Published in final edited form as:

Speech Commun. 2011 March 1; 53(3): 379–389. doi:10.1016/j.specom.2010.11.001.

A Method for Transcribing the Manual Components of Cued Speech

Jean C. Krause^{*}, Katherine A. Pelley-Lopez, and Morgan P. Tessler

Department of Communication Sciences and Disorders, University of South Florida, Tampa, FL

Abstract

Designed to allow visual communication of speech signals, Cued Speech consists of discrete hand signals that are produced in synchrony with the visual mouth movements of speech. The purpose of this paper is to describe a method for transcribing these hand signals. Procedures are presented for identifying 1) the steady-state portion of the cue to be analyzed, 2) the cue's handshape, and 3) the cue's placement. Reliability is evaluated, using materials from 12 cuers that were transcribed on two separate occasions (either by the original rater or a second rater). Results show very good intra-rater and inter-rater reliability on average, which remained good across a variety of individual cuers, even when the cuer's hand gestures were heavily coarticulated. Given its high reliability, this transcription method may be of benefit to applications that require systematic and quantitative analysis of Cued Speech production in various populations. In addition, some of the transcription principles from this method may be helpful in improving accuracy of automatic Cued Speech recognition systems.

Keywords

transcription; coarticulation; cue accuracy; handshape identification; placement identification

1. Introduction

Speech communication in the visual modality is typically quite difficult (e.g. Erber, 1975; Grant et al., 1998). On average, recognition of words in sentences is less than 50%, even for highly practiced speechreaders (Auer and Bernstein, 2007). In response to this issue, a number of approaches for supplementing the visual speech signal have been proposed (Upton, 1968; Ebrahimi and Kunov, 1991; Massaro et al., 2009). One such approach, Cued Speech, is the focus of this paper.

Developed by Cornett (1967), Cued Speech calls for discrete hand signals to be produced in synchrony with the visual mouth movements of speech. As shown in Figure 1, the American English system uses eight handshapes to distinguish among 25 consonants and six placements to distinguish among 15 vowels. Handshapes and placements each represent multiple phonemes, but the set of phonemes assigned to a particular cue group are, by design, easily distinguishable via speechreading. As a result, experienced users of Cued

© 2010 Elsevier B.V. All rights reserved

^{*}Corresponding author. Tel: +1 813 974-9798; fax: +1 813 974-0822. jeankrause@usf.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Speech can obtain near-perfect reception of words (Nicholls and Ling, 1982) and everyday sentences (Uchanski et al., 1994).

To date, most work examining Cued Speech has focused either on its perception (e.g. Alegria and Lechat, 2005; Leybaert and Lechat, 2001) or its role in language learning (e.g. Leybaert and Alegria, 2003), while technological efforts have explored the feasibility of automatic Cued Speech generation, synthesis, and/or recognition systems (Cornett et al., 1977; Duchnowski et al., 2000; Attina et al., 2004; Gibert et al., 2005). However, all such work has assumed or required that the manual components of Cued Speech (i.e. the “cues”) be produced clearly and correctly. Yet not all cuers are likely to achieve this level of precision; Likert scale ratings assessing the performance of Cued Speech interpreters, for example, show some degree of variation in cue accuracy and/or intelligibility (Krause et al., 2008). Such ratings provide valuable descriptive information regarding the *quality* of cue production, but the *quantity* of cue errors typically exhibited by Cued Speech users remains largely unknown.

Before such quantitative data can be collected, however, it is first necessary to establish a reliable method for transcription of cues. Such a transcription method would not just be useful for quantifying cue errors and accuracy levels of individuals and populations of Cued Speech users (e.g. parents, interpreters, and other professionals who use Cued Speech); in all likelihood, it would be as valuable to the study of Cued Speech as the method of phonetic transcription is to the study of acoustic speech. Cue transcriptions could be used to identify common error types and their frequency of occurrence or to investigate more basic scientific questions related to Cued Speech. For example, with cue transcriptions, it would be possible to explore factors that affect accuracy of cue production (e.g. speaking rate, number of syllables, etc.), or to examine the effects of cue errors on receptive processing of Cued Speech. The insights gained from this type of research would not only increase our scientific understanding of Cued Speech production and perception, but could also lead to improvements in the practice of Cued Speech, perhaps in the form of new training methods informed by the research findings.

