores, due to differences in cueing rate, speechreadability, or prosody.

Chapter Two

Method

The main focus of this study is to characterize how the intelligibility of a transliterated message, as measured by the percentage of words correctly perceived by expert Cued Speech receivers, varies with the accuracy level of the cued message produced by the transliterators. In addition, the effect that the professional experience of each transliterator has on this relationship is examined, and the relationship between lag time and intelligibility is also explored. Finally, in situations where two stimulus items were cued with the same accuracy percentages, but did not result in the same intelligibility scores, the possible role of other factors is noted, including such factors as the transliterator's mouth clarity, the rate at which the sentence was cued, the type of accuracy errors produced (e.g., omissions of entire words versus individual cues), the synchronization between cues and mouthshapes, etc.

Two types of intelligibility are measured: *original message* (OM) intelligibility, or the percentage of the original spoken message that is correctly received, and *transliterated message* (TM) intelligibility, or the percentage correctly received of just that portion of the message that was actually cued by the transliterator. OM intelligibility is an overall measure of intelligibility and was used to analyze the accuracy-intelligibility relationship. It captures how much access deaf receivers actually receive from a transliterated lecture. However, the analysis of OM intelligibility alone has limitations. It cannot determine what portion of the unintelligible information can be attributed to omissions made by the transliterator, as opposed to receiver errors, nor can it differentiate between how much information the receivers might have filled in from context versus how much information they were given at least some access to by the transliterator. By analyzing the intelligibility of only those words which the transliterator actually mouthed and cued correctly, or at least partially correctly, TM intelligibility is a better a measure of how much of the transliterator’s message was successfully received.

Participants

Eight expert Cued Speech receivers were recruited to participate in intelligibility tests for this thesis. Although the results of this study will have implications for children who use Cued Speech transliterators in educational settings, children were not used because perception abilities and language skills, even for older children, may still be developing; therefore, only participants 18 years of age or older were included. In addition, all participants were required to be high school graduates; to pass a language screening; and to present with no known visual acuity problems, as participants were required to view video recordings on a computer monitor.

The Expressive Written Vocabulary section of the Test of Adolescent and Adult Language Third Edition (TOAL-3; Hammil, Brown, Larsen, and Wiederholt, 1994) was used to screen for basic proficiency in written English. The Written Vocabulary Section of the TOAL-3 requires examinees to correctly construct a written sentence for each vocabulary word given. In isolation, this section of the TOAL-3 functions as a language screening tool, with normative data provided for normal-hearing children and adults.

Because the normative data does not include deaf and hard of hearing individuals, this screening procedure ensured that all participants, deaf or hearing, possessed an English proficiency level on par with typical high school graduates. Participants were required to score within one standard deviation of age-appropriate averages. All participants who were screened met this requirement (see Appendix A, Table A1).

Expert Cued Speech receivers were recruited via advertisements sent to several regional and national Cued Speech organizations. To qualify as an “expert” Cued Speech receiver, each participant was required to meet the following criteria: 1) introduced to Cued Speech as a communication mode before age 10, 2) used Cued Speech as communication mode receptively (or receptively and expressively) at home (with at least one parent) and at school (through a teacher or transliterator), and 3) used the Cued Speech communication mode for at least 10 years. Additionally, participants were required to pass a visual-only receptive Cued Speech screening. Receptive Cued Speech skills were screened by presenting videos of conversational sentences that were cued with 100% accuracy for participants to view and transcribe. Participants viewed the cued sentences on the computer monitor and were asked to transcribe each sentence word-for-word by writing their responses. Five conversational sentences obtained from a list of Clarke sentences (Magner, 1972) were presented via Cued Speech (no audio), and only participants who correctly transcribed 90% or more of the words in these five Clarke sentences were included in this study. Because these individuals participated in the portion of the experiment that does not incorporate audio information, there were no exclusionary criteria based on hearing level. However, each expert cue receiver was required to complete a hearing and communication background survey, which asked them to classify their hearing levels as normal hearing or as mild, moderate, severe, or profound hearing loss and to provide some basic information about their communication background. All background information collected on participants, including the results from this survey and from the CS screening tests, are summarized in Table A1 of Appendix A.

In a companion study, 12 normal hearing individuals were also employed as control subjects in order to determine the baseline intelligibility of the lecture material used in the experiment (Tope, 2008). Background information on these individuals is summarized in Table A2 of Appendix A.

Materials

The video materials utilized in this study were taken from the videotaped performances of 12 Cued Speech transliterators (CSTs) of varying experience levels who participated in the larger grant study (Krause, 2006). The experience level of each transliterator has been classified for that study as “novice,” “experienced,” or “veteran.” The background information and corresponding experience level classifications for the 12 CSTs is summarized in Table 1. Classifications were based on responses to written questions regarding level of education, relevant certifications, amount of continuing education (in hours per year) and experience (in years) as a Cued Speech transliterator. The experience categories were defined (Krause, 2006) as follows:

1) **novice**: minimal certification or no certification, with work experience of less

than the equivalent of one full-time year

2) **experienced**: minimal certification with less than the equivalent of three full-time years of work experience, or no certification with 3-5 years of experience
3) **veteran**: highest level of certification and/or more than 5 years of experience.

Table 1

Transliterator Experience

	Work Experience as a CST	Hours Per Week Transliterating	Relevant Certifications	Study
CST1	1 year	2 hrs. 20 min.	Certified Edu. CST by State	Husaim & Tessler (2006)
CST2	10 yrs.	35 hrs.	Certified Edu. CST by State	Husaim & Tessler (2006)
CST3	4 yrs.	2 hrs. 15 min.	Certified Edu. CST by State	Husaim & Tessler (2006)
CST4	22 yrs.	35 hrs.	Certified Edu. CST by State	Husaim & Tessler (2006)
CST5	15 yrs.	35 hrs.	Certified Edu. CST by State	Lindsay (2006)
CST6	15 yrs.	30 hrs.	Certified Edu. CST by State	Pelley (2006)
CST7	9 yrs.	30 hrs.	Natl. Certified (TSC)	Tessler (2007)
CST8	5 yrs.	35 hrs.	None	Tessler (2007)
CST9	15 yrs.	32.5 hrs.	None	Tessler (2007)
CST10	18 yrs.	35 hrs.	None	Tessler (2007)
CST11	6 yrs.	5 hrs.	None	Tessler (2007)
CST12	20 yrs.	3 hrs.	Natl. Certified (TSC)	Tessler (2007)

The CSTs transliterated a lecture about plants that was based on a 25-minute educational film entitled *Life Cycle of Plants* (Films for the Humanities, 1989). The original film contained video images with audio narration and was chosen because it was 1) part of the materials used by the University of South Florida's Educational Interpreting Program, 2) designed to be used in a high school setting, and 3) consistent in vocabulary and pacing throughout the audio narration. The audio narration was broken into three sections of roughly equal length, and each section was re-recorded at a normally paced presentation rate (i.e. speaking rate) with deliberate pauses between phrase boundaries. The resulting recordings were slowed by an expansion factor of 1.25 to create a version of the materials at a "slow-conversational" presentation rate and sped up by a compression factor of 0.8 to create a version of the materials at a "fast-conversational" presentation rate. Each transliterator was then presented with the three sections of the lecture at three different conversational presentation rates (one section per presentation rate): slow-conversational, measured at 88 words-per-minute (wpm); normal-conversational, measured at 109 wpm; and fast-conversational, measured at 137 wpm. The transliterations elicited were filmed using a digital video camera and saved to a computer disk for analysis. Although each CST was exposed to every section and every presentation rate, only 4 CSTs transliterated any particular section at a given rate because the materials were counterbalanced across presentation rates. Of the video materials available at each of the three rates, only the transliterations elicited at the slow-conversational rate were used in this study in order to 1) avoid any uncontrolled impact on intelligibility due to varying rates of transliteration, 2) ensure the clips available contained a wide range of accuracy performances (avoiding lower ceilings in accuracy performance which may occur at faster conversational rates), and 3) model after the best-case transliteration setting, where the speaker's rate is at a slower conversational rate with many pauses.

In anticipation of intelligibility experiments such as this study, the video performance of each transliterator was then edited into short video clips, using Adobe Premiere Pro 1.5. The video performances were segmented at phrase boundaries (one phrase per video clip), resulting in roughly 80 video clips per transliterator at the slow-conversational rate. When possible, clip boundaries were aligned with visual "break points" in the video that corresponded to the deliberate pauses in the audio recording. In other words, visual "break points" were points in the video where the transliterator had finished cueing the phrase or sentence and paused before cueing the next phrase or sentence, thus retaining the speaker's pause from the audio recording. There were times, however, when the transliterator: 1) cued a liaison at a natural break point (thereby connecting two separate phrases together with cues), or 2) became unsynchronized, erroneously producing a cue from the boundary of one phrase or sentence with the mouth movements of another (making it impossible to divide the two sentences without the confusion of cue information belonging to a neighboring sentence). In these cases, two phrases or short sentences were either combined (if the resulting combination did not exceed 12 words) or divided at alternate break points, provided that the modified clips were semantically appropriate and did not contain more than 12 words. This upper limit on phrase length was instituted in order to maximize the likelihood that participants would be able to remember the words in the video clip long enough to be able to write them down. Given that most people can remember seven, plus or minus two, unrelated meaningful bits of information (Miller, 1956), it is reasonable to expect participants to remember 12 word phrases, considering that each phrase contains less than or equal to 7 content words and that the words in each clip are related and frequently organized into units (for example, the words within a prepositional phrase, such as “at the bottom of,” would most likely not need to be remembered for the individual pieces, but as a single unit).

Preparation of Stimuli. In order to analyze the relationship between accuracy and intelligibility, and the secondary effects of lag time and experience on this relationship, it was necessary to note the accuracy score and lag time associated with each clip, as well as the experience level of the transliterator who produced the clip. Accuracy data for each clip was derived from an existing database of cue-by-cue accuracy measurements created for previous studies in our laboratory (Pelley, et al., 2006; Tessler, 2007). In this database, every cue produced by each transliterator was classified as a correct cue, a substitution (cue containing incorrect handshape and/or incorrect placement), an omission (deletion of a required cue), or an insertion (introduction of an unnecessary cue). Accuracy data for each clip were then computed in Microsoft Excel using these cue-by-cue classifications, with formulas created to determine the accuracy score for the specific portion of the database corresponding to the cues in the clip. In addition to the accuracy data per clip, key word accuracy data for each clip was also derived from the existing accuracy data in order to allow for analysis of the relationship between key word accuracy and key word intelligibility. Key words were identified by a panel of transliteration experts in part of the larger project as “content words that need to be in the script for full comprehension” (Kile, 2005). Key word accuracy data for each clip was calculated in Excel by computing the cue-by-cue accuracy percentages for each key word, then averaging the overall key word accuracy percentages for each of the key words found within the same clip.

A similar database for lag time measurements was partially completed for other studies (Park, 2005; Smart, 2007). That database consisted of the measurements for eight of twelve transliterators, with lag times calculated for the beginning word, ending word, and middle syllable in each phrase or sentence. Corresponding measurements were completed for the remaining four transliterators, and lag time data for each clip was computed by averaging the three (beginning, middle, and end) individual lag times within each video clip. If one or more of the three lag times within a clip could not be calculated because the word was skipped by the transliterator, then the average lag time was based on the one or two available lag times for the clip.

Selection of Stimuli. In all, roughly 900 clips elicited at the slow-conversational rate were available for use in this study (3 sections x 4 transliterators per section x 75 phrases per section/transliterator). These clips included approximately four instances (one per CST) of each of the roughly 225 phrases in the audio narration. A subset of the video clips was selected and assembled into four stimulus blocks such that all phrases from the audio narration could be presented to the receiver in order. Thus, each stimulus block consisted of approximately 225 clips. The video clips in each stimulus block were not only selected in order to display all phrases from the narration, but also with the primary objective of selecting clips with a variety of accuracy scores ranging from near 0% to near 100%. Within this range of accuracy scores, clips were selected to sample values across the accuracy spectrum, in order to provide a continuous accuracy variable along the x-axis for comparison with the continuous dependent variable, intelligibility. To achieve this range, a total of nine accuracy levels were targeted (10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%), with the goal of selecting 20 to 30 clips at each accuracy level for each stimulus block. Clips qualified for selection at a given accuracy level when the clip's accuracy score was within +/-5% of that accuracy target level.

A secondary goal in stimulus selection was that the 20 to 30 clips selected at a given accuracy level would also be distributed as equally as possible across location in lecture. In addition, clips at each accuracy level were composed of a variety of transliterators, when possible, to allow for analysis of transliterator experience level on the accuracy-intelligibility relationship. Similarly, clips with a variety of lag times were also selected, so that the effect of different lag times on the accuracy-intelligibility relationship could be analyzed. In order to manage these competing constraints during clip selection, a master spreadsheet was constructed to display all relevant information regarding each clip: the transliterator's experience code, the lecture section (1, 2, or 3), and the target phrase or sentence, as well as the accuracy score and average lag time corresponding to the clip. Color coding was applied to values within each characteristic (accuracy, key word accuracy, experience, CST, and lag time) that were in short supply, with yellow, orange, and red indicating severity of shortage (calculated from histograms of the values available within each characteristic). The spreadsheet also denoted any phrases that were omitted entirely by a particular transliterator as well as any clips that were created from alternate break points.

As clips were selected, a second spreadsheet displayed running totals for each stimulus characteristic within various ranges as well as the ideal number of clips corresponding to each of those ranges, for reference. Running totals were constantly reviewed as the experimenter attempted to meet as many of the stimulus selection goals as possible. Clips were selected from the most constrained accuracy target ranges first and then went in order either from lowest accuracy range to highest (Stimulus Blocks 1, 3, 4, since more variety remained at the higher accuracy ranges) or the reverse (Block 2). Within a target range, color coding was used to select clips with values that were severely needed whenever possible.

In summary, individual transliterated video clips were selected 1) in order to display all phrases from the narration to the receiver in order, 2) in order to display 20 to 30 clips at each accuracy level, 3) in order to distribute accuracy scores across location in the lecture to the maximum extent possible, 4) while using a variety of transliterators at different experience levels, 5) with clips that contain a variety of lag times. Using this stimulus selection procedure, four stimulus blocks were created so that each stimulus block consisted of a different set of video clips. The stimulus blocks were counterbalanced across participants in order to minimize the effect of context and the impact of coincidental pairings of difficult phrases with poorer accuracies.

