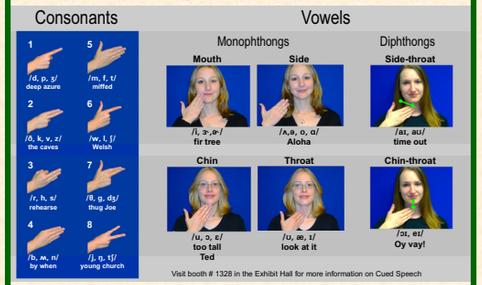
1aSC12. Transmission and reception of visual speech signals produced by Cued Speech transliterators

Katherine Pelley and Jean C. Krause (jkrause@cas.usf.edu)
Department of Communication Sciences and Disorders, University of South Florida, Tampa, FL

-Background-

Cued Speech (Cornett, 1967)

- Visual-only speech communication system used by some deaf individuals
- Hand "cues" are produced in synchrony with the mouth movements of speech
- Cues disambiguate visually confusable phonemes (i.e. visemes)
  - Eight handshapes represent groups of visually distinct consonants
  - Six placements represent groups of visually distinct vowels



- When used correctly, Cued Speech allow for near-perfect reception of everyday connected speech (Uchanski et al., 1994)

→ Some deaf people who use Cued Speech rely on **Cued Speech transliterators** as a means of accessing spoken information

-Accuracy vs. Intelligibility-

Cued Speech transliterator = an "interpreter" who uses Cued Speech

- Factors affecting interpreter intelligibility are unknown (Kluwin and Stewart, 2001)
- In investigating these factors, **two** aspects of the interpreted message must be considered:

- **Cued Speech** transliterators are attractive candidates for study
- One-to-one correspondence between spoken phonemes and cued phonemes means that both aspects can be easily quantified:
  - **Accuracy:** proportion of signal correctly ***transmitted*** by the transliterator
  - **Intelligibility:** proportion of signal correctly ***received*** by the deaf consumer

-Experiment 1: Accuracy-

Purpose

How does the accuracy of Cued Speech transliterators, measured by percent-correct cues produced, vary with:
- ***Speaking rate:*** slow, normal, fast
- ***Lag time:*** average delay between spoken and cued signals

Effect of experience was also examined

Participants

- Twelve (12) Cued Speech transliterators (CSTs), assigned to one of three categories based on level of experience
  - Two (2) **"novice"** CSTs - minimal or no certification and less than (the equivalent of) one year of experience
  - One (1) **"experienced"** CSTs - minimal certification with less than three years of work experience, or no certification with 3-5 years of experience
  - Nine (9) **"veteran"** CSTs - highest level of certification and/or more than five years of experience

Materials

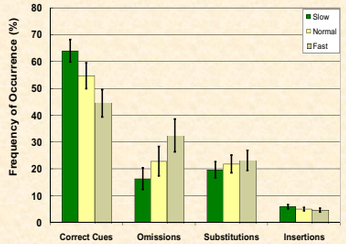
- Video recordings of the cued messages produced when each participant transliterated materials at three different speaking rates
  - Transliterators were presented with audio recordings of an 8th grade "lecture"
  - The lecture was presented in three segments, each at a different (conversational) speaking rate
    - **slow:** 88 wpm (speech expanded by a factor of 1.25)
    - **normal:** 109 wpm (original)
    - **fast:** 137 wpm (speech compressed by a factor of 0.8)
  - Speaking rate counterbalanced across segments

Procedures

- Transliterations were viewed in slow motion using Adobe Premiere Pro 1.5, and each cue produced was classified in one of four production categories:
  - **Correct cues**
  - **Omissions**
  - **Substitutions**
  - **Insertions**

Relationship to speaking rate

- On average, across all experience levels...
  - Correct cues: negative relationship with speaking rate
  - Omissions: positive relationship with speaking rate
  - Substitutions and insertions: no effect of speaking rate

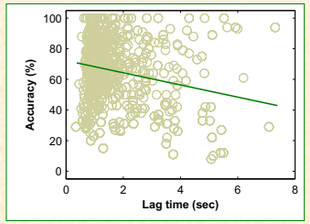


→ **Accuracy decline caused mostly by increased omissions**

- In addition, the negative relationship between accuracy and speaking rate was exhibited by all transliterators

Relationship to lag time

- **At the phrase level,** lag time is **inversely correlated** with accuracy (Spearman's rho = - .235, p<0.000)
  - However, it accounts for only 4% of the variance:



Overall Results

- **On average**, across all transliterators and speaking rates...
  - Correct cues occurred most frequently (54%)
  - Omissions were the most frequent type of error (24%)

Production category	Frequency of occurrence
Correct cues	54%
Omissions	24%
Substitutions	22%
Insertions	5%

- Accuracy of **individual CSTs** ranged from
  - 29% to 84% on average across rates
  - 40% to 89% at the slow-conversational rate

-Experiment 2: Intelligibility-

Purpose

How does the intelligibility of Cued Speech transliterators, measured by percent-correct words received, vary with:
- ***Accuracy:*** percent-correct cues transmitted
- ***Lag time:*** average delay between spoken and cued signals

Participants

- Eight (8) "expert" Cued Speech receivers
  - Profoundly deaf individuals with at least 10 years of experience using Cued Speech
  - Exposed to Cued Speech before age 10
  - Passed CS receptive screening (>90% reception of 5 sentences cued with 100% accuracy)