A major concern in collecting this type of quantitative data, though, is the reliability of the transcription method. Given the prescribed nature of Cued Speech, this concern may (at first) appear minimal. Indeed, transcription of cues and assessment of cue accuracy can be extremely straightforward – but only when cues reach target values. The process becomes considerably more complicated when the cuer's hand gestures are heavily coarticulated. Ensuring reliability in these situations is the focus of this paper. It requires either a technological solution (e.g. based on automatic cue recognition) or a well-specified procedure for performing transcriptions by hand. Although Cued Speech technology is rapidly improving, automatic recognition systems have not been tested on cuers whose manual gestures are heavily coarticulated. Moreover, recognition rates on clearly articulated cues are still well below 100% (Heracleous et al., 2009), which is insufficient for obtaining reliable measurements of cue accuracy. Therefore, a method for transcribing the manual components of Cued Speech by hand is required.

This paper describes one such method, for cases when the cuer's target word sequence is known. A lab manual containing additional pictures, detailed instructions, and examples is available at <http://uweb.cas.usf.edu/~jkrause/CS-Transcription.pdf>. Here, general procedures are explained in three major areas: 1) determining the video frame that contains the “steady-state” cue to be analyzed, 2) identifying the cue's handshape, and 3) identifying the cue's placement. In addition, recommendations are provided for classifying each cue transcribed as a correct production, a substitution, an omission, or an insertion. Finally, inter-rater and intra-rater reliability of the method are evaluated.

2. Terminology and Overview

In this paper, a method is described for transcribing the *articulated cue* sequence, or the sequence of cues actually produced by the cuer. For some judgments, that cue sequence is compared to the *target cue* sequence, or the sequence of cues required to represent the utterance appropriately in Cued Speech (i.e. the “expected” cue sequence). In so doing, each articulated cue is effectively judged for accuracy, or “scored” (classified) as a correct production, a substitution, or an insertion.

Although transcribing the articulated cue sequence can be quite involved, the target cue sequence is relatively simple to derive, especially when the cuer's exact words are known. Each word (or short phrase) has, at most, a small number of acceptable variations in the cue sequence, which reflect possible differences in dialect (e.g., /ɛg/ or /eg/ for “egg”) or pronunciation (e.g., /ðə/ or /ði/ for “the”)¹. Denoting the target cue sequence (and its acceptable variants, if applicable) in writing can easily be accomplished with cue notation (Cued Speech Association UK, 2009), which assigns a number to each handshape (1–8, see Figure 1) and a letter of the alphabet to each placement (“m” for mouth, “c” for chin, “t” for throat, “s” for side, “sf” for side-forward, and “sd” for side-down). For example, the target cue sequence for the word “ladder” would be denoted *6t1m*, indicating that the appropriate cue sequence for the word is handshape 6 at the throat placement (for /læ/) followed by handshape 1 at the mouth placement (for /dæ/).

Transcribing the articulated cue sequence can be similarly straightforward – if the cuer's hand makes clear transitions between steady-state cues, and if the cues themselves have well-specified handshapes and placements. If, however, the cues are formed poorly or coarticulated to the extent that one or both of these conditions is not met, then the reliability of the transcription and the corresponding cue accuracy judgments can be affected. In these situations, we propose that the following method be used to transcribe the articulated cue sequence and assess its accuracy.