Presentation Sessions

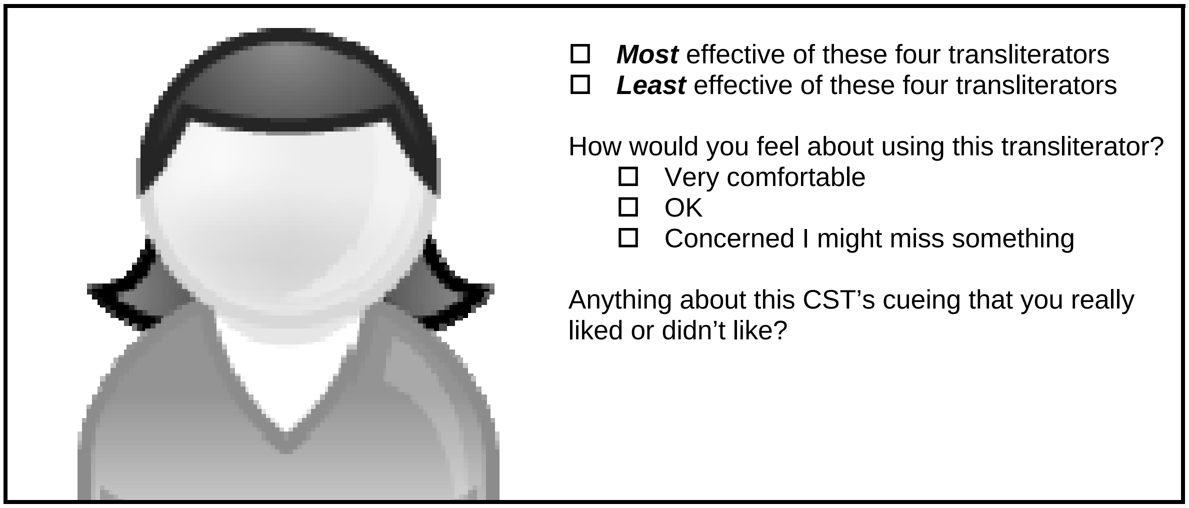
All participants were tested individually at a computer in a sound-treated room at the University of South Florida. Presentation sessions were conducted in two 2-hour sessions, with one 15-minute and two 10-minute breaks per session. Participants were also encouraged to take breaks as necessary to maximize attention.

The original film *Life Cycle of Plants* was presented (without audio) in short segments (i.e. scenes) on a computer monitor. After each segment, the film was paused, and one or more stimulus items were presented. The stimulus item(s) corresponded to the audio narration for that segment of the film, so that the film segment that preceded them served to provide relevant context and simulate the classroom environment. Each stimulus item consisted of one phrase (i.e. video clip) of the transliterated message. Given the goals of stimulus selection, consecutive stimulus items were not necessarily produced by the same transliterator. Nonetheless, the available materials dictated that stimulus items selected for each of the three lecture sections could be drawn from a subset of only four CSTs, which afforded cue receivers a chance to become familiar with the CSTs throughout the course of a lecture section. Cue receivers were instructed to type what the transliterator cued verbatim and were only permitted to view each segment of the educational film and corresponding stimulus items one time.

All receivers controlled the rate of presentation of the stimuli via a user interface implemented in Matlab, a computer software tool. The interface consisted of a window for displaying the video clips, a response blank for collecting their responses, and a "play" button to click when they were ready for the next item.

Receiver Ratings and Subjective Impressions

Cued Speech receivers were also asked to rate transliterators and to provide their subjective impressions regarding each transliterator's performance. After each section of the lecture was completed, receivers were given a short survey that included the pictures (for reference) of the four transliterators they had just finished viewing. Figure 1 shows the format of the survey which was used to collect this data was from each participant. At the conclusion of the experiment, each receiver was also asked to review all 12 transliterators and to circle any that were "highly effective" and place an "X" next to any that were "highly ineffective." No limits (minimum or maximum) were placed on the number of transliterators that could be placed in either category.



☐ ***Most*** effective of these four transliterators
☐ ***Least*** effective of these four transliterators

How would you feel about using this transliterator?
- ☐ Very comfortable
- ☐ OK
- ☐ Concerned I might miss something

Anything about this CST's cueing that you really liked or didn't like?

Figure 1. Format of survey used to collect each receiver's ratings and subjective impressions of the 12 transliterators.

Scoring

Intelligibility scoring included analysis of both OM intelligibility and TM intelligibility. For original message intelligibility, two types of scores were calculated to determine the intelligibility of 1) all words and 2) key words only (the same key words as those used for key word accuracy, described earlier). Percent-correct intelligibility scores were tabulated by examining the agreement between the typed responses and the original spoken messages corresponding to the transliterated phrases. When scoring all words, each word was required to be exactly correct, with the exception of obvious errors (explained below). When scoring key words only, morphological errors involving the addition or deletion of affixes were considered acceptable, such as the omission of "-ing" from "scurrying," provided that the stem of the word was perceived correctly. In both types of scoring, credit was given for obvious spelling and typographical errors as well as homophonous words, but lexical errors (such as "grow," instead of "thrive") were considered incorrect in order to exclude situations where the receiver did not have access to a word because of poor accuracy, but filled in the gaps given the context of the information (because intelligibility, not comprehension, was the measure of interest).

For transliterator message intelligibility, only key word scores were measured. Participant responses were judged against only those key words from the original message that the transliterator actually attempted to cue (i.e. any key words the transliterator provided by mouthing and correctly cueing at least some portion of the key words).

Participant responses were graded first by an autoscoring program in Matlab, which automatically scored an entire sentence as “correct” when the participant’s response contained all of the words to be scored (all words or key words) with exact spelling. The program also generated an Excel file containing any sentence that was not entirely correct. This file was then hand graded by the experimenter, and credit was given for simple typographic errors and errors in spelling that did not change the word phonologically (for example, the entry of a homophone such as “there” for “their”). For key word grading, the experimenter also gave credit for morphological errors that involved the addition or deletion of an affix.

Data Analysis

OM and TM percent-correct intelligibility scores from the experiment were calculated for each receiver and transliterator as well as for individual stimulus items. Transliterator scores were used to obtain an overview of the accuracy-intelligibility relationship (by comparing intelligibility scores with accuracy averages for each transliterator), and receiver scores were used to look for variability between receivers. The OM scores for individual stimulus items were compiled to construct three scatterplots and corresponding psychometric functions for transliterator intelligibility as a function of accuracy: 1) all word intelligibility vs. accuracy, 2) key word intelligibility vs. accuracy, and 3) key word intelligibility vs. key word accuracy. Key word intelligibility was plotted as a function of both accuracy and key word accuracy in order to focus on the effect of transliterator error patterns (key word errors vs. errors in other words) on the accuracy-intelligibility relationship. For each of these three sets of data, the strength of the relationship between intelligibility and accuracy was measured using a Spearman's rho correlation (a non-parametric test of correlation was used because the data were not normally distributed). In addition, the shape of the function was explored by determining what type of function appeared to fit the data best (linear, polynomial, probit, etc.) in a least-squares sense.

The resulting functions were also analyzed with regard to factors such as transliterator experience and transliterator lag time. The three accuracy-intelligibility psychometric functions were plotted for each experience group (veteran, experienced, and novice) separately in order to examine the effect of experience on the shape of the functions. Then, the effect of lag time on the shape of the accuracy-intelligibility psychometric functions was examined by plotting these functions for several different lag times (e.g. "short", "medium", and "long"). In addition, the relationship between lag time and intelligibility was investigated, first by constructing a psychometric function with lag times for each stimulus item plotted along the x-axis and corresponding all word OM intelligibility scores plotted along the y-axis and then by measuring the strength of the relationship by conducting a Spearman's rho correlation between the two variables. Finally, for stimulus items with the same accuracy, but different intelligibility scores, further analysis, including analysis of subjective receiver responses, was informally conducted regarding the effect of cueing rate, transliterator error patterns (omission of whole-word versus the omission of multiple cues from several words), synchrony of the transliterator's mouth and hand, clarity of the transliterator's speech movements, stylistic differences in positioning, handshape, etc.

Chapter Three

Stimulus Selection

In selecting stimuli for each of the four stimulus blocks, the highest priority was to construct stimulus blocks containing a continuous distribution of the full range of possible accuracy and key word accuracy values. Following the primary goal, other selection factors included selecting clips containing a variety of experience levels and lag time values such that each participant would view approximately the same proportion of clips for each transliterator, experience level, and lag time interval. Using the stimulus selection procedures, four stimulus blocks were assembled. As shown in Table 2, each stimulus block consisted of approximately 225 video clips (225 clips for Stimulus Block 1, 224 clips for Stimulus Block 2, 227 clips for Stimulus Block 3, and 224 clips for Stimulus Block 4). Although each stimulus block represented a transliteration of all phrases from the film narration, the number of total clips varied across blocks depending on the number of times within a block that two consecutive spoken phrases were combined into one video clip (for reasons given in the previous section).

Table 2

Stimulus Block Composition

	Number of clips				
	Total	Novel Clips	Duplicates	Triplicates	Quadruplicates
Block 1	225	225	0	0	0
Block 2	224	147	77	0	0
Block 3	227	123	48	56	0
Block 4	224	92	57	28	47

Because it was also necessary for each stimulus block to contain all phrases from the narration in order, there were, of course, limitations in the number of clips available for selection within each of these selection factors. Most notably, fewer clips were available at some accuracy levels (especially in the 0% to 35% accuracy range) than others. Similarly, there were limitations to the number of clips available for each transliterator because many transliterators did not cue every sentence from the lecture. Experience level limitations were due to the unequal number of novice, experienced, and veteran transliterators who participated in this study (2 novice, 1 experienced, and 9 veteran). Available lag time values were the result of measurements of transliterator behaviors and could not be manipulated. As a result of these limitations, it was necessary to use some clips in more than one stimulus block. Still, even by the fourth stimulus block that was assembled, 92 of the 224 clips selected were unique clips that had not been used in any previous stimulus block. The overlap between blocks is summarized in Table 2.

Accuracy

The scatterplots in Figure 2 show the distribution of accuracy and key word accuracy values for each video clip selected within each block. The number of unique clips (shown as green diamonds) visibly decreases with the addition of each new stimulus block. Nonetheless, these figures demonstrate that a continuous range in accuracy values from 0% to 100% is well-represented in each of the stimulus blocks. In order to obtain this range, however, nearly all clips available with accuracy values from 0% to 35% were used in all four stimulus blocks due to the low proportion of clips containing cued phrases with accuracy totals in this range.

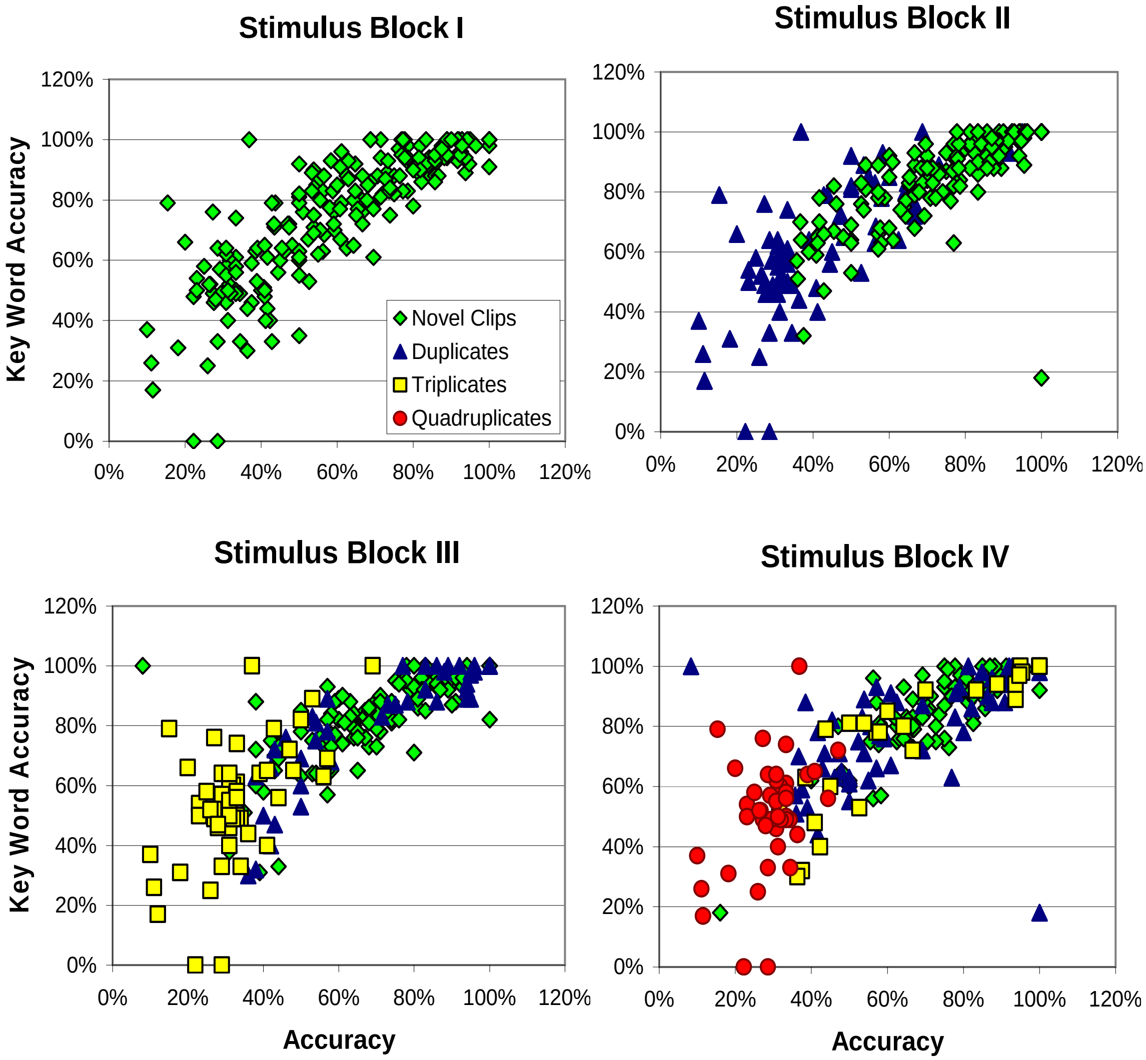


Figure 2. Scatterplots of accuracy and key word accuracy for each stimulus block. Novel video clips are shown as green diamonds, duplicate clips are shown with blue triangles, triplicate clips are shown as yellow squares, and quadruplicate clips are shown as red circles (representing clips that were used in all four stimulus blocks).

