Using landmark measure to evaluate effective clear speech

Suzanne E. Boyce1, Jean Krause1, Sarah Hamilton1, Rajka Smiljanic2, Ann Bradlow, Ahmed Rivera Campos1, Joel MacAuslan1

1University of Cincinnati, 2University of South Florida, 3University of Texas, 4Northwestern University, 5S.T.A.R. Corporation

Introduction
A number of studies have established that normal native speakers of a language know how to improve their intelligibility to listeners under intelligibility-challenging conditions. (Uchanski, 2005). This “Clear Speech” speaking style is significantly more intelligible to listeners; the average Clear Speech benefit is 15-17% to normal-hearing listeners in noise and to hearing impaired listeners in quiet (Uchanski, 2005). This is roughly the equivalent of a 5 dB improvement in signal/noise ratio (Uchanski, 2005). Many of these studies have reported that measures associated with articulatory “precision” are greater in Clear Speech, including: (1) an increase in vowel space, (2) increased consonant-to-vowel intensity differences, (3) stronger stop bursts, among others (Bradlow & Bent, 2002; Krause & Bradia, 2002, 2004; Smiljanic & Bradlow, 2005).

A consistent finding in studies of Clear Speaking style is that there are significant speaker-to-speaker differences (Ferguson, 2004) and that some speakers are more intelligible than others when producing Clear Speech (Bradlow & Bent, 2002; Krause & Bradia, 2004; Smiljanic & Bradlow, 2005). It is not immediately clear what distinguishes the better speakers, but one strong possibility is that these speakers produce more of the acoustic characteristics that distinguish Clear from Conversational Speech. The ability to detect when a speaker’s speech patterns are mostly likely to be intelligible would obviously be helpful in training clinicians, teachers and public safety workers to be more effective communicators.

In previous work, we have explored the use of the SpeechMark™ landMark-based computer program to detect the acoustic characteristics of Clear speaking style. Landmark-based speech analysis takes advantage of the fact that articulatory gestures, such as the onset and offset of frication, voicing, etc., show characteristic patterns of abrupt change in the speech signal (Stevens et al., 1992). These patterns are detected by an automated computer program (set to a specific threshold of change over time, and assigned to a particular type of landmark). The onsets and offsets of landmarks also allow for automatic detection of pauses, speaking time, and voicing time. We use a form of the landmark analysis system of Liu (1995) based on Stevens et al. (1992) that detects three types of abrupt landmarks plus landmarks corresponding roughly to the acoustic center of a vowel. The Speech signal is automatically partitioned into 5 frequency bands plus a voicing band. Abrupt landmarks are identified as points where the classifier has switched from not recognizing the particular landmark type to recognizing that landmark type. The points are then labeled with clusters that coincide in a specific pattern. These landmark patterns are identified by comparison of “coarse” and “fine” spectral detail.

Hypotheses
In this paper, we focus on the problem of detecting the best (the most intelligible) exemplars of Clear Speaking style.

Phase I Hypothesis: Clear and Conversational speech recordings known to differ in intelligibility are characterized by different patterns of landmark-based measures.

Phase II Hypothesis: The BEST Clear Speech speakers will show a more extreme version of the landmark patterns that differentiate Clear from Conversational Speech.

Methods and results

Phase I: The SpeechMark™ measurement system was applied to the five different collections of speech recordings described above (see Table 1). Three measures based on the SpeechMark™ system were computed for each data collection: (1) Total Number of Landmarks per Sentence, (2) Total Duration of Sentence, and (3) Total Number of Landmark Clusters corresponding to syllabic units in a Sentence (i.e., Syllabic Mark). Measure (3) is a rough measure of the number of landmarks per sentence that correspond to phonotactically possible syllables of English according to a set of rules that reflect distributional characteristics in the speech signal (Fall et al., 2002).