This transcription method assumes that the materials to be transcribed were filmed using a high-quality digital video camera and that the transcriber can view the video recording in slow motion or frame-by-frame as needed. As long as these conditions are met, the transcription method is largely independent of the framing of the camera shot; measurements are normalized according to the size and location of the cuer's anatomical structures in the video recording (i.e. location of mouth corner, width of finger, etc.). To maintain precision in the measurements, however, a tight shot is recommended. One procedure for achieving a tight shot is to place the top of the cuer's forehead at the top edge of the video frame and then to adjust the zoom so that roughly two-thirds of the cueing forearm is captured for cues produced at the throat and side placements.

3. Transcription Method

Before an articulated cue can be transcribed, the steady-state portion of the cue must be identified. In heavily coarticulated materials, however, the duration of the steady-state can be very brief, especially relative to the video frame rate of most digital cameras (33 ms, or 30 frames per second). Because these brief (i.e. reduced) articulations are most perceptible when the phoneme sequence is expected (or similar to what is expected) from context, a two-part procedure is necessary to determine which manual gestures should be transcribed. The first sub-procedure is used to identify, preliminarily, the video frame containing the

¹See Cornett & Daisey, 2001 or <http://ne-cs.org/Art/> for a complete description of the procedure for determining cue sequences in American English.

steady-state portion of the next possible cue. The second sub-procedure is used to refine this judgment by evaluating the cue in context; the “cue under review” is examined in relation to the target cue sequence in order to determine whether it constitutes a perceptible cue that should be transcribed and scored (as a correct cue, substitution, or insertion), or whether it should be disregarded as a transitional gesture (i.e. an intermediate gesture between two cues). Each of these steps is described in more detail below.

3.1. Step 1: Identify Possible Cue for Inspection

The flow chart in Figure 2 provides an overview of the procedures involved in identifying, preliminarily, the video frame that contains the steady-state portion of the next possible cue. Each video frame is examined, and the procedure begins when any of the following conditions occur:

1. **Contact** – One or more finger(s) in the handshape are in contact with a placement.
2. **Constant location (steady)** – A consistent handshape and location are maintained for at least 100 ms (i.e. 3 or more video frames). Alternatively, a consistent handshape is maintained for 100 ms in conjunction with a change in direction of movement; very slight changes in position of the handshape's contact finger (typically the longest finger in the handshape, see Sec. 3.3) serve to triangulate a constant location over the course of 3 consecutive video frames, with no accompanying arm movement.
3. **Constant location (brief)** – A consistent handshape and location are maintained for at least 33 ms, but less than 100 ms (i.e. handshape appears in focus for 1–2 video frames).
4. **Change in handshape** – The handshape is changed to a different Cued Speech handshape, according to the specifications outlined in Figure 4, while none of the video frames containing the new handshape satisfy any of the first three conditions above.
5. **Change in placement** – The placement is changed to a different Cued Speech placement, according to the specifications outlined in Figure 5, while none of the video frames containing the new placement satisfy any of the first three conditions above.

In the first two cases, the immediate conclusion is that the cue under review has reached steady-state. It therefore passes to the next sub-procedure for further inspection, where it is always transcribed and scored (i.e. judged for accuracy). In the latter three cases, subsequent logic is applied to determine whether the (relatively) transient cue under review *could* represent a reduced steady-state gesture and thus merits further inspection, or whether it can be disregarded as a transitional gesture.

To make this determination, the handshape and placement of the cue under review are compared to the handshape and placement of the previous cue. If *both* differ from the previous cue, the cue under review passes to the next sub-procedure for further inspection. If *one or the other* differs from the previous cue, the cue under review also passes to the next sub-procedure for further inspection – unless the handshape or placement change is part of a transition gesture that will later reach steady-state. If *neither* the handshape nor the placement is different from the previous cue, the possibility that the cue under review is the second of two identical cues in a row is examined. In this case, a clear indicator that the cue is repeated (e.g. a flick, tap, or well-defined movement) is generally required for the cue under review to be passed to the next sub-procedure for further inspection; at the side

placement, however, the requirement that this indicator be present is waived when two repeated cues would be expected from context.