Experience

As seen in Table 3, the number of clips representing a particular experience category was similar for each of the four stimulus blocks, although the number of clips in each of the three categories was quite different. Given that there was only one transliterator classified as “experienced” and two transliterators classified as “novice,” while nine transliterators were classified as “veteran,” it was not possible to select the same number of clips in each experience category. Rather, each experience category was represented proportionally by distributing clips as equally as possible among CSTs. With approximately 225 clips per stimulus block and 12 CSTs, the ideal distribution across CSTs would call for approximately 19 clips per CST. Thus, the ideal experience distribution for each stimulus block would be 38 clips in the novice category (19 clips x 2 CSTs), 19 clips in the experienced category (19 clips x 1 CST), and 171 clips in the veteran category (19 clips x 9 CSTs). As Table 3 shows, every stimulus block contained sufficient representation of both the veteran and the experienced categories as well as a high percentage of the ideal number of clips for the novice category, with a minimum of 32 novice clips per stimulus block.

In addition, Appendix B lists the number of stimulus clips per CST for each stimulus block, which confirms that the clips selected within experience category were well-distributed across the CSTs in that category. While it was not possible to select exactly 19 clips per CST and balance all factors involved (accuracy, key word accuracy, CST, experience, and lag time), it is important to note that the number of clips per CST for all four stimulus blocks remained within the range of 13 to 26 clips, with only about 5% of the CST clip totals deviating by more than 4 clips from the ideal total of 19 clips.

Table 3

Number of Clips per Experience Category

	Ideal	Block 1	Block 2	Block 3	Block 4
Novice	38	41	35	32	33
Experienced	19	21	25	21	26
Veteran	171	169	171	178	172

Lag time

Overall, the clips available for stimulus selection contained a wide range of lag time values. These values were not uniformly distributed, as the available lag time values resulted from transliterator behavior and could not be manipulated for the purposes of the study. Clips were therefore chosen to represent the underlying distribution, with each range of lag time values represented proportionally; because a majority of the clips available for selection contained lower lag time values (from approximately 0.75 seconds to 1.75 seconds), a majority of the selected clips contained these lag time values as well. As shown in Figure 3, no one stimulus block contained a disproportionate number of clips at any specific range of lag time values. As a result, the number of clips representing a particular range of lag time values was similar for each of the four stimulus blocks, although the number of clips in each range varied widely.

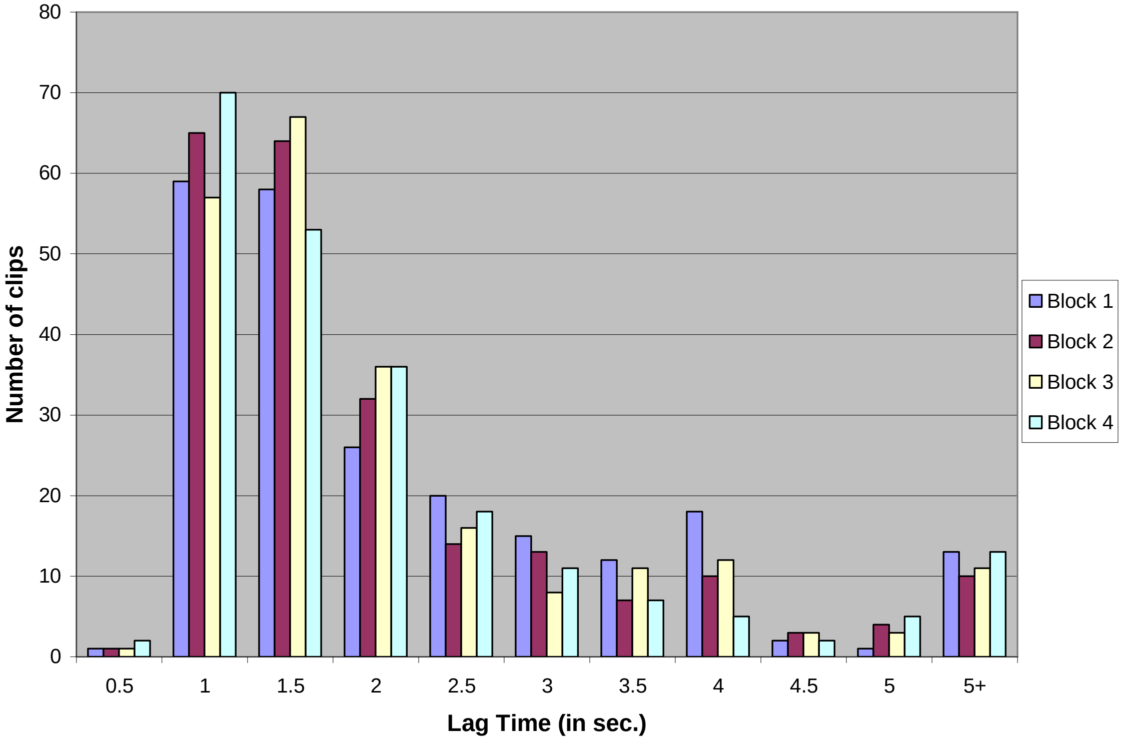


Figure 3. The number of clips selected for each 0.5-second range of lag time values in each stimulus block.

Individual CST Representation

Finally, the accuracy values corresponding to the stimuli selected for each transliterator were examined in order to determine whether a full range of accuracy values was obtained per CST. Figure 4 shows that while every CST was represented by a wide range of accuracy values, only about half were represented throughout the full range of possible accuracy values. This can most likely be explained by a lack of available clips produced within certain accuracy ranges by individual CSTs. However, another possibility is that the clips selected may not have sampled the full range of an individual CST's accuracy values. If this were the case, a discrepancy between the average accuracy of the stimuli selected for a particular CST and that CST's overall accuracy would be expected. However, Table 4 shows that the average accuracies of the stimuli selected for each CST were, in general, highly reflective of their overall accuracies.

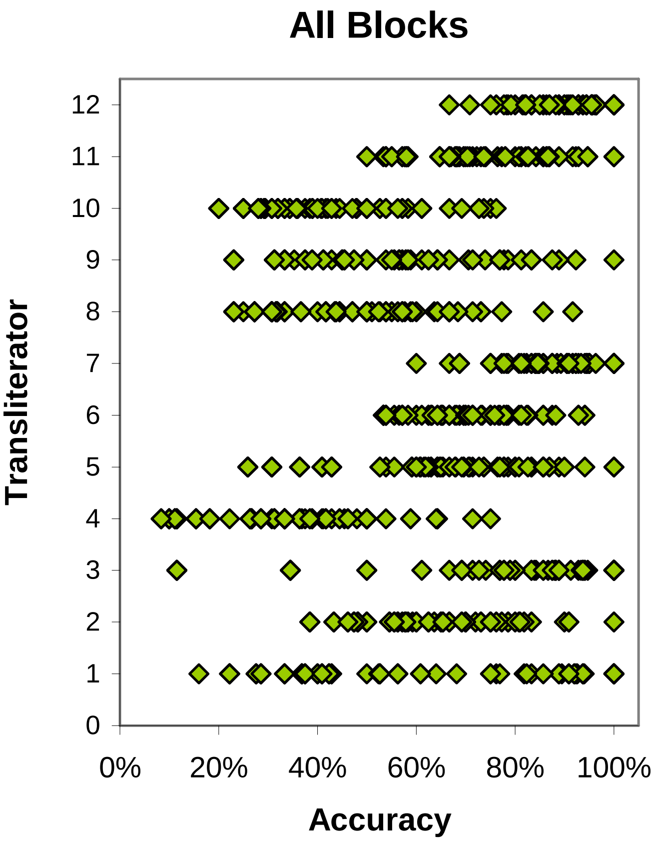


Figure 4. Accuracy distribution of selected stimuli for each transliterator (CST1 through CST12) across all stimulus blocks.

Table 4

Accuracy for Selected Stimuli and Full CST Performances

	Accuracy	
Transliterator	Selected Stimuli (%)	Full CST Performance (%)
CST1	68	55[a]
CST2	67	66[a]
CST3	81	64[a]
CST4	40	40[a]
CST5	69	68[b]
CST6	71	71[c]
CST7	86	84[d]
CST8	51	53[d]
CST9	59	55[d]
CST10	47	49[d]
CST11	73	73[d]
CST12	90	89[d]

[a] Husaim & Tessler (2006) [b] Lindsay (2006) [c] Pelley (2006) [d] Tessler (2007)

The only two transliterators whose accuracy was not well-represented by the stimulus selection were the two novice transliterators (CST1 and CST3). In these two cases, the average accuracy of the stimuli selected was much higher than the true overall accuracy of the transliterator: the average accuracy of the stimuli selected for CST1 was 68%, 13 percentage points higher than his/her actual overall accuracy (55%), and the average accuracy of the stimuli selected for CST3 was 81%, 17 percentage points higher than his/her true overall accuracy score (64%). This disparity was due to the high percentage of whole-word omissions produced by these two novice transliterators, who often omitted entire phrases. When a phrase was omitted, there was no clip available for selection. If these “non-clips” (skipped phrases) had been eligible for stimulus selection, the average accuracy of the stimuli selected for each of these two transliterators would have been much lower.

Chapter Four

Intelligibility Results

Table 5 summarizes the physical characteristics (accuracy, key word accuracy, and lag time) of the stimuli presented in the experiment, as well as the corresponding intelligibility results for each expert cue receiver. The information regarding physical characteristics confirms that the stimuli selected were indeed well-balanced across receivers with respect to accuracy scores, key word accuracy scores, and lag times. Each participant viewed a stimulus block that averaged 61% accuracy across all clips presented (with varying accuracy scores 0% to 100%). The average key word accuracy of the stimulus clips was also approximately the same for each participant, with overall key word averages of 75% and 76% for each participant's stimulus block. Lastly, the average lag time of the stimulus clips presented to participants varied only slightly, with average lag times per full stimulus block ranging from 1.8 seconds to 1.98 seconds, a difference of less than 10%.

Table 5

Stimulus Characteristics and Intelligibility by Receiver (Averaged Over Stimulus Block)

	Characteristics of Stimuli Received			Intelligibility		
Receiver	Accuracy (%)	Key Word Accuracy (%)	Lag Time (sec)	OM All Word (%)	OM Key Word (%)	TM Key Word (%)
CS01	61	75	1.98	79	83	89
CS02	61	76	1.80	72	78	83
CS03	61	75	1.98	76	84	89
CS04	61	76	1.80	72	79	84
CS05	61	75	1.84	69	75	80
CS06	61	75	1.82	74	76	81
CS07	61	75	1.84	68	70	76
CS08	61	75	1.82	65	70	74
Average	61	75	1.86	72	77	82
Range	8-100	0-100	0.34-7.31	0-100	0-100	0-100

The overall intelligibility (averaged across all transliterators and all receivers) for these stimuli was 72% for all words in the original message, 77% for key words in the original message, and 82% for key words in the transliterated message. Even though large differences in absolute performance levels were observed between cue receivers (15 percentage points between the highest and lowest intelligibility scores obtained by individual expert cue receivers in each of the three intelligibility measures), the relative intelligibility of these three measures was also observed in all individual cue receiver averages. That is, all cue receivers obtained the highest overall scores for TM key word intelligibility and the lowest overall scores for OM all word intelligibility, regardless of their absolute performance levels. Moreover, the average OM intelligibility score obtained by each cue receiver was considerably higher than the average accuracy (61%) in all cases and higher than the average key word accuracy (76%) in roughly half of the cases. The average TM intelligibility scores, ranging from 74% to 89%, were even higher than OM intelligibility scores, demonstrating that Cued Speech reception was generally high for any words that the transliterator attempted to cue[1]. Even so, TM intelligibility scores did not approach the intelligibility scores of normal hearing listeners, who received, on average, 98% of all words in the original message and 99% of key words (Tope, 2008), given audio stimulus items based on the film's narration (i.e. spoken with 100% accuracy).

Transliterator Intelligibility Differences

As Table 6 shows, the intelligibility scores for individual transliterators followed the same pattern as the overall intelligibility results: TM intelligibility scores were the highest of the three intelligibility measures, followed by OM key word intelligibility, and finally, OM all word intelligibility. In addition, intelligibility scores for individual transliterators were substantially higher than accuracy for most of the group: CST4 scored 14 percentage points higher on intelligibility than accuracy (only 40% on accuracy but 54% on all word intelligibility), CST6 scored 15 percentage points higher on

---

[1] TM intelligibility here is averaged across CSTs and would be higher for some individual CSTs, presumably those with the highest accuracy.

intelligibility than accuracy (71% accuracy and 86% on all word intelligibility), CST9 and CST11 scored 17 percentage points higher on intelligibility than accuracy (59% and 73% on accuracy, but 76% and 90% on all word intelligibility, respectively), and CST8 scored 23 percentage points higher on intelligibility than accuracy (51% all word accuracy, but 74% all word intelligibility). Interestingly, the CST with the highest TM intelligibility (CST1) was not one of the transliterators with the highest accuracy or highest OM intelligibility. This large difference between OM and TM intelligibility scores for CST1 can be attributed to the high amount of paraphrasing used by this transliterator (a novice).