Materials

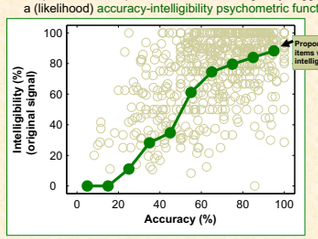
- Drawn from videos collected for Experiment 1
- ~2700 phrases excised from transliterator videos
- Only those elicited at the **slow-conversational rate** (~88 wpm) were considered (~900 possible stimuli, or ~75 clips per CST)
- Four stimulus blocks selected, such that
  1. The entire film narration (240 excised videos) could be presented phrase-by-phrase (in order) to each participant
  2. The range of accuracy scores was as well-distributed between 0% and 100% as possible

Presentation sessions

- Stimulus items presented **one phrase at a time**
- Stimuli periodically interspersed with excerpts from the original film, presented for context
- Each stimulus presented **once** on a computer monitor
  - Participant controlled pace via computer interface
  - Participant typed response **verbatim**

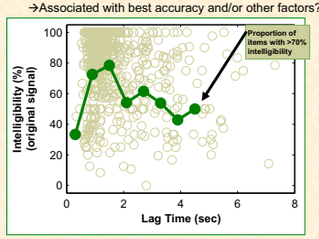
Relationship to accuracy

- **On average, intelligibility (72%) substantially higher than accuracy (54%)**
  - Also **higher for most (8) individual CSTs** (5 - 23 points)
  - Accuracy ~ equal to intelligibility for 4 CSTs:
    - Both novices (CST1, CST3)
    - Two veterans (CST7, CST12) – possible ceiling effect
- **For individual stimuli,** the relationship between accuracy and intelligibility is more variable. However, ...
  - Accuracy accounts for 26% of the variance in intelligibility and more when experience is controlled
    - Novices: 56% of intelligibility variance
    - Veterans: 28% of intelligibility variance
  - Proportion of data points with >70% intelligibility suggests a (likelihood) accuracy-intelligibility psychometric function:



Relationship to lag time

- **For individual stimuli,** the relationship between lag time and intelligibility is not linear. However, ...
  - The >70% intelligibility likelihood lag time - intelligibility psychometric function suggests an **optimal lag time**:
    - →**1-1.5 seconds**
    - →Associated with best accuracy and/or other factors?



Overall Results

- On average, across all receivers and transliterators...
  - 72% of all words in the original signal were received

Production category	Intelligibility
All words – original signal	72%
Key words – original signal	77%
Key words – transmitted signal	82%

- Intelligibility obtained by **individual receivers** varied up to 15 percentage points
  - 65% - 79% for original signal (all words)
  - 70% - 83% for transmitted signal key words
  - 74% - 89% for transmitted signal key words

-Conclusions-

- **Accuracy of "typical" CSTs is substantially lower than 100%**
  - Some highly experienced veteran CSTs are quite accurate
    - 2 veteran CSTs were above 85% at the slow-conversational rate
  - However, many "typical" (i.e. working, randomly selected) CSTs are not
    - 7 veteran CSTs: 40% to 73%, even at the slow-conversational rate
    - 3 less-experienced CSTs: 51% to 81%, with accuracy dropping markedly as speaking rate increased

  →Increased transliterator training and professional development opportunities should be created to address these issues in working transliterators
- **Accuracy plays a large role in intelligibility**
  - Accounts for 26% of the variance in this experiment
  - May account for more if accuracy measurements can be refined
    - Many substitutions are likely to be partially correct (e.g. right handshape, wrong placement)
    - No partial credit was awarded
- **Lag time also affects intelligibility** ("optimal" lag time is 1-1.5 seconds)
- **Other factors observed here that are likely to affect intelligibility:**

Transliterator factors
- Visual prosody
- Speechreadability
- Facial expressions and non-manual behaviors
- Cueing style: clear vs. conversational and highly co-articulated

Receiver factors
- Degree of reliance on speechreading vs. cuereading (making cueing errors easier or more difficult to tolerate)
- Current frequency of Cued Speech use

-Future Work-

- Refine accuracy analysis (e.g. partial credit for substitutions)
- Quantify effects of rate on intelligibility and accuracy x rate interactions
- Isolate and analyze other transliterator factors, such as speechreadability, that may also affect intelligibility
- Extend experiments to other visual communication options used by deaf individuals: Signing Exact English, American Sign Language, etc.
- Compare accuracy-intelligibility psychometric functions across communication options in order to
  - Increase understanding of intelligibility of visual signals
  - Gain insight into modality-independent aspects of perception

-Acknowledgments-

The authors wish to thank Morgan Tessler, Dana Husaim, and Jessica Lindsay for assistance in accuracy measurement and analyses, Wendy Park and Jane Smart for work on lag time measurements and stimulus preparation, Kendall Tope for assistance in stimulus creation, and Danielle Milanese for donated transliteration services used to develop practice items. In addition, we thank Catherine Rogers, Kelly Lamar Crain, and Patricia Blake-Rahter for helpful technical discussions throughout the process.

Financial support for this work was provided in part by a grant from the National Institute on Deafness and Other Communication Disorders (NIH Grant No. 5 R03 DC 007355).