3.2. Step 2: Inspect and Transcribe (if applicable)

Figure 3 provides an overview of the inspection procedures used to determine whether the cue under review (identified in Step 1) is transcribed and scored (i.e. judged for accuracy), or disregarded as a transitional gesture. A cue under review that has definitively reached steady-state (by making contact with a placement or lasting at least 100 ms) is always transcribed and scored; otherwise, context (as specified by the target cue sequence) is considered to determine whether or not it should be transcribed. If context suggests that the (relatively) transient cue under review is either a correct production or a partially correct substitution, then it is transcribed and scored when either or both of the following conditions is met: 1) the correct handshape is achieved for any length of time (i.e. the cue may stop briefly at a placement or simply pass through it), or 2) the correct placement is achieved for at least 33 ms (1 video frame). If, on the other hand, context suggests that the transient cue under review is *not* part of the expected cue sequence, then it is generally disregarded as a transitional gesture – unless it is produced at an “established” placement, that is, at the exact same location as the previous cue or the next cue. In that case, the transient cue under review does not qualify as a transitional gesture (since the hand is not moving); therefore, it must be transcribed and scored.

3.3. Handshape and Placement Identification

When analyzing or transcribing the cue under review, identification of handshape and placement is typically straightforward; however, it can be difficult when the cue is heavily coarticulated or when the cuer's form impedes clarity. For those situations, the procedures used to resolve handshape and placement are summarized in Figures 4 and 5, respectively. Effectively, they establish boundaries between cue groups so that handshape and placement can be identified reliably, even when the cue under review is a reduced or intermediate gesture.

As Figure 4 shows, boundaries between handshapes are based largely on which fingers are considered “up” and which are considered “down.” This determination is made somewhat differently for different digits. For the thumb, it is based solely on whether or not the tip of the thumb is visible. For the fingers, a larger number of factors is taken into account. First, a finger is always considered down when half or more of it has been removed from view. Second, a finger is considered down, with one exception, when it is 1) retracted (i.e. flexed such that the fingernail is no longer visible or the two most distal knuckles are palmar to the middle finger) and 2) separated from the other (up) fingers in the handshape; the exception is when handshape 8 is a possibility. Third, even if it is not separated, the index finger is considered down any time it is retracted and handshape 5 is a possibility. In all other situations, a finger is considered up – except for the case when the index finger meets any of the “down” conditions while touching a placement (in which case it cannot be classified as up or down).

If it is difficult to determine whether a finger is up or down because of a visibility issue beyond the cuer's control (e.g. blurry video frame, shadow, etc.), then the up/down determination for the finger is made to maximize the accuracy score (i.e. the cuer is given the “benefit of the doubt”). If, however, the up/down determination is difficult because of a problem with the cuer's form, then the handshape under review is coded as handshape “9” (i.e. a non-intelligible handshape). A handshape 9 is also coded for any gesture that could easily be interpreted as two or more different Cued Speech handshapes. Otherwise, the handshape under review is identified as the Cued Speech handshape that it most resembles,

from among those handshapes which have the same fingers (or a subset of the same fingers) up.

Similarly, a placement under review is identified as the Cued Speech placement that its location most resembles. To make this judgment, boundaries between placements are used to establish an acceptable region for each placement. As Figure 5 shows, these regions are centered around the ideal location for each placement (specified by the National Cued Speech Association (1994)), with slight variations to region boundaries depending on whether or not the cue makes contact with (i.e. touches) the face or neck. When the cue does make contact, the finger that touches is used to judge the cue's location; otherwise, the "contact" finger for the handshape² (National Cued Speech Association, 1994) is generally used. The one possible exception is handshape 8, for which location is judged using a secondary contact point (the point midway between the two extended fingertips) instead of the contact finger if (and only if) doing so improves the accuracy score.