Table 6

Stimulus Characteristics and Intelligibility by CST (Averaged Over Stimulus Block)

	Characteristics of Selected Stimuli			Intelligibility		
CST	Accuracy (%)	Key Word Accuracy (%)	Lag Time (sec)	OM All Word (%)	OM Key Word (%)	TM Key Word (%)
CST1	68	70	3.30	66	66	95
CST2	67	81	1.66	75	79	79
CST3	81	85	3.42	76	82	86
CST4	40	59	3.41	54	66	79
CST5	69	85	1.92	77	86	88
CST6	71	84	1.10	86	88	88
CST7	86	95	1.13	87	91	91
CST8	51	75	1.21	74	78	78
CST9	59	75	1.10	76	79	79
CST10	47	67	0.76	52	60	60
CST11	73	85	1.41	90	94	93
CST12	90	96	1.10	87	91	91

Although each expert cue receiver saw a full stimulus block with the same accuracy values on average (i.e. 61%), Table 6 shows that the average accuracy for the stimuli selected from each CST was not the same. This was because the accuracy scores available from each CST were different. Given these differences, the transliterators' average intelligibility scores provide a rough indication of the relationship between accuracy and intelligibility. In general, the transliterators' intelligibility scores followed accuracy, with transliterators who had the higher accuracy averages also obtaining higher intelligibility scores and transliterators with lower accuracy averages obtaining lower intelligibility scores. However, this pattern did not hold for all transliterators.

Deviations from the pattern may be easiest to see by comparing the rank order of transliterators based on accuracy averages with the rank order of transliterators based on average intelligibility. These rankings are shown in Table 7. While group ranking was generally preserved between accuracy averages and intelligibility scores for a majority of the transliterators, both of the novice transliterators had lower intelligibility rankings than accuracy rankings. While most transliterators exhibited higher intelligibility than accuracy, the accuracy averages of the *these* transliterators were nearly the same as their intelligibility scores, with CST1 averaging 68% in accuracy, but only 66% in all word intelligibility, and CST3 averaging 81% in accuracy and only 82% in all word intelligibility. As a result, CST1's intelligibility ranked lower than three transliterators with lower accuracy rankings, and CST3 ranked lower in intelligibility than four transliterators with lower accuracy averages. In contrast, CST9 and CST11 had intelligibility ranks that were considerably higher than expected given their accuracy ranks. With CST9 averaging only 59% in accuracy, but 76% in all word intelligibility and CST11 averaging 73% in accuracy, but a high 90% intelligibility, each was ranked higher in intelligibility than three transliterators with higher accuracy ranks.

Table 7

Transliterator Rankings by Accuracy, Intelligibility, and Subjective Receiver Ratings

Ranking	Accuracy	OM All Word Intelligibility	Subjective Receiver Ratings
1st	CST12 (90%)	CST11 (90%)	CST7 (9.375)
2nd	CST7 (86%)	CST12 (87%)	CST11 (8.75)
3rd	CST3 (81%)	CST7 (87%)	CST12 (6.875)
4th	CST11 (73%)	CST6 (86%)	CST3 (5)
5th	CST6 (71%)	CST5 (77%)	CST6 (4.375)
6th	CST5 (69%)	CST9 (76%)	CST5 (3.75)
7th	CST1 (68%)	CST3 (76%)	CST1 (2.5)
8th	CST2 (67%)	CST2 (75%)	CST2 (2.5)
9th	CST9 (59%)	CST8 (74%)	CST9 (1.875)
10th	CST8 (51%)	CST1 (66%)	CST4 (1.25)
11th	CST10 (47%)	CST4 (54%)	CST8 (0)
12th	CST4 (40%)	CST10 (52%)	CST10 (0)

Table 7 also shows subjective rankings for each transliterator, derived from receiver ratings. To obtain the rankings, point values were assigned to each participant's responses (1.25 points for each "Very Comfortable" rating, 0.625 points for each "Okay" rating, and 0 points for each "Concerned" rating), yielding a composite rating from 0 (when all eight receivers rated the transliterator with "Concerned") to10 (when all eight receivers rated the transliterator with “Very Comfortable”). For the eight CSTs who obtained similar rankings in both accuracy and intelligibility, subjective rankings based on receiver ratings were also similar to these two rankings. Of the remaining four transliterators, three had subjective rankings that were similar to accuracy rankings (including CST1 and CST3 who ranked *lower* on intelligibility than accuracy and CST9 who ranked *higher* on intelligibility than accuracy). However, the subjective ranking of CST11 (who ranked first in intelligibility despite ranking fourth in accuracy) was most similar to his/her intelligibility ranking.

More details regarding each expert cue receiver’s ratings and subjective impressions of each of the twelve transliterators is summarized in Table 8. The data show a high amount of receiver agreement for ratings of the transliterators with the highest and lowest overall intelligibility. For example, the transliterators with the best intelligibility scores, CST11, CST7, and CST12, received the highest subjective rating scores and were chosen by a majority (if not all) of receivers as “highly effective” transliterators. Similarly, the transliterator with the lowest intelligibility score, CST10, also received the worst intelligibility rating score and was unanimously chosen by all eight receivers as a “highly ineffective” transliterator. The greatest amount of variability in participant ratings was found for CST3 and CST6, where some participants found the transliterators to be conveying a message with which they were “Very Comfortable” with, some found the message to be “Okay,” and some participants were “Concerned [They] Missed Something” from the CST’s transliteration.

Table 8

Receiver Ratings of each CST, Ranked by Intelligibility Scores

		Ratings (Number of Receivers)			Number of Times Selected	
Transliterator (Intelligibility)	Composite Rating (0-10 pts)	Very Comfortable (1.25 pts)	OK (0.625 pts)	Concerned (0 pts)	Highly Effective	Highly Ineffective
CST11 (90%)	8.75	6	2	0	8	0
CST7 (87%)	9.375	7	1	0	8	0
CST12 (87%)	6.875	4	3	1	6	1
CST6 (86%)	4.375	2	3	3	2	0
CST5 (77%)	3.75	0	6	2	1	2
CST3 (76%)	5	2	4	2	2	0
CST9 (76%)	1.875	0	3	5	0	3
CST2 (75%)	2.5	1	2	5	1	4
CST8 (74%)	0	0	0	8	0	8
CST1 (66%)	2.5	0	4	4	0	3
CST4 (54%)	1.25	0	2	6	0	4
CST10 (52%)	0	0	0	8	0	8

Finally, a few other idiosyncrasies in receiver ratings should be mentioned. First, while CST12 was selected by 6 out of 8 receivers as “highly effective,” 1 receiver found this transliterator to be “highly *in*effective.” Also, CST8 received an overall receiver rating score of 0 (indicating 8 out of 8 receivers felt they missed something from the message due to the CST's performance) and was unanimously selected as a "highly ineffective" transliterator, despite achieving 74% intelligibility. This is in contrast to two transliterators with lower intelligibility scores (CST1 and CST4 scored 66% and 54%) who were rated more highly (at least some of the receivers felt "OK" with the message conveyed by CST1 and CST4) and less frequently selected as "highly ineffective" transliterators.

Accuracy-Intelligibility Functions

As described earlier, the primary objective of this paper is to characterize the relationship between Cued Speech transliterator accuracy and the intelligibility of the messages received by expert cue receivers. Toward this end, scatterplots relating the accuracy and intelligibility of individual stimulus items for the three combinations of measures were examined (see Figure 5): 1) accuracy vs. OM all word intelligibility, 2) accuracy vs. OM key word intelligibility, and 3) key word accuracy vs. OM key word intelligibility. Although the data points in each of these three scatterplots are widely distributed, they do suggest a positive relationship between accuracy and intelligibility. Spearman's rank order correlations were performed, confirming that a statistically significant positive correlation exists between accuracy and intelligibility for each of the three functions. This non-parametric test of correlation was employed because the data are not normally distributed. As is visible on the graphs, the distribution of the accuracy-intelligibility functions is skewed toward 100%, given that a substantial number of stimuli reached the maximum intelligibility values.

The accuracy-intelligibility correlation was strongest for the relationship between accuracy and all word intelligibility of the original message and least strong for the relationship between accuracy and key word intelligibility of the original message, with Spearman’s rho values of 0.478 for all word accuracy vs. OM all word intelligibility (p=0.000), 0.384 for all word accuracy vs. OM key word intelligibility (p=0.000), and 0.472 for key word accuracy vs. OM key word intelligibility (p=0.000). The variation in *accuracy* accounted for 26% of the variation in OM all word intelligibility scores and 22% of the variation in OM key word intelligibility scores, while the variation in *key word accuracy* accounted for 25% of the variation in OM key word intelligibility scores.

In order to help analyze the concentration of data points within each intelligibility scatterplot, Figure 5 also shows the mean and mode intelligibility values for each 20 percentage point accuracy interval from 0 to 100%. The mean reflects the general underlying linear relationship between accuracy and intelligibility. The mode, however, shows the most frequently occurring intelligibility values for a given accuracy range were sometimes far from the mean. This discrepancy suggests that other trends beyond the linear accuracy-intelligibility relationship may exist but are difficult to see due to the copious number of overlapping data points on each graph.

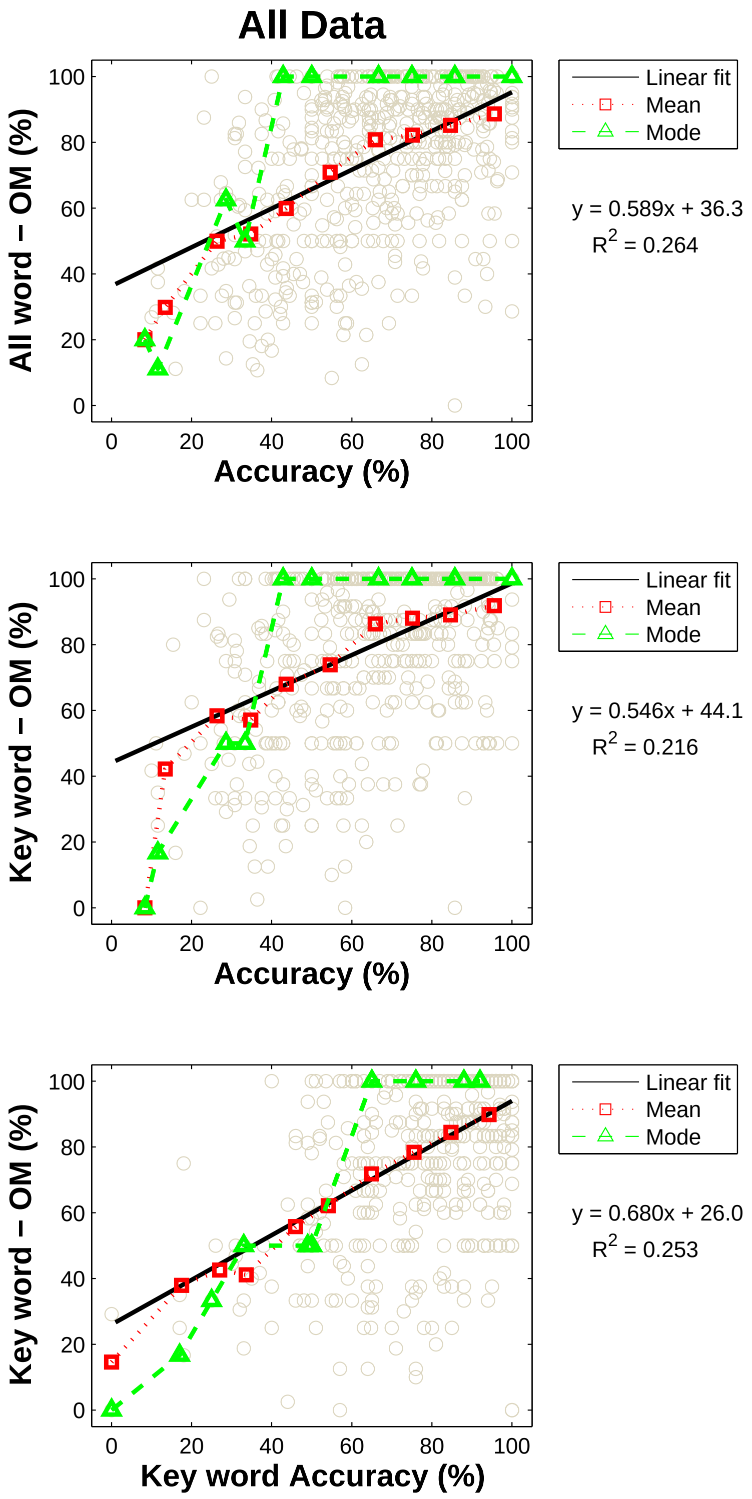


Figure 5. The accuracy-intelligibility relationship, with mean and mode intelligibility scores shown for each 10-point accuracy interval.

To further examine trends in concentrations of the data points, Figure 6 shows the likelihood that, given a data point with a certain accuracy value, receivers were able to figure out at least 70% of the message. The 70% threshold was chosen for this graph because in educational settings it represents a letter grade of “C,” the minimum passing grade. The likelihood value was determined by calculating the proportion of all data points in each 10-point accuracy interval that had greater than 70% intelligibility. This likelihood value thus approximates the probability that a receiver obtains an intelligibility score at or above 70% for any stimulus item in that accuracy interval. As such, the likelihood values provide some indication of the intelligibility mode as well as the intelligibility mean for each accuracy interval.

The likelihood function appears somewhat sigmoidal in shape, showing little change in intelligibility for changes in accuracy at the extreme ends of the accuracy scale (0-40% accuracy and 75%-100% accuracy) and a steeper slope in the middle of the scale where increases in accuracy cause more dramatic increases in intelligibility. Moreover, all three accuracy-intelligibility likelihood functions share characteristics of the same general shape (with similar slope and similar left-right shifts). However, the three functions appear slightly different at the higher accuracy values, as the accuracy vs. key word OM intelligibility function appears to reach a plateau in intelligibility scores for accuracy values above 60%, while the other two functions continue to increase in intelligibility for this accuracy range.

