Objective Data on Clear Speech: Does it Help in Training Audiology Students?

Boyce, S. E.*, Balvalli, S. N.*, MacAuslan, J. +, Clark, J. C.*, Martin, D.*

*University of Cincinnati, +Speech Technology and Applied Research Corp.

INTRODUCTION

Typical speakers instinctively use a "CLEAR" speaking style when they are instructed to "speak as if your listener is hearing impaired" or "speak as if your listener is not a native speaker of your language". CLEAR speech is more intelligible to hearing impaired listeners by about 17% (1, 2, 3). The ability to automatically detect differences between a speaker's ordinary speech patterns and their most intelligible speech, would clearly be helpful in clinical training and telemedicine applications.

Here we describe, a Landmark-based computer program (4, 5) to detect articulatory differences between "CLEAR" and "CONVERSATIONAL" styles of speech. Landmark-based speech analysis takes advantage of the fact that important articulatory events, like voicing, frication etc. show characteristic patterns of abrupt change in the speech signal. These patterns are detected by an automated computer system and assigned to a particular type of Landmark.

Most audiologists acquire the skill of 'clear speech' as a result of long experience. Typically, audiologists in training are instructed that CLEAR speech is both slower and louder than ordinary speech. In addition, acoustic analysis of naturally produced clear speech suggests that (among other strategies) speakers "sharpen" their articulatory precision to be closer to the canonical, phonemically distinctive shapes of words and to enhance perception. Experiments with clear speech to normal-hearing listeners and to hearing-impaired listeners has shown that neither slowness nor loudness is necessary for improved intelligibility to normal-hearing listeners (1, 3, 6). In contrast, several researchers have found that a slower delivery with greater pausing is disproportionately helpful to hearing-impaired listeners (7).

An objective method to train audiologists to speak clearly is necessary. We report results from a preliminary study of clear speech training for AuD students, using the Landmark-based software to provide objective measures of speech production correlated with intelligibility. The aim of the tool is to increase the rate at which audiologists in training acquire and deploy their clear-speaking skills with patients.

RATIONALE: Audiology students will improve their ability to communicate with patients if they receive intensive instruction on clear speech.

HYPOTHESIS: Instruction is more effective when feedback is based on objective measures of speech production correlated with increased intelligibility (the Landmark system) -

- number of Landmarks (LMs) (8)
- number of distinctly produced syllable clusters (syls)
- total duration of speech (duration)
- number of pauses (pauses)

METHODS

Methods: Automatic Landmark Analysis

We used a form of the Landmark analysis system based on (9, 10) that detects three types of Landmarks:

1. g: glottis. Marks the time when the vocal folds transition from not vibrating to freely vibrating (+g) or vice-versa (-g).
2. s: syllabicity. Marks sonorant consonantal releases (+s) and closures (-s).
3. b: burst. Designates frication onsets or affricate or stop bursts (+b) and points where aspiration or frication ends (-b) due to a stop closure. (Indicated from simultaneous abrupt changes in frequency bands.) These are never voiced.

The speech signal is automatically partitioned into 5 frequency bands plus voicing. Landmarks are identified as points where abrupt changes in the spectrum at particular frequency bands of a particular type coincide. Sequences of Landmarks that represent syllabic groupings are then identified and tabulated.

Measures based on Landmark Analysis:

i. Total Number of Landmarks per second of speech

ii. Total of time intervals between +b and +g Landmarks (≈ VOT)

iii. Principal components analysis of Landmark clusters corresponding to syllabic units (a measure of syllabic complexity). Note that a CVC such as "cab" or "pat" will show up as *(+b -b +g -g -b)* when the consonants are clearly articulated, but as *(+b -b +g -g)* when the final consonant is unreleased. A single vowel syllable will show up as *(+g -g)*.

Table 1 shows the tabulation of syllable cluster types. Some of the most common types are: *+g+s-s, +b+g-s-g, +g+s-s-g, +g-s-g-b, +s-s-g, +b-b+g+s-g.*

Syllable Type	Total Count	Mean Duration
+g-g	12	0.242
+g+s	1	0.49
+g-s	1	0.064
+s-g	3	0.144
+s-s	2	0.008
+b+g-g	2	0.081
+b+g+s	1	0.144
+b+g-s	1	0.176
+g-g-b	4	0.268
+g+s-g	18	0.337
+g+s-s	4	0.139
+g-s-g	7	0.22