For cues that do make contact, differentiating the placements is relatively straightforward. If the finger touches below the chin, the placement is identified as a throat placement; if it touches above the bottom of the chin but below the lowest point of the lower lip, the placement is identified as a chin placement. Otherwise, the placement is generally identified as a mouth placement – except when it obstructs the mouth. Any cue that obstructs the mouth is coded as placement "X" (i.e. a non-intelligible placement). For cues that do not make contact, the same criteria are used, but a horizontal criterion is subsequently applied. Any cue at throat or chin level that is located more than one finger's width away from midline (toward the side placement) is identified as a side placement. Similarly, any cue at mouth level that is located more than one finger's width away from the corner of the mouth (toward the side placement) is also identified as a side placement. Otherwise, the original identification of placement (mouth, chin, throat, or "X") is retained.

4. Reliability

In order to assess the reliability of transcriptions produced using this method, a large set of cues produced by 12 different cuers was transcribed on two separate occasions, and agreement between the two transcriptions was analyzed. The second transcription was completed either by the original rater or a different rater so that both intra-rater and inter-rater reliability could be assessed.

4.1. Cued speech materials

The materials that were transcribed were elicited from 12 cuers who worked as classroom interpreters using Cued Speech. As shown in Table 1, the cuers varied widely, both in amount of experience as classroom interpreters (1 – 20 years) and in relevant professional certifications (4 with no certification; 6 with state certification; 2 with national certification). The 12 cuers were asked to transliterate³ a high-school level lecture about plants. Derived from the audio narration used in an educational film (Films for the Humanities, 1989), the 8-minute lecture (presented at a relatively slow speaking rate, 88 words-per-minute) required that each cuer produce roughly 1400 cues in order to represent the material correctly in Cued Speech.

²The National Cued Speech Association defines the contact finger, or "pointer" finger for a handshape as the finger that should be used to make contact with a placement. For handshapes 1–7, the designated contact finger is the longest extended finger in the handshape. For handshape 8, the contact finger varies with placement; it is the middle finger at the mouth and side placements and the index finger at the chin and throat placements.

³Transliterate is the preferred term when information is transferred between two forms of the same language - here, between spoken English and cued English.

Given the cuers' differing experience levels and the difficulty of the subject matter presented, the resultant materials varied widely in quality of Cued Speech production. As expected, many cuers produced cue sequences that were heavily coarticulated. Furthermore, the 12 cuers varied noticeably in both cue clarity and cue accuracy. As one measure of the variation, eight expert Cued Speech users, who were Deaf and accustomed to receiving information through Cued Speech, were asked to watch excerpts from each of the 12 cuer's materials and to indicate how comfortable they were with each individual's cueing. Composite quality ratings calculated from their responses (1.25 points for each "Very Comfortable" response, 0.625 points for each "Okay" response, and 0 points for each "Concerned I might miss something" response) confirmed a wide variation in cue quality across the 12 cuers. The ratings ranged from 0 (for two cuers who received a "Concerned" rating from all eight experts) to 9.375 (for one cuer who received seven "Very Comfortable" ratings and one "OK" rating) and were very nearly uniformly distributed between these two extremes.

4.2. Analysis and Results

To assess *intra-rater* reliability, 10% of the materials produced by each of eight cuers were re-transcribed by the same rater who initially transcribed the materials; approximately 1100 cues were re-transcribed in total. To assess *inter-rater* reliability, 20% of the materials produced by each of four cuers were re-transcribed by a different rater than the one who initially transcribed the materials; again, approximately 1100 cues were re-transcribed in total. In both cases, 1) the segment that was transcribed/re-transcribed was randomly selected (i.e., a random number generator was used to determine which target cue would be the starting point of the segment for each of the cuers), 2) the rater who re-transcribed the materials was blind to the first transcription, and 3) the second transcription occurred at least two months after the first transcription.

Reliability was then evaluated by determining how well the two sets of transcriptions agreed, both in number of cues and in their identity. For cue number, reliability was measured by calculating the proportion of cues that were in alignment across corresponding transcriptions (%-alignment). The number of cues in alignment was derived from the scoring information that accompanied the transcriptions (which linked the transcribed cues to the target cue sequence, categorizing each as a correct cue, substitution, omission, or insertion). For cue identity, reliability was measured by calculating the proportion of aligned cues that were in agreement (%-agreement). As shown in Table 2, intra-rater and inter-rater reliability was similar and very good for both measures. On average, 95% of cues aligned across transcriptions, and 91% of those cues were in agreement.