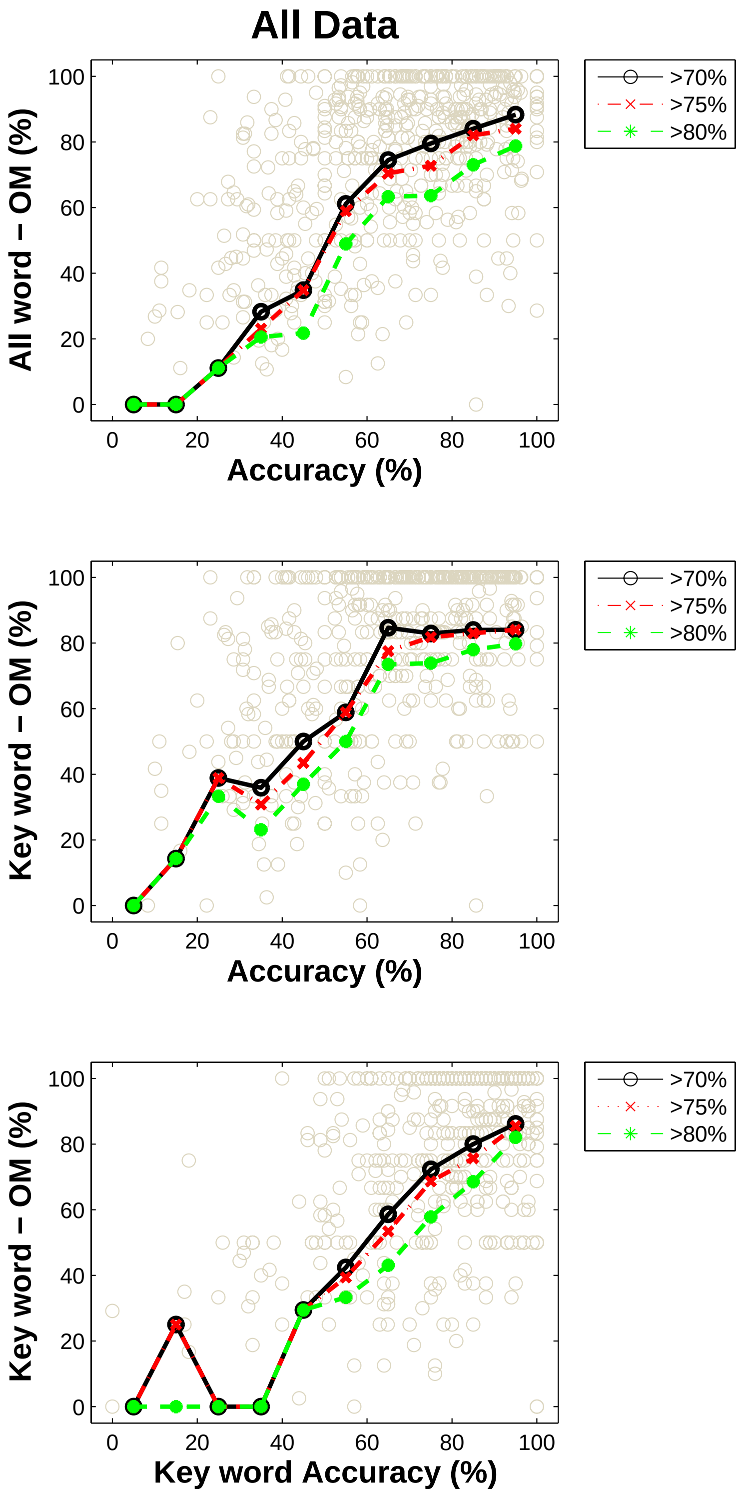


Figure 6. The accuracy-intelligibility relationship, with the proportion of data points that reach 70% or higher intelligibility shown for each 10-point accuracy interval.

Individual differences. Accuracy-intelligibility functions for individual receivers and transliterators are shown in Appendix C. Figures C1 and C2 show similar accuracy-intelligibility functions for all receivers, with differences mostly in left-right shift indicating differences in absolute performance levels between receivers. Figure C3 and C4 demonstrate that using stimulus items from many individual transliterators was helpful in widening the range of accuracy values that could be used to characterize the overall relationship between accuracy and intelligibility. Because stimulus items from some transliterators were restricted in accuracy range, the full accuracy-intelligibility function could not have been characterized with data from these transliterators only. For example, the range of accuracies utilized as stimulus items for CST11 and CST12 was restricted to higher accuracy values only, causing a ceiling effect in intelligibility (around 90% for CST11 and 80% for CST12). A limited range in intelligibility data also happens for CST2, CST6, CST9, and CST10. However, the data for each of these transliterators show neither a floor nor a ceiling effect and demonstrate a positive relationship between accuracy and intelligibility, with each characterizing a different portion of the overall accuracy-intelligibility psychometric function. Together, and with data from the remaining transliterators, a large portion of the accuracy-intelligibility function was characterized.

Effect of experience. In order to examine the effect of experience on the shape of the accuracy-intelligibility psychometric functions, the three likelihood functions were re-plotted, with data separated by experience group (veteran, experienced, and novice). As shown in Figure 7, the positive relationship between accuracy and intelligibility is apparent for both the novice and veteran groups, but no relationship is apparent for the experienced group. However, the experienced group consisted of only one transliterator, and number of stimulus items available from any one transliterator is not enough data to characterize the relationship. Although it is not possible to determine what relationship exists between accuracy and intelligibility for experienced transliterators, sufficient data are available to suggest similarities and differences between the novice and veteran groups. In comparing the novice and veteran accuracy-intelligibility functions, their slopes appear very similar, but where 70% accuracy corresponds to 50% intelligibility in the novice function, only 50% accuracy is needed to achieve 50% intelligibility in the veteran function. Thus, while intelligibility improves with accuracy at the same rate for both groups, the accuracy-intelligibility function of the novices is shifted to the right in comparison to veterans, indicating that task difficulty is higher for receiving information from novice transliterators than veteran transliterators, even when accuracy is controlled.

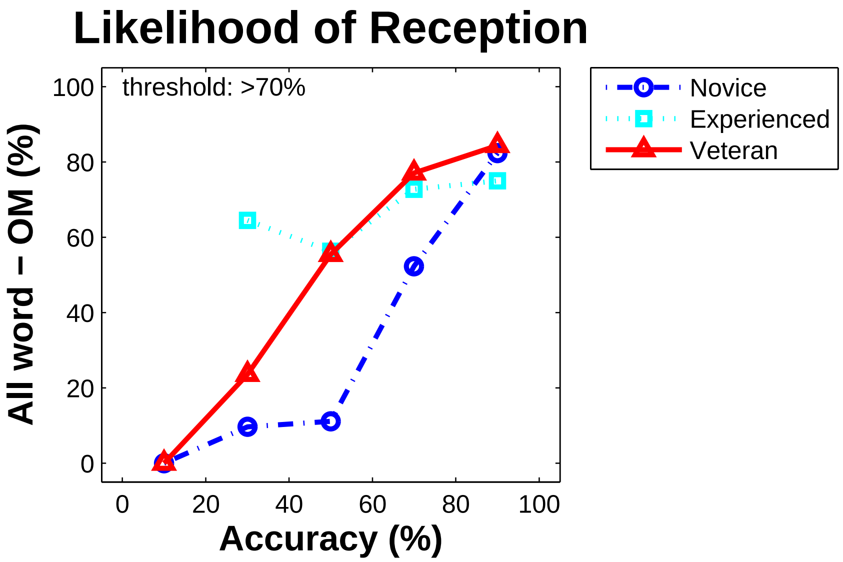


Figure 7. Accuracy-intelligibility likelihood functions plotted for each experience category (each showing the proportion of data points that reach 70% or higher intelligibility for a given accuracy interval).

Effect of lag time. As predicted, no large effects of lag time on intelligibility were found in this study. Figure 8 shows no clearly definable differences in the shape of the accuracy-intelligibility function when plotted separately for different lag time ranges. No difference in task difficulty is shown for these lag time ranges, which is evidenced by the lack of left-right shift between most of the functions. It is possible that the functions for lag time ranges of 0-1 seconds and 1-2 seconds may be somewhat to the left of the other functions (suggesting that receiving information from transliterators with shorter lag times may be an easier task than those with longer lag times, even when accuracy is controlled), or that there is a difference in slopes between the lag time ranges graphed (suggesting differences in the amount of variability associated with various lag time ranges). However, additional statistical analyses are needed to investigate these possible differences, and it is apparent that the differences, if any exist, are small.

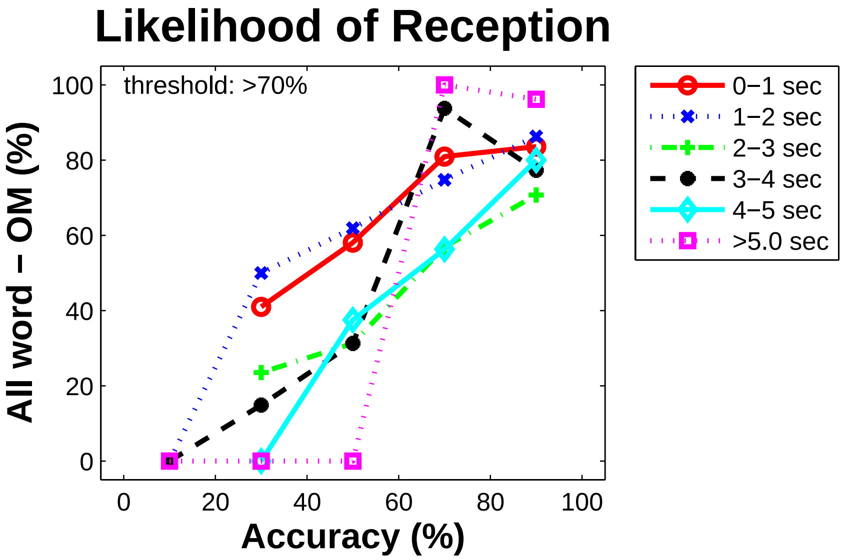


Figure 8. Accuracy-intelligibility likelihood functions plotted for each 1-second lag time range (each showing the proportion of data points that reach 70% or higher intelligibility for a given accuracy interval).

Lag Time-Intelligibility Function

Finally, the relationship between lag time and intelligibility was investigated by plotting lag time as the independent variable and OM all word intelligibility score as the dependent variable for each stimulus item in order to construct a scatterplot of the lag time-intelligibility relationship. A weak, negative relationship was observed, and the strength of the relationship between the two variables was measured by conducting a

Spearman's rank order correlation. However, no statistically significant correlation was found (Spearman's rho value of -0.041, p=0.319).

Although no linear relationship was found, the density of data points suggested a curvilinear relationship. To investigate this possibility, the likelihood function, representing the likelihood that stimulus items have an intelligibility score greater than 70%, was again calculated (see Figure 9). The general shape of the lag time-intelligibility function shows that intelligibility scores of greater than 70% were most likely to occur when lag times were between 1 and 1.5 seconds. Of the stimulus items with lag times in this range, 70% to 75% were associated with intelligibility scores of greater than 70%; this range can therefore be considered an optimal lag time range for CS transliterators. As lag time increased beyond this range, the decline in intelligibility was somewhat steep for lag times between 1.5 and 2 seconds (decreasing from 75% likelihood that the intelligibility cutoff is achieved in the optimal lag time range to only 55% likelihood when the lag time is 2 seconds). Further increases in lag times above 2 seconds, however, showed a more gradual decline in intelligibility scores (ultimately dropping as low as 45% likelihood at 4 second lag times). The steepest decline in intelligibility occurred for lag times that were shorter than those in the optimal lag time range. Stimulus items with lag times of less than 1 second had the lowest likelihood of reaching the intelligibility threshold (only 35% of stimulus items with lag time below 1 second had intelligibility scores of 70% or greater, compared to 70% of stimulus items with 1 second lag times, from the optimal lag time range).

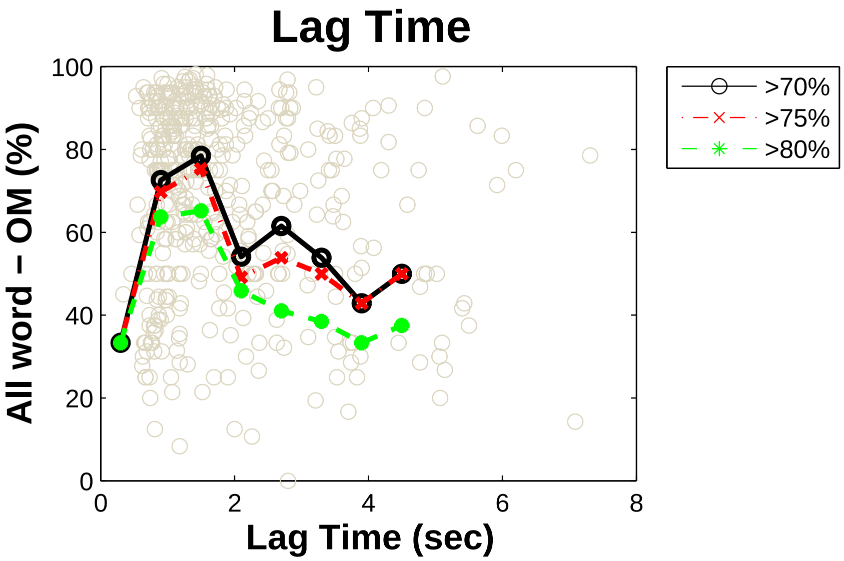


Figure 9. Lag time-intelligibility likelihood functions, each showing the proportion of data points that reach 70% or higher intelligibility scores for each 0.5-second lag time range.

Chapter Five

Discussion

The results of this study show that message intelligibility for typical Cued Speech transliterators is 72% of all words in the original message, on average, when conveying educational materials designed for high school settings. Intelligibility is higher for key words in the original message (77% on average) and highest for key words in the transliterated message (82% on average). Yet even the highest intelligibility score (TM-key word) for the receivers who obtained the highest intelligibility (CS01 and CS03), 89%, was less than the analogous measure obtained from normal hearing listeners (when presented with audio stimulus items of 100% accuracy). The intelligibility of key words in the original spoken message was 99.8% for monolingual listeners and 99% for bilingual listeners (Tope, 2008). To understand the causes of this intelligibility gap between normal hearing listeners and expert cue receivers, it is first necessary to identify the factors affecting message intelligibility of Cued Speech transliterators and quantify their contributions to intelligibility.