Table 1: Characteristics of Experiments Used for Analysis

<table>
<thead>
<tr>
<th>Reference article</th>
<th># talkers</th>
<th># listeners per talker</th>
<th>Total Number of Landmark Clusters (% of Stimuli)</th>
<th>Mean Number of Landmarks per Stimulus (SD)</th>
<th>Gain ± SEM in # Landmarks (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bradlow &amp; Bent (2002)</td>
<td>2</td>
<td>10</td>
<td>25±4</td>
<td>5±3</td>
<td>17±4 (94)</td>
</tr>
<tr>
<td>Krause &amp; Bradia (2002)</td>
<td>5</td>
<td>10</td>
<td>20±4</td>
<td>10±4</td>
<td>15±1 (96)</td>
</tr>
<tr>
<td>Smiljanic &amp; Brada (2008)</td>
<td>6</td>
<td>10</td>
<td>24±6</td>
<td>10±4</td>
<td>24±5 (94)</td>
</tr>
<tr>
<td>Smiljanic &amp; Bradlow (2008)</td>
<td>6</td>
<td>10</td>
<td>24±5</td>
<td>10±4</td>
<td>24±5 (94)</td>
</tr>
<tr>
<td>Boyce et al. (2011)</td>
<td>6</td>
<td>10</td>
<td>27±5</td>
<td>10±4</td>
<td>27±5 (94)</td>
</tr>
</tbody>
</table>

Table 2: Correspondence between intelligibility and landmark measures. # of landmarks. # of Syllable Clusters, and Total Duration of the Sentence, and intelligibility across all talkers of all productions described in Table 1. All measures shown plotted through in a standard error of the mean. **B & B refers to the 2 talkers of Bradlow & Bent, BKB refers to the 31 talkers of Krause & Bradia. In Phase II, 3 talkers of Smiljanic & Bradlow et al. and 4 talkers of Boyce et al. were chosen to produce high Clear Speech and 5 talkers of Smiljanic & Bradlow et al. were chosen to produce Low Clear Speech. In Phase II, the listeners were normal-hearing and the tests of intelligibility were performed in quiet.

FIGURE 1. Initial spectral analysis of an utterance: voicing (bottom) and five frequency bands’ energy waveforms.

FIGURE 2. Mean Number of landmarks vs. Mean Number of Syllable Clusters for all talkers (B & B) and for talkers of Smiljanic & Bradlow et al.

FIGURE 3. Mean Number of landmarks vs. Mean Number of Syllable Clusters for four talkers in Boyce et al. predicted to be best, median, and worst at producing Clear Speech, as compared to other talkers.

Discussion
Phase I: Table 2 shows the relationship between the intelligibility data and the landmark measures as, roughly, the amount of benefit we can assign to each of the landmark measures. Mean Number of Landmarks, Mean Number of Landmark Clusters, and Total Duration that talkers who show the greatest intelligibility gains in Clear Speech would produce more Landmarks and more Syllabic Clusters per sentence than any other talkers of the same study. (Note that talkers in the Krause & Bradia study were chosen to be particularly good producers of Clear Speech while talkers from the Boyce et al. study were chosen to range from "best" to "worst"). We plotted the two most significant SpeechMark™ measures, Mean Number of Landmarks and Mean Number of Syllable Clusters, against one another for the 17 talkers from all of the data collections for which we have listener intelligibility data. The talkers are divided into "best" vs. "other" groups by their % intelligibility ranking.

Phase II: Figure 2 and 3 show that SpeechMark™ measures parallel this difference in intelligibility. Both figures show separation between "Best" and "Other" talkers; Figure 3 shows that of the 4 talkers in the Boyce et al. study, the predicted "best" and "worst" talkers were most separated, while the predicted “middle” talkers lie in between. The pre-selected Krause & Bradia talkers showed the strongest Landmark pattern among the "best" talkers. In contrast, the three talkers from the Boyce et al. study who were selected to produce mediocre or ineffective Clear Speech show the lowest concentration of SpeechMark™ measures. Thus, this figure is probably representative of the range of variability in Clear Speaking style among people who have not been trained on Clear Speaking style or selected for natural ability.

Conclusion
Our conclusion is that Landmark system measures can reliably detect differences between Clear Speaking style and Conversational Speaking style. Further, these results give us confidence that the Landmark measures provide a reliable and accurate model of effective Clear Speaking style as produced by “best” vs. “worst” talkers.

References