A close inspection of cues that were not in agreement revealed that the vast majority of disagreements differed by only one parameter. Specifically, 27% of disagreements differed only by handshape, 32% differed only by placement, and 29% differed only by movement, with only 11% of disagreements differing by two or more parameters. Given the 9% disagreement rate for aligned cues (see Table 2), the implication is that less than 1% of cues in the aligned transcription differed by two or more parameters. Thus, the corresponding transcriptions were in very good agreement.

4.3. Individual Differences

Table 3 shows the individual reliability measures for each of the 12 cuers whose materials were transcribed, along with their composite quality ratings (derived from evaluations by eight expert Cued Speech users, described earlier). Despite noticeable variations in clarity and accuracy (quality ratings ranging from 0.0 to 9.4 on a 10-point scale), individual differences in reliability were relatively small. All individual %-alignment measures were

within just 4 percentage points of the overall average (95%), ranging from 91% to 99%, which demonstrates very high reliability (in all 12 cases) for number of cues transcribed. Similarly, the 12 individual %-agreement measures were within 7 percentage points of the overall average (91%), ranging from 84% to 98%. Even though this range was slightly larger than the range for %-alignment, reliability of the %-agreement measure remained strong in all 12 cases. Moreover, the %-agreement values for the individual cuers were not related to their quality ratings (Spearman's $\rho = 0.1$, $p = 0.76$), which means that the identity of the transcribed cues was largely independent of cue quality. In other words, this transcription method can be reliably applied to large numbers of cuers with a wide variety of skill levels.

Given this level of reliability, a number of applications are possible. Perhaps the most straightforward one involves further analysis of the Cued Speech materials transcribed for this study. The 12 cuers represent a relatively specialized population (classroom interpreters who use Cued Speech), and it would be interesting to analyze their cue accuracy levels and error patterns, both as individuals and as a group. To do so, however, requires transcription of substantially more material. Even though more than 2200 cues were transcribed (and re-transcribed) for the reliability analysis, these transcriptions are based on short excerpts from each cuer's transliteration, often as short as 4–5 sentences (<200 cues). Such short excerpts are not necessarily representative of the individual's cue accuracy (i.e. accurate cuers sometimes have inaccurate cue sequences, and inaccurate cuers sometimes have accurate ones); therefore, transcriptions of longer samples are needed in order to analyze individual differences in cue accuracy or error patterns.

For our next study, we therefore plan to transcribe and analyze the entire set of Cued Speech materials collected from the 12 cuers, which consists of over 16,000 cues (~1400 cues $\times$ 12 cuers). With reliability of the transcription method established, we believe that this time-intensive undertaking can produce valuable information regarding cue accuracy levels and error patterns of interpreters who use Cued Speech. Indeed, some interesting patterns have already emerged from a preliminary analysis of the short excerpts transcribed for the reliability assessment. The group, on average, produced just 64% of cues correctly. Substitutions were the most frequent type of error at 20%, and the most frequent substitution was *4t* for *4s*, accounting for 27 of 406 substitutions (twice as many as the next most frequent substitution). These preliminary findings suggest that, at least in some situations, cue accuracy of classroom interpreters could be substantially improved. If this result is confirmed when we analyze the entire set of materials, a description of common error patterns could also be obtained from the transcriptions. This information may be helpful in the development of new training methods aimed at improving the practice of classroom interpreters who use Cued Speech.