In this study, the primary factor under investigation for its effect on intelligibility was accuracy. As hypothesized, the relationship between accuracy and intelligibility was positive and nonlinear. Plots of accuracy-intelligibility psychometric functions constructed from the data collected showed that as accuracy increased, intelligibility increased. The same accuracy-intelligibility relationship was generally exhibited when plotted by receiver and by transliterator, though there were individual differences in the left-right shift of the resulting psychometric functions.

A second hypothesis of this paper was that the shape of the accuracy-intelligibility functions would be affected by experience, with flatter slopes and left-shifts in the functions for veteran transliterators as compared to novices. This hypothesis proved partially true for Cued Speech transliterators. The psychometric functions of veteran transliterators were shifted to the left, as compared with novices, but there was no significant difference in the slopes of the functions for either group. Unlike the novices and veterans, the "experienced" category showed no relationship between accuracy and intelligibility. With only one transliterator (CST8) categorized as "experienced" rather than as a "novice" or "veteran," however, it is unknown whether s/he is typical of the middle experience category or an outlier. Data from more transliterators in this category will be necessary before the accuracy-intelligibility relationship for experienced transliterators can be determined.

Another hypothesis of this paper was that the effect of lag time on the accuracy-intelligibility psychometric function would be minimal (because lag time changes are more likely to affect accuracy alone, rather than the relationship between accuracy and intelligibility). As predicted, no effect of lag time was found on the accuracy-intelligibility psychometric function. However, lag time (the amount of time the transliterator lags behind the speaker) was an independent factor associated with changes in transliterator message intelligibility. Specifically, the psychometric function relating lag time and intelligibility indicated an "optimal" lag time range (where intelligibility scores greater than 70% occurred most frequently) for lag times between 1 and 1.5

seconds. Given that lag time changes are most likely to affect accuracy, this range of lag time values is expected to be associated with higher accuracy values as well.

Finally, it was hypothesized that even when experience and lag time were controlled, variability due to other factors would be evident in the psychometric function. As such, two stimulus items with the same accuracy could have different intelligibility scores, due to differences in cueing rate, speechreadability, prosody, or other factors. This hypothesis was true: for any given accuracy value, there was a large range in the resulting intelligibility values, demonstrating a high degree of variability unaccounted for by accuracy alone.

Other Sources of Variability

Because accuracy only accounted for approximately 26% of the variability in intelligibility, it is apparent that other factors also play a role in intelligibility. Many such factors are likely to stem either from differences in transliterators or differences in receivers. Therefore, possible sources of both transliterator variability and receiver variability merit further consideration. Sources of transliterator variability include any variability in intelligibility that is due to subtle differences in behavior, style, or performance of transliterators (i.e. between two transliterators or even within a single transliterator's performance). Sources of receiver variability, on the other hand, include the variability in intelligibility that occurs due to receiver-specific factors that affect message reception either between two receivers (e.g., receptive cueing fluency), or within a receiver (e.g., attention).

Transliterator variability. Possible sources of transliterator variability are numerous and include the transliterator's cueing mechanics (handshape formation, placements, and transitional cueing movements), speechreadability, error types, facial expressions, cueing rate, synchronization between mouth and cues, and prosody (i.e. timing and emphasis). Other sources of transliterator variability may include intra-transliterator factors (e.g. fatigue, attention, nervousness), differences in cue selection (e.g. cueing unstressed /i/, as in "funny" as /i/ versus /I/), and dialect (dialect may be apparent in selection of cues as well as mouth movements).

In this study, some of these factors obviously played a role in transliterator intelligibility, at least in the cases of the four transliterators whose intelligibility did not closely follow accuracy. CST11, for example, ranked fourth in accuracy at 73%, but achieved the highest intelligibility (90%) of any transliterator, surpassing even CST12, the transliterator with the highest accuracy (90%), whose intelligibility was 87%. It should, however, be noted that the accuracy difference between these two transliterators may be somewhat exaggerated. The reason for this is that the accuracy measurements were completed on a strict grading system, with no credit given for substitutions, even when part of the target cue (handshape or placement) was correctly produced. While CST12 produced few substitutions, CST11 often produced errors of substitution that resulted from a tendency to "hypocue" (approximate, but not fully achieve the intended handshape or placement). As such, CST11 was given no credit for 22% of his/her performance due to errors of substitution. If these substitutions were each partially correct, CST11 would have been deserving of an accuracy score of 84%, instead of 73%

under the strict grading system. Although such a scenario would account for part of CST11's higher intelligibility, it still does not explain why CST11 would be *more* intelligible that CST12, whose accuracy was 90%. Therefore, CST11 must have achieved high intelligibility results by capitalizing on other skills.

A close inspection of CST11's performance reveals several strengths which may have played a role in increased intelligibility: 1) highly visible/clear speech, 2) excellent facial expressions and other non-manual information (use of eyebrows, head leaning, and showing lists on the hand) to show questions, emphasize important points, and convey the tone of the message, and 3) effective use of available time while still keeping up with the message (i.e., CST11 capitalized on speaker pauses, slowing down slightly in order to better show syllable, word, or sentence stress). While CST12 demonstrated visible/clear speech and facial expressions, he/she did not capitalize on speaker pauses, but instead followed immediately behind the spoken messages, demonstrating pauses equal to speaker pauses. As a result, the message prosody and word stress were less pronounced, and the cueing rate was effectively faster. Because of this difference, CST11 had a longer average lag time (1.4 seconds) than CST12 (1.1 seconds, a lag time average that is at the lower end of the optimal lag time range of 1 to 1.5 seconds). It is therefore possible that CST11 ranked highest in intelligibility and was unanimously identified as a "highly effective" CST by all eight receivers because of better message prosody, slower effective cueing rate, and/or more optimal lag time.

Two other transliterators whose performance may lend insight into sources of transliterator variability are CST1 and CST3, the two novice transliterators whose intelligibility was worse than expected given their accuracy scores. While veteran transliterators generally achieved intelligibility scores that were significantly higher than their accuracy averages, the intelligibility of these two transliterators was below their accuracy averages, with CST1 averaging 68% accuracy and 66% intelligibility and CST3 averaging 81% accuracy and 76% intelligibility. Aside from experience level, specific factors noted by the experimenter that were likely to have negatively impacted intelligibility for both of these two transliterators include slow cueing rate, misleading facial expressions (concentrating, confused, or discouraged facial expressions), and poor timing and rhythm within words and sentences (extraneous pausing, poor demonstration of word or syllable emphasis, and poor conveyance of the importance of words). The slow cueing rate of both novices resulted in average lag times much longer than the optimal lag time range of 1 to 1.5 seconds (CST1 had an average lag time of 3.4 seconds and CST3 had an average lag time of 3.5 seconds). Additionally, it is possible that the cueing rates were *too slow* at times and may have caused receivers to have difficulty keeping words in working memory. Between CST1 and CST3, CST1 generally exhibited clearer visible speech, while CST3 was slightly less clear, held his/her hand at an atypical angle and had many false starts. Combined with extraneous pausing, these false starts resulted in a misleading rhythm that may have obscured word boundaries (in one case, a receiver perceived “surgically precise” as “surge eucalyptus”). All or most of these transliteration behaviors are likely to reflect the novice transliterators’ inability to cope with the cognitive load and/or physical demands of the task of transliterating at a conversational speed. More practice transliterating at faster speeds and/or experience would likely improve the intelligibility scores of these two novices. Of course, additional experience and practice alone would not necessarily improve all of the above factors, unless more training was also provided to raise awareness of these issues.

Finally, discrepancies between the intelligibility score and receiver ratings for one transliterator (CST8) may also be an indication of additional sources of transliterator variability. Based on receiver ratings, CST8 ranked last (tied with CST10; both received a composite rating of 0 on a scale of 0 to 10) and was unanimously identified as a "highly ineffective" transliterator, despite achieving 74% intelligibility and ranking 9th out of the 12 transliterators in intelligibility. Several transliteration behaviors were noted as possible factors that may explain why the receiver impressions for this transliterator were poorer than would be expected based on intelligibility alone. First, CST8 often did not show visibly discernable placements as his/her hand consistently remained in front of the chin (for side, throat, and chin cues) during a majority of the transliteration, regardless of the consonant or consonant-vowel combination being produced. Many of the receivers complained about this aspect of CST8's cueing, both on the survey and in conversations at the breaks during the experiment. Second, CST8 regularly used unusual and sometimes misleading mouthshapes, frequently producing lip rounding in words that should not contain it (for example, "slightest"). Next, CST8 exhibited poorer synchronization between mouth and cues than other transliterators, which became apparent when his/her performance was edited into individual sentence video clips. CST8 was the transliterator with the highest number of combined clips due to poor break points between phrases; frequently, the transliterator's mouth was still articulating the sound(s) of the previous sentence while beginning to cue a word from the next sentence or vice versa. Lastly, CST8's cue pacing and lack of facial expressions appeared less effective at conveying the natural rhythm and stress of speech than other transliterators.

Although CST8 still produced a message that was 74% intelligible overall, receivers were more aware of CST8's issues (or more bothered by them), than they were for two other transliterators with lower accuracy: CST1 had only 66% intelligibility and CST4 had only 54% intelligibility, but both were only identified as "highly ineffective" transliterators by half the group, whereas CST8 was unanimously chosen by all eight receivers as "highly ineffective." This difference in ratings, which can not be attributed to differences in intelligibility, suggests either a conscious level of awareness of CST8's shortcomings or a general frustration on the part of the receiver when viewing CST8. While the other two transliterators were less intelligible, it is likely that the receiver was sure about what cues these transliterators produced, whether the cues were correct or in error. However, viewing CST8's transliteration likely left the receivers with more uncertainty regarding what cues had been produced, causing them frustration and/or resulting in a higher cognitive load as they tried to determine what they were supposed to receive from CST8.

Receiver variability. Possible sources of receiver variability include the receiver's experience and comfort level with receptive cueing and as well as processing strategies used in cue reading. Because deaf individuals encounter more diverse communicative environments than do hearing individuals, these individuals are likely to have less consistent exposure to their chosen communication modes than hearing individuals.

While at least one participant (CS01) reported using cueing all the time with most members of his/her family, several of the receivers in this study reported that they are now (in their adult lives) primarily in non-cueing environments, either with hearing people who do not cue or with deaf people in settings where sign language is the primary mode of communication (for work or with deaf friends/spouses who do not cue). At least one participant (CS03) reported feeling "rusty" with receptive Cued Speech, saying s/he relied heavily on speechreading during the testing. When receivers are less experienced with receptive cueing or have with reduced comfort levels due to lack of recent use, more emphasis may need to be placed on clearer speech movements (to aid in speechreading), slower rate, and possibly other stylistic differences.

Regardless of comfort level, there may also be differences between cue receivers in processing strategies used for cue reading. Some receivers may rely more heavily on information from the lips (and prefer transliterators who exhibit a high degree of lip clarity), while others may rely more heavily on information from the cues (and prefer transliterators who exhibit a high degree of cue clarity). The former group would be expected to make errors more frequently that are consistent with the lips but incongruent with the cues produced, while the latter group would be expected to do the reverse (decoding the cues with less reliance on mouth clarity, more frequently making errors that are consistent with the cues but incongruent with the mouthshapes produced). Of course, for all receivers, some of both error types may occur, and both types of errors were indeed noted in this study. For example, CS01 perceived the word "ficus" (5s-5t2t3s) as "fights" (5s-5t5s3s). Both words are similar visually based on mouthshapes alone, but the cues distinguish the words; thus, CS01 followed the lips more than the cues, resulting in this error. On the other hand, CS03 followed the cues, rather than the lips, when perceiving the word "life" (cued 6s-5t5s) as "light" (also cued 6s-5t5s) – these words are cued identically, but it is not difficult to tell them apart through speechreading. Although both types of errors were observed, a cursory inspection of receiver responses suggests that the overwhelming majority of the errors made by the eight receivers in this study were such that responses tended to correspond more to the lips than the cues. However, generalizations about the receivers' processing strategies are difficult to make because errors in cue production (made by the transliterator) may have influenced receiver errors, even for receivers who use a processing strategy that typically relies more heavily on cues than on mouthshapes (analyzing cue-by-cue accuracy data to determine the type of transliterator errors made when words were not transliterated with 100% accuracy and aligning this information with receiver responses was beyond the scope of the present study).

Role of Training and Certification

Because a wide range of performances (in accuracy and in resulting intelligibility) was exhibited for the transliterators in the veteran category, a closer examination of the experience backgrounds of the veterans is warranted. Nine of the twelve transliterators recruited for this study qualified as "veteran" transliterators (defined as transliterators with the highest level of certification and/or more than 5 years work experience), and Table 9 illustrates that there were substantial differences between these individuals with respect to their years of transliterating experience, current weekly hours transliterating, and relevant certifications.

Table 9

Accuracy, Intelligibility, and Experience Profiles for Veteran CSTs, Ranked by Intelligibility

	Work Experience as a CST	Hours Per Week Transliterating	Relevant Certifications	Accuracy (%)	Intelligibility (%)
CST11	6 yrs.	5 hrs.	None	73	90
CST7	9 yrs.	30 hrs.	Nationally Certified (TSC)	86	87
CST12	20 yrs.	3 hrs.	Nationally Certified (TSC)	90	87
CST6	15 yrs.	30 hrs.	Educational CST Certified by State	71	86
CST5	15 yrs.	35 hrs.	Educational CST Certified by State	69	77
CST9	15 yrs.	32.5 hrs.	None	59	76
CST2	10 yrs.	35 hrs.	Educational CST Certified by State	67	75
CST4	22 yrs.	35 hrs.	Educational CST Certified by State	40	54
CST10	18 yrs.	35 hrs.	None	47	52

Although no one experience factor alone appears to explain differences in intelligibility, several important trends are evident. First, the two veteran transliterators with the fewest years of experience (CST7 and CST11) were among the highest in intelligibility, while the two veteran transliterators with the lowest intelligibilities (CST4 and CST10) were two of the three most-experienced veterans. One explanation of the intelligibility differences between these groups of transliterators is that training methods may have improved over the last 20 years. This explanation seems likely considering that the first formal Cued Speech transliterator training classes were not offered until 1985 (Krause, Schick, & Kegl, in press). If the training that transliterators received approximately 20 years ago was indeed more varied and continuing skill development training unavailable, this would also explain why the three veterans with the highest number of years experience, varied from among the most intelligible (CST12) to among the least intelligible (CST4, CST10).