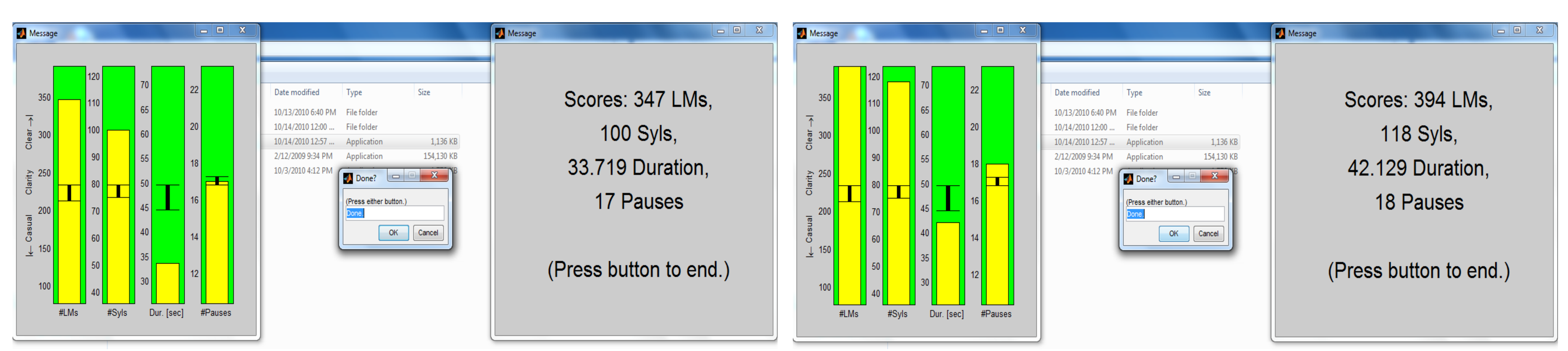
Table 1.* *Count of Syllable Cluster Types.

Methods: Present Experiment

PARTICIPANTS: Seven, 3rd year AuD students
MATERIALS AND INSTRUMENTS: BKB sentences were recorded (using a microphone) on a laptop which had the automatic computer program (the Landmark system)
PROCEDURE: The participants produced BKB sentences under two situations:

1. Conversational speech
2. "Clear" speech (as if talking to someone in the presence of background noise).

At the end of each recording, a graph appeared on the laptop screen providing information about the clarity of their speech in terms of LMs, syls, duration and pauses.



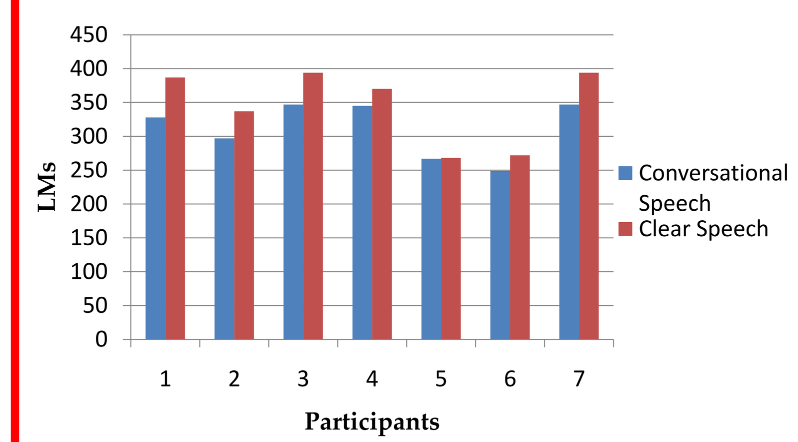
Example: Visual Display of Conversational and Clear Speech for one participant using the software (Proportions shown in the left-hand graph are derived from an independent study of 24 typical speakers of American English producing the BKB sentences in conversational vs. clear speaking styles)

Students were told that if their speech was clearer the second time, they should expect to see an increase in one or more of the above variables.

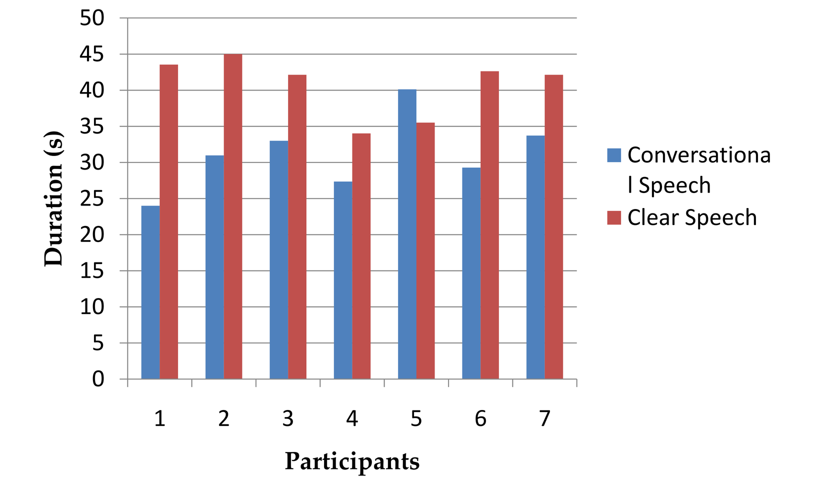
RESULTS

Wilcoxon Signed Rank test was administered [Level of significance = 0.05] and results indicated that the computer based program caused a significant increase in LMs ($T = 0$, $p < 0.05$) and duration ($T = 1$, $p < 0.05$) but not for pauses ($T = 4.5$, $p = 0.05$) and syls ($T = 3$, $p = 0.05$).