5. Summary and Conclusions

When a cuer's target word sequence is known, the method proposed here is a highly reliable method for transcribing the manual components of Cued Speech. The method remains reliable across a variety of individual cuers, even when the cuer's hand gestures are heavily coarticulated. It should be noted, however, that transcriptions produced by this method describe only the manual components of the cued message. The corresponding mouth movements are assumed and are not explicitly evaluated during transcription. For a more complete transcription of the entire Cued Speech signal, it would be necessary to develop additional procedures for denoting mouth movements and the relative timing between mouth movements and cues. At present, however, we have opted to focus on the manual components of Cued Speech, as these are the components that are unique to cueing; mouth movements, which are also used in speech communication (i.e. without accompanying

cues), are typically well-practiced by talkers/cuers and tend to be less variable. Thus, the transcription method described here provides an important tool for any application which requires an analysis of the manual components of Cued Speech. For example, the method could be used to quantify cue accuracy levels in various populations (e.g. parents, interpreters, and other professionals who use Cued Speech) as well as to investigate factors that affect cue accuracy (e.g. speaking rate). In addition, some of the transcription principles outlined may be of benefit to systems aimed at automatic recognition of Cued Speech. For those who are interested in using this transcription method in whole or in part, a lab manual with further methodological details is available for download at <http://uweb.cas.usf.edu/~jkrause/CS-Transcription.pdf>.

Acknowledgments

The authors wish to thank Dana Herrington and Jessica Vick for many helpful technical discussions, and Joe Frisbie for the cue chart used in Figure 1. Financial support for this work was provided in part by a grant from the National Institute on Deafness and Other Communication Disorders (NIH grant number 5 R03 DC 007355).

References

- Alegria J, Lechat J. Phonological processing in deaf children: When lipreading and cues are incongruent. *Journal of Deaf Studies and Deaf Education*. 2005; 10(2):122–133. [PubMed: 15778209]
- Attina V, Beutemps D, Cathiard M-A, Odisio M. A pilot study of temporal organization in cued speech production of french syllables: rules for a cued speech synthesizer. *Speech Communication*. 2004; 44:197–214.
- Auer ET, Bernstein LE. Enhanced visual speech perception in individuals with early-onset hearing impairment. *Journal of Speech, Language, and Hearing Research*. Oct..2007 50:1157–1165.
- Cornett RO. Cued speech. *American Annals of the Deaf*. 1967; 112:3–13. [PubMed: 6043101]
- Cornett RO, Beadles R, Wilson B. Automatic cued speech. *Processing Aids for the Deaf*. 1977:224–239.
- Cornett, RO.; Daisey, ME. *The Cued Speech resource book for parents of deaf children*. National Cued Speech Association; Cleveland, OH: 2001.
- Cued Speech Association UK. How to write cues: Cue script and cue notation. 2009. http://www.cuedspeech.org.uk/uploads/documents/cuescript_and_cue_notation.pdf
- Duchnowski P, Lum DS, Krause JC, Sexton MG, Bratakos MS, Braida LD. Development of speechreading supplements based on automatic speech recognition. *IEEE Transactions on Biomedical Engineering*. Apr..2000 47:487–496. [PubMed: 10763294]
- Ebrahimi D, Kunov H. Peripheral vision lipreading aid. *IEEE Transactions on Biomedical Engineering*. Oct..1991 38:944–952. [PubMed: 1837008]
- Erber NP. Auditory-visual perception of speech. *Journal of Speech and Hearing Disorders*. 1975; 40(4):481–492. [PubMed: 1234963]
- Films for the Humanities. *Life cycle of plants*. 1989
- Gibert G, Bailly G, Beutemps D, Elisei F. Analysis and synthesis of the three-dimensional movements of the head, face, and hand of a speaker using cued speech. *Journal of the Acoustical Society of America*. Aug.; 2005 118(2):1144–1153. [PubMed: 16158668]
- Grant KW, Walden BE, Seitz PF. Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *Journal of the Acoustical Society of America*. 1998; 103(5):2677–2690. [PubMed: 9604361]
- Heracleous P, Aboutabit N, Beutemps D. Lip shape and hand position fusion for automatic vowel recognition in cued speech for french. *IEEE Signal Processing Letters*. May; 2009 16(5):339–342.
- Krause JC, Kegl JA, Schick B. Toward extending the educational interpreter performance assessment to cued speech. *Journal of Deaf Studies and Deaf Education*. 2008; 13(3):432–450. [PubMed: 18042791]