Another factor which may explain intelligibility differences between the three highly experienced veterans is the level of certification obtained. While CST10 has no certification and CST4 was certified as an educational CST by the state (both of whom obtained the lowest intelligibility scores), CST 12 (who was highly accurate and highly intelligible) has *national* certification. In fact, two of the three most intelligible transliterators have obtained national certification, while those with state certification varied greatly in respect to their accuracy and intelligibility scores. Standards for state certification are more varying and can be less rigorous than national certification, with many transliterators automatically granted state certification based on their number of years of experience (under a state's "Grandfather Clause"). As a result, the difference in requirements for state versus national certification may be great.

While differences in certification may explain the difference in intelligibility for the transliterators with the most years of experience, certification status alone does not always predict intelligibility. Both CST11 and CST12 obtained the highest intelligibility scores of all nine veterans; however, CST11 has no certification, while CST12 is nationally certified. Since CST training methods have changed over time, perhaps CST11's initial training focused on additional factors that affect intelligibility, such as facial expression, synchronization, or lip movement, where it is possible that CST12's training did not. However, it is worth noting that although CST11 scored high intelligibility, he/she did not score nearly as high in accuracy. This difference in accuracy may be related to the lack of certification by CST11and suggests that transliterator evaluation processes should be focused on intelligibility, rather than accuracy, particularly considering that it is unknown whether or not intelligibility differences exist in transliterators due to differences in physical speed, language, or speech skills (all variables that are difficult for tests to measure).

Conclusions

Accuracy plays a large role in intelligibility, but there are many other factors that affect transliterator message intelligibility. Sources of transliterator variability point to a number of factors (e.g. visual speech clarity, facial expression, non-manual makers, and cueing rate) that caused some transliterators to be higher or lower in intelligibility than would be predicted by accuracy-intelligibility relationship. Sources of receiver variability (e.g. current comfort level with receptive cueing and processing strategies for cue reading, with heavier reliance on either lips or cues) also caused some receivers to perform differently, obtaining higher or lower intelligibility scores given stimulus items of similar accuracy.

While the field of interpreting cannot control for differences in receiver performance, interpreting standards can be introduced to reduce sources of variability between transliterators and improve overall intelligibility. In general, greater transliterator experience was found to have a positive effect on intelligibility. When comparing the performance on novices and veterans, the accuracy-intelligibility functions shifted to the left for veterans in compared to novices, given that higher intelligibility scores were obtained with lower accuracy scores for veterans than for novices. However, even when experience is controlled, much of the variance in intelligibility remains unexplained by accuracy (44% for novices, 72% for veterans). Therefore, it is important to isolate and quantify the contribution of other factors, such as those identified here (speechreadability, facial expressings, cueing rate, etc.), in future studies.

Future Work

While the use of 12 transliterators was sufficient for characterizing the accuracy-intelligibility relationship, it cannot be assumed that the results of these 12 individuals are representative of the performances of all Cued Speech transliterators. Therefore, future studies should expand the number of transliterators included. Special attention should be paid to recruit more transliterators who will qualify as “novice” and “experienced” transliterators, as there were only two novices and one experienced transliterator for the current study. Attention should also be paid to recruiting more nationally certified transliterators to ensure that the two nationally certified transliterators utilized in this study are representative of the skills of other nationally certified transliterators. Because a majority of the transliterators employed for this study were from a limited number of states, more transliterators with state certification from a variety of states should also be recruited in future studies. The inclusion of transliterators certified by other states would provide information regarding the variability of state standards.

Because differences in transliterator cueing rate were thought to be a factor in the intelligibility results of this study, future research is needed to quantify the effect of rate on message intelligibility. Without even recruiting more participants, an “optimal cueing rate” could be determined based on the existing data if cueing rate measurements were made for each stimulus item and correlated with the intelligibility results of this study. Future intelligibility experiments should also be constructed to draw from the full database of video clips (containing these 12 CSTs transliterating at slow, normal, and fast presentation rates). The effect of presentation rate could then be investigated by determining psychometric functions (analogous to those found in this study for intelligibility and accuracy, experience, and lag time) at each of the three presentation rates.

Finally, it is important to conduct similar experiments for communication modes other than Cued Speech. The quantitative analysis of the factors affecting intelligibility of transliterator messages is especially important for any communication mode utilized by deaf students in educational settings. This study is part of a larger study by Krause

(2006) that aims to complete similar intelligibility experiments for other communication modes, including Signed Exact English (SEE II), Conceptually Accurate Signed English (CASE, also sometimes referred to as Pidgin Signed English), and eventually American Sign Language (ASL).

References

Beasley, D. S., Bratt, G. W., & Rintelmann, W. F. (1980). Intelligibility of time-compressed sentential stimuli. *Journal of Speech and Hearing Research, 23,* 722-731.

Clarke, B. R., & Ling, D. (1976). The effects of using Cued Speech: A follow-up study. *The Volta Review, 78,* 23-34.

Cokely, D. (1986). The effects of lag time on interpreter errors. *Sign Language Studies*, *53*, 341-375.

Cornett, R. O. (1967). Cued Speech. *American Annals of the Deaf, 112,* 3-13.

Films for the Humanities (Producer). (1989). *The Life Cycle of Plants* [Film]. (Available from Films Media Group, P.O. Box 2053, Princeton, NJ 08543-2053)

Fischer, S. D., Delhorne, L. A., & Reed C. M. (1999). Effects of rate of presentation on the reception of American Sign Language. *Journal of Speech, Language, and Hearing Research, 42,* 568-582.

Gustason, G. (1990). “Signing Exact English.” In *Manual communication: Implications for education* (p. 108-127). H. Bornstein. Ed. Washington, DC: Gallaudet University Press.

Hammil, D. D., Brown, V. L., Larsen, S. C., & Wiederholt, J. L. (1994). *Test of adolescent and adult language* (Third edition). Austin, TX: Pro-Ed.

Husaim, D., & Tessler, M. (2006). The effect of speaking rate, experience, and lag time on Cued Speech transliterator accuracy. Unpublished honor’s thesis, University of South Florida, Tampa, Florida.

Jones, B. E., Clark, G. M., & Soltz, D. F. (1997). Characteristics and practices of sign language interpreters in inclusive education programs. *Exceptional Children, 63*(2), 257-268.

Kluwin, T. N., & Stewart, D. A. (2001). Interpreting in schools, a look at research. *Odyssey, Winter/Spring,* 15-17.

Krause, J.C. (2006). Personal communication.

Krause, J.C., Schick, B., & Kegl, J.A. (in press). "A version of the Educational Interpreter Performance Assessment for Cued Speech transliterators: Prospects and Significance. In *Cued Speech and Cued Language Development of Deaf Students* (Chapter 15). San Diego, CA: Plural Publishing, Inc.

LaSasso, C. J., & Metzger, M.A. (1998). An alternate route for preparing deaf children for BiBi programs: The home language as L1 and Cued Speech for conveying traditionally spoken languages. *Journal of Deaf Studies and Deaf Education, 3,* 264-289.

Leybaert, J. & Charlier, B. L. (1996). Visual speech in the head: the effect of Cued Speech on rhyming, remembering, and spelling. *Journal of Deaf Studies and Deaf Education, 1,* 234-248.

Magner, M.E. (1972). A speech intelligibility test for deaf children. Northhampton, MA: Clarke School for the Deaf.

Marschark, M., Sapere, P., Convertino, C., & Seewagen, R. (2005). Access to postsecondary education through sign language interpreting. *Journal of Deaf Studies and Deaf Education, 10* (1), 38-50.

Miller, G.A. (1956) The magical number seven plus or minus two: Some limitations on our capacity for processing information. *Psychological Review, 63*, 81-97.

Nicholls, G. H., & Ling, D. (1982). Cued Speech and the reception of spoken language. *Journal of Speech and Hearing Research, 25,* 262-269.

Park, W.E. (2005). Personal communication.

Pelley, K.A., Husaim, D., Tessler, M., Lindsay, J., & Krause, J.C. (2006). The effect of speaking rate and experience on Cued Speech Transliterator accuracy. Unpublished Poster Session, ASHA, Miami, Florida.

Reed, C. M., Delhorne, L.A., Durlach, N. I., & Fischer, S. D. (1990). A study of the tactual and visual reception of fingerspelling. *Journal of Speech and Hearing Research, 33,* 786-797.

Scheetz, N. A. (2001). *Orientation to Deafness* (Second Edition). Needham Heights, Massachusetts: Allyn and Bacon.

Schick, B., Williams, K., & Bolster, L. (1999). Skill levels of educational interpreters working in public schools. *Journal of Deaf Studies and Deaf Education, 4*(2), 145-155.

Schick, B. & Williams, K. (2001). Evaluating interpreters who work with children. *Odyssey, Winter/Spring*, 12-14.

Schick, B., Williams, K., & Kupermintz, H. (2006). Look who's being left behind: Educational interpreters and access to education for deaf and hard-of-hearing students. *Journal of Deaf Studies and Deaf Education, 11* (1), 3-20.

Smart, J. (2007). Personal Communication.

Strong, M. & Rudser, S. F. (1985). An assessment instrument for sign language interpreters. *Sign Language Studies, 49* (Winter), 343-362.

Strong, M. & Rudser, S. F. (1986). The subjective assessment of sign language interpreters. *Sign Language Studies, 53* (Winter), 299-313.

Tessler, M. (2007). Personal Communication.

Tope (2008). The Effect of Bilingualism on L2 Speech Perception. Unpublished Undergraduate Honors thesis. University of South Florida; Tampa, Florida.

Traxler, C. B. (2000). The Stanford Achievement Test, ninth edition: National norming and performance standards for deaf and hard-of-hearing students. *Journal of Deaf Studies and Deaf Education, 5:4* (Fall), 337-348.

Wandel, J. E. (1989). Use of internal speech in reading by hearing and hearing impaired students in Oral, Total Communication, and Cued Speech programs. Unpublished doctoral dissertation, Columbia University, New York.

Wilson, R. H. & Strouse, A. L. (1999). "Auditory measures with speech signals," *Contemporary Perspectives in Hearing Assessment.* Needham Heights, MA. Allyn & Bacon.

Winston, E. (1989). "Transliteration: What's the message?" *The Sociolinguistics of the Deaf Community.* C. Lucas Ed. San Diego, CA: Academic Press.

Appendices

Appendix A: Participant Information

Table A1

Cued Speech Receivers

	CS01	CS02	CS03	CS04	CS05	CS06	CS07	CS08
Age	27	30	39	34	25	28	20	33
Gender	M	F	M	F	M	F	F	F
Education	B.A.	Some college	B.A.	PhD	Some college	Some college	Some college	B.A.
Hearing level	Profound	Profound	Profound	Severe to Profound	Profound	Profound	Profound	Profound
CS Screening	100%	100%	100%	100%	100%	100%	100%	95%
TOAL-3 Percentile Score	95	95	98	84	75	75	84	75
First Language	English	English	English	English and French	English	English	English and ASL	English
Age first exposed to CS	2	5.5	3	2	2	10 months	1	3
CS use - Home	Y	Limited	Y	Y	Y	Y	Y	Y
CS use - School	Y	Y	Y	Limited	Y	Y	Y	Y
# Yrs CS	25	24	36	32	23	27	19	30
Pref. Comm. Mode	English (oral and cued)	No response	English (oral and cued)	Spoken English ASL	Cued Speech	Cued Speech	Cued Speech/ English	Cued Speech and Sign Language
Other Lang. fluent	Signed English	None	None	ASL	ASL	Sign Language	Spanish	None
Age exposed (other lang.)	18	N/A	N/A	17	17	19	12	N/A

Appendix A (Continued)

Table A2

Background and Scores (given 100% Accurate Stimuli) for Normal Hearing Listeners (Tope, 2008)

	Bilingual?	Education	TOAL-3 Percentile Score	All Word Intelligibility (%)	Key Word Intelligibility (%)
BL01	Y	M.A.	95	98.66	99.48
BL02	Y	Some College	84	93.23	96.62
BL03	Y	Ph.D.	84	97.65	99.22
BL04	Y	B.A.	84	99.26	99.87
BL05	Y	Some College	84	99.46	99.87
BL06	Y	B.A.	91	98.72	99.48
BL07	Y	Some College	63	95.50	97.92
BL08	Y	B.A.	98	99.53	99.74
ML01	N	Some College	95	99.40	99.87
ML02	N	Some College	84	99.19	99.87
ML03	N	B.A.	95	99.66	100.00
ML04	N	Some College	75	99.06	99.48

Appendix B: Number of clips per CST

	Ideal	Block 1	Block 2	Block 3	Block 4
CST1	19	21	19	19	17
CST2	19	16	14	21	16
CST3	19	20	16	13	16
CST4	19	23	22	25	22
CST5	19	19	21	20	20
CST6	19	17	19	18	16
CST7	19	17	17	14	16
CST8	19	21	25	21	26
CST9	19	21	19	20	19
CST10	19	18	25	24	25
CST11	19	20	21	17	20
CST12	19	18	13	19	18