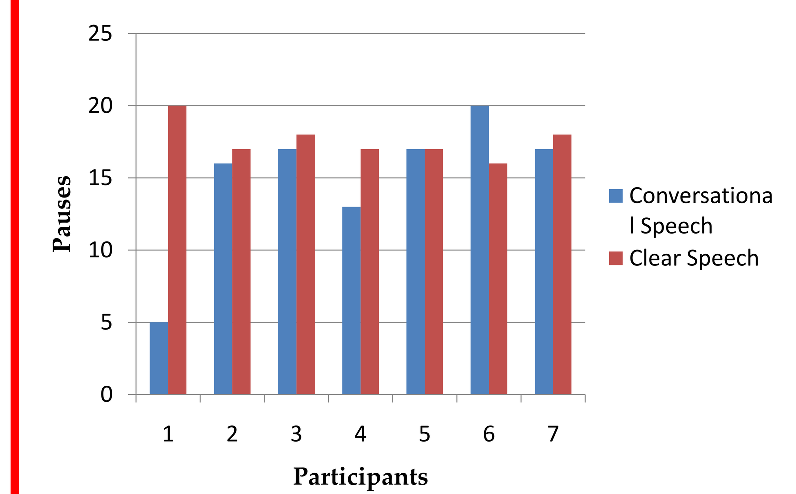
1. LMs:



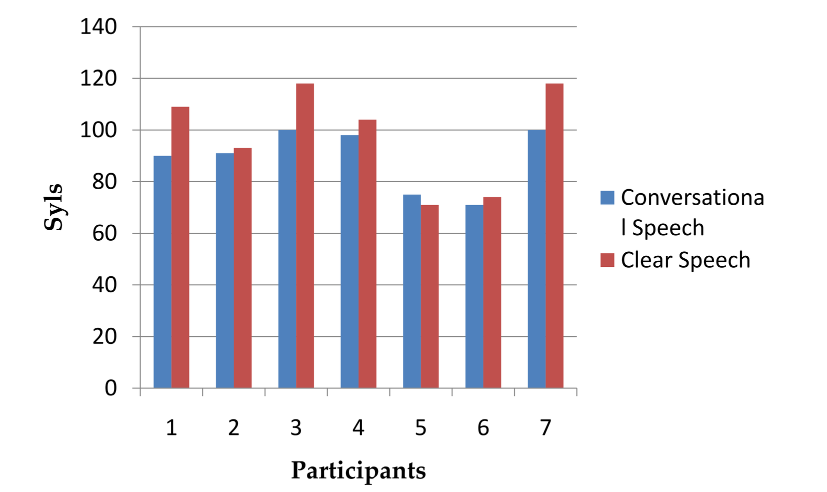
2. Duration:



3. Pauses:



4. Syls:



CONCLUSION

Previous research has indicated that the differences between CLEAR and CONVERSATIONAL speaking style can be automatically and reliably detected via a Landmark Analysis system. Results of this study are encouraging considering that the Landmark analysis system based software has the potential to training clinicians and student clinicians to speak clearly with patients. Future directions include expanding the speech materials used for training to common clinical situations that can easily be trained using a script, such as test instruction, explaining an audiogram, etc.

We envision that this device will not only be of great value in training audiologists, but also be very helpful in training SLPs, educators and significant others of individuals with hearing impairment, auditory processing disorders, speech impairments and learning disabilities.

REFERENCES

1.Shaw, B. & Krause, J. (2006). Clear Speech at Normal Rates: Intelligibility for Older Hearing-Impaired Adults. Paper presented at American Speech-Language-Hearing Association Convention.
2. Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. Journal of Speech and Hearing Research, 29, 434-446.
3. Krause, J. C, & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. Journal of the Acoustical Society of America, 115, 362-378.
4. Boyce, S., Bradlow, A., & MacAuslan, J. (2005). Landmark analysis of clear and conversational speaking styles. Paper presented at the150th meeting of the Acoustical Society of America, Minneapolis, Minnesota.
5. Boyce, S., Bradlow, A., MacAuslan, J. & Smiljanic, R. (2007). Automatic Detection of Differences between Clear and Conversational Speech. American Speech-Language-Hearing Association Convention, Boston.
6. Uchanski, R. M., Choi, S. S., Braida, L. D., & Durlach, N. I. (1996). Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. Journal of Speech and Hearing Research, 39, 494-509.
7.Uchanski, R. M. (2005). Clear Speech, In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of Speech Perception*, Blackwell Publishers, Malden, MA, p. 207-235.
8. Stevens, K .N. (1991). Speech perception based on acoustic Landmarks: Implications for speech production. Perilus XIV. Papers from the symposium, Current phonetic research paradigms: Implications for speech motor control.
9. Liu, S. (1995). Landmark detection in distinctive feature-based speech recognition. Unpublished Ph.D. dissertation. Cambridge, MA.
10.Stevens, K.N., Manuel, S.Y., Shattuck-Hufnagel, S., & Liu, S. (1992). Implementation of a model of lexical access based on features. In J.J. Ohala, T.M. Nearey, B.L. Derwing, M.M. Hodge, & G.E. Wiebe (Eds), *Proceedings of the 1992 International Conference on Spoken Language Processing. Edmonton: University of Alberta.*