- Leybaert, J.; Alegria, J. The role of cued speech in language development of deaf children. In: Marschark, M.; Spencer, PE., editors. *Oxford Handbook of Deaf Studies, Language, and Education*. Oxford University Press; 2003. p. 261-274. chapter 19
- Leybaert J, Lechat J. Phonological similarity effects in memory for serial order of cued speech. *Journal of Speech, Language, and Hearing Research*. Oct..2001 44:949–963.
- Massaro, DW.; Carreira-Perpinan, MA.; Merrill, DJ. iglasses: An automatic wearable speech supplement in face-to-face communication and classroom situations. In: LaSasso, C.; Leybaert, J.; Crain, KL., editors. *Cued Speech and Cued Language Development of Deaf and Hard of Hearing Children*. Plural Publishing Inc.; 2009. chapter 20
- National Cued Speech Association. Guidelines on the mechanics of cueing. *Cued Speech Journal*. 1994; 5:73–80.
- Nicholls G, Ling D. Cued speech and the reception of spoken language. *Journal of Speech and Hearing Research*. 1982; 25:262–269. [PubMed: 7120965]
- Uchanski RM, Delhorne LA, Dix AK, Braida LD, Reed CM, Durlach NI. Automatic speech recognition to aid the hearing impaired: Prospects for the automatic generation of cued speech. *Journal of Rehabilitation Research*. 1994; 31(1):20–41.
- Upton HW. Wearable eyeglass speechreading aid. *American Annals of the Deaf*. 1968; 113:222–229. [PubMed: 5651479]

Figure 1.

Assignment of consonants to handshapes and vowels to hand placements in Cued Speech for American English.

Figure 2.
Flow chart summarizing procedures used to identify, preliminarily, the video frame that contains the steady-state portion of the next possible cue.

Figure 3.
Flow chart summarizing procedures used to determine whether a cue under review (identified in Step 1) is transcribed and scored or disregarded as a transitional gesture.

Figure 4.
Flow chart summarizing procedures for identifying handshape.

Figure 5.
Flow chart summarizing procedures for identifying placement.

Table 1

Classroom interpreter (transliterater) experience with Cued Speech (in years) and relevant certifications for each of the 12 cuers. TSC signifies the Transliteration Skills Certificate, issued by TECUnit, the national certifying body for interpreters who use Cued Speech.

	Experience	Certification
CST1	1	State-level
CST2	10	State-level
CST3	4	State-level
CST4	22	State-level
CST5	15	State-level
CST6	15	State-level
CST7	9	National-level: TSC
CST8	5	None
CST9	15	None
CST10	18	None
CST11	6	None
CST12	20	National-level: TSC

Table 2

Summary of reliability data. N represents the total number of cues between the two transcriptions, while M represents only those cues that could be aligned.

Type of Reliability	Cue number		Cue identity	
	N	%-alignment	M	%-agreement
Intra-rater	1156	95	1100	92
Inter-rater	1159	94	1090	91
Overall	2315	94.6	2190	91.1

Table 3

Composite quality ratings (10-point scale) and reliability measures for the 12 individual cuers. Composite quality ratings of accuracy and clarity were derived from ratings by eight expert Cued Speech users who were Deaf and accustomed to receiving information through Cued Speech. Range for each measure (across the 12 cuers) is listed in parentheses.

	Composite quality rating	Reliability	
		%-alignment	%-agreement
CST1	2.5	97	97
CST2	2.5	92	85
CST3	5.0	99	98
CST4	1.3	91	87
CST5	3.8	94	91
CST6	4.4	94	89
CST7	9.4	96	93
CST8	0.0	94	93
CST9	1.9	93	91
CST10	0.0	92	91
CST11	6.9	98	84
CST12	6.9	98	96
	(0.0 – 9.4)	(91–99)	(84–98)