Appendix C: Individual Data

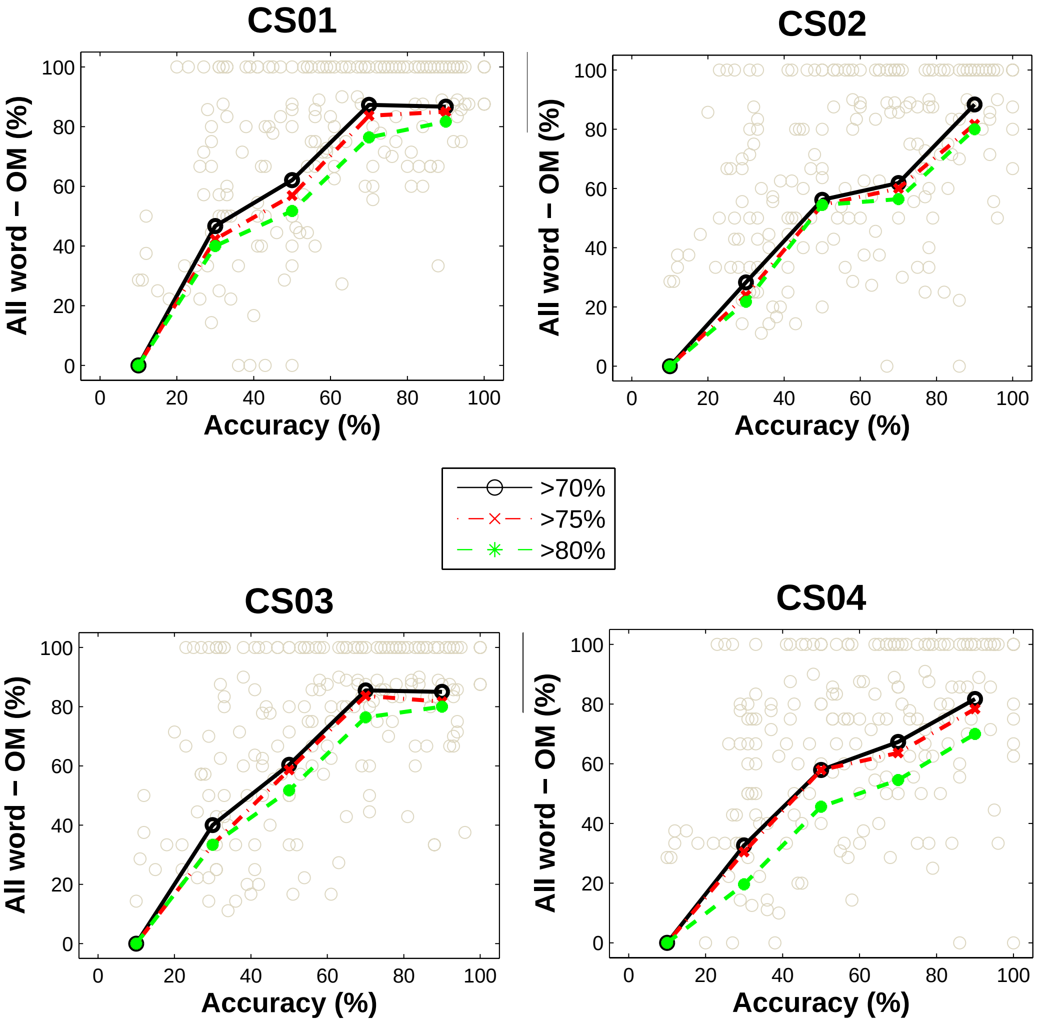


Figure C1. Accuracy-intelligibility likelihood functions plotted for each expert receiver, CS01 through CS04 (each showing the proportion of data points that reach 70% or higher intelligibility for a given accuracy interval).

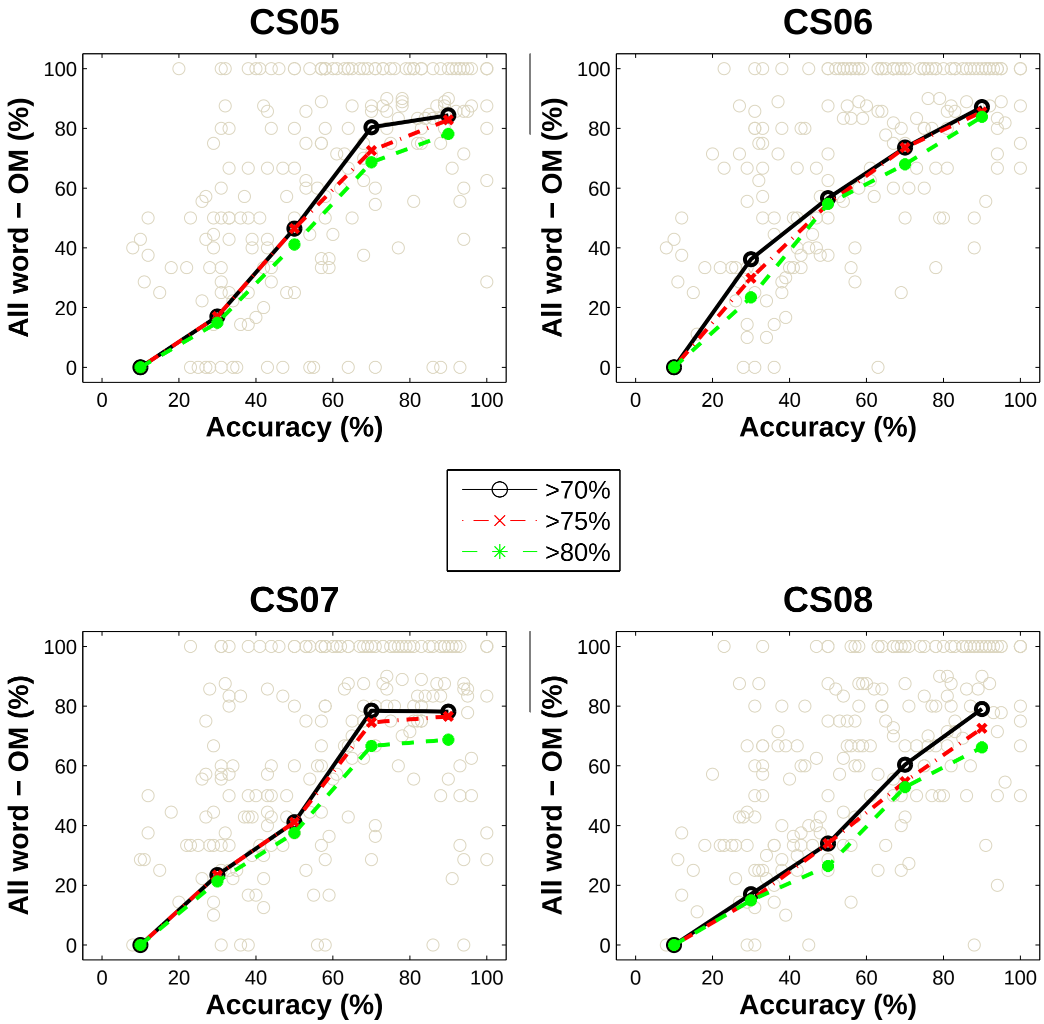


Figure C2. Accuracy-intelligibility likelihood functions plotted for each expert receiver, CS05 through CS08 (each showing the proportion of data points that reach 70% or higher intelligibility for a given accuracy interval).

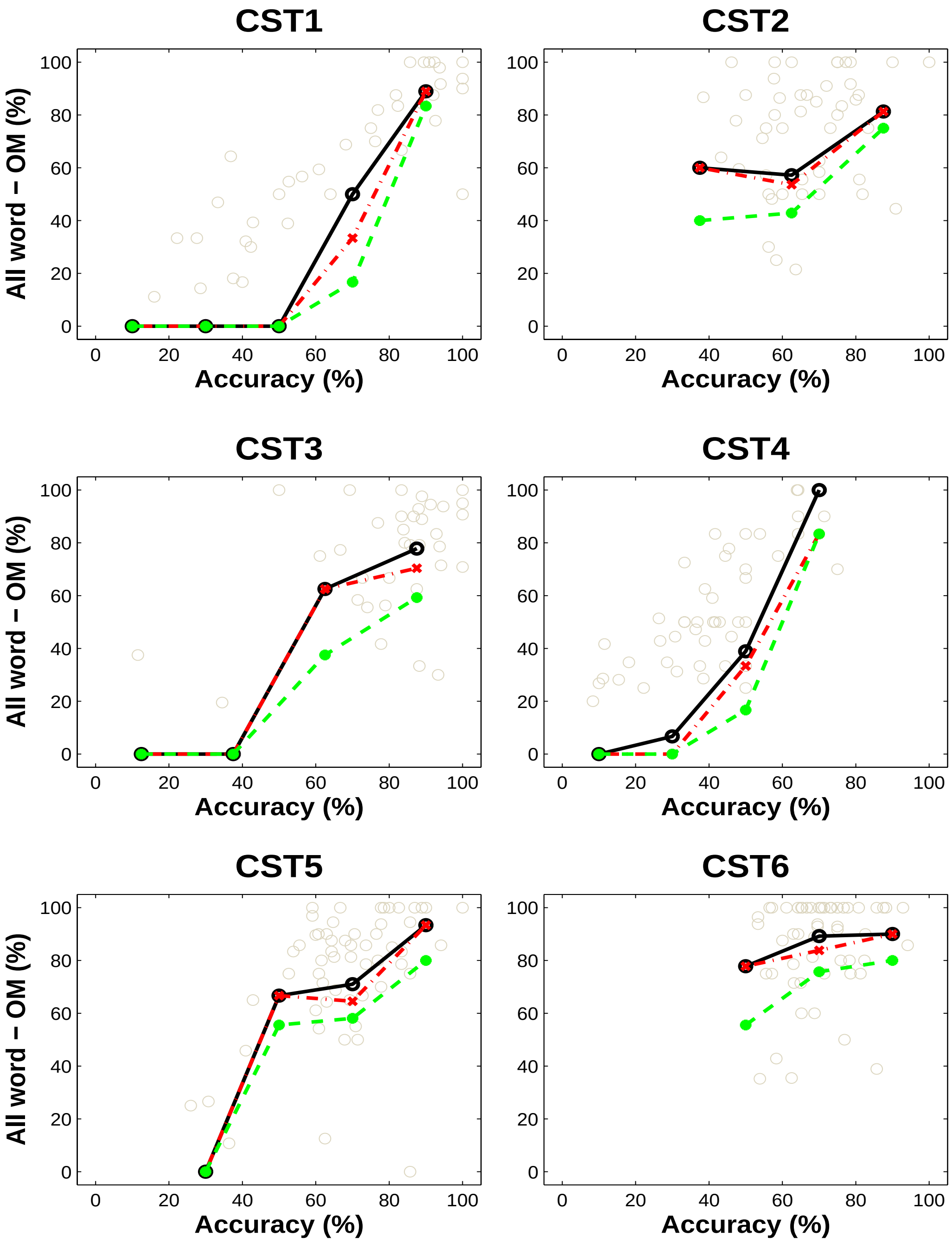


Figure C3. Accuracy-intelligibility likelihood functions plotted for each transliterator, CST1 through CST6 (each showing the proportion of data points that reach 70% or higher intelligibility for a given accuracy interval).

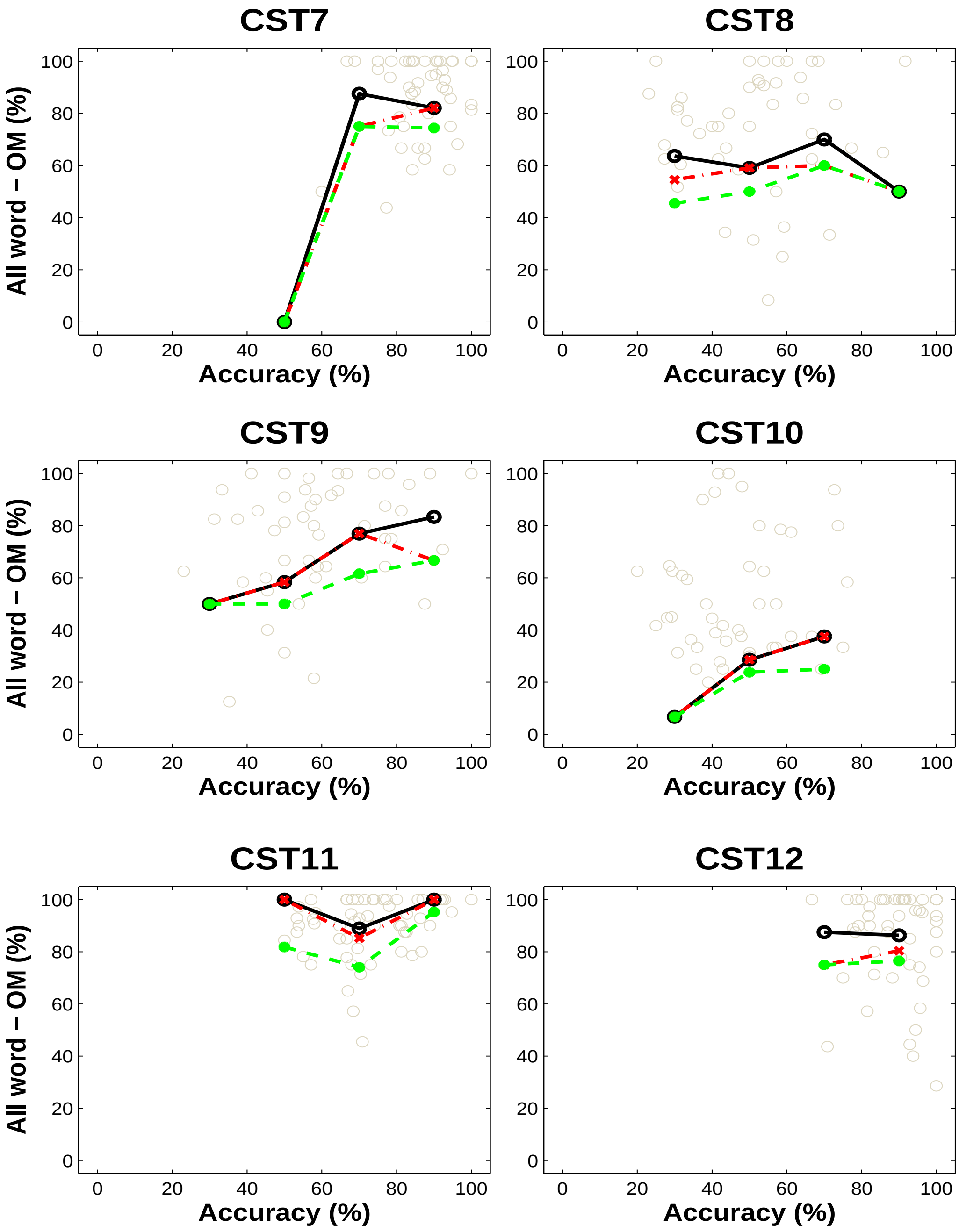


Figure C4. Accuracy-intelligibility likelihood functions plotted for each transliterator, CST7 through CST12 (each showing the proportion of data points that reach 70% or higher intelligibility for a given accuracy interval).