Calibration of Target Word Embedded in a Carrier Phrase

Emily M. Chu[1,2] and Robert S. Schlauch[1]

1. Department of Speech-Language-Hearing Sciences
University of Minnesota, Twin Cities
2. Wayzata High School

Abstract

Scientists disagree over the approach for calibrating the level for words in standardized recordings of word-recognition tests. The same concern applies to monitored-live-voice presentations of speech in clinical settings. One approach is to have the talker produce calibrated levels for a carrier phrase (e.g., "you will say") and to let the target word fall naturally. The justification for this approach is that words vary naturally and predictably in their level. Another approach is to equate each word for its root-mean-square (RMS) level. This approach is preferred to control for idiosyncratic differences in production levels. To assess this issue, RMS levels for words and carrier phrases from an online source (NIST) for the Modified Rhyme Test were analyzed for 9 talkers. The entire utterance (carrier phrase plus target word) had equal RMS levels (within 1 dB). The detailed analysis of this sample of target words, which revealed significant inter-talker level differences that are larger than the differences predicted by speech acoustics. The strengths and limitations of each approach for setting the level of the target word will be discussed.

I. Introduction

Background:

- There is much debate over the proper procedure for calibrating stimuli for testing word recognition in a carrier phrase

1. Schlauch (2007) reports that many researchers equate words for their rms level in a word-recognition task. In the case of a word embedded in a carrier phrase, should the word or the entire phrase be used for calibration?
2. Some researchers believe "the carrier phrase is to help the talker maintain consistent vocal effort" and that "the most accurate calibration, would be on the carrier phrases without the target word (keyword). Calibrating on the carrier phrase with the target word as long as enough phrases so that the variations induced by the target words average out, is a second best but easier to use and therefore a more practical solution." (R. Mckinley, personal communication, February 5, 2019) This preserves the natural variability of word intensity due to speech acoustics, i.e. an /a/ sound will have a higher intensity than an /i/ sound.
3. Many audiology textbooks and other speech science reference materials state that the carrier phrase should be allowed to peak at zero on a VU meter (Berger, 1971), letting the target word fall naturally in relation to the carrier phrase. The target word should be allowed to peak where it will, as words vary in power (Martin, 1991).

Purpose of Study:

- To perform an acoustical analysis of a set of standard stimuli for the Modified Rhyme Test to investigate factors and characteristics of recorded stimuli that can affect the level of experimental control in a study
- An additional behavioral study was performed using two talkers with the largest rms level differences to examine potential behavioral correlates to acoustic calibration factors.

II. Method

- Words from the Modified Rhyme Test (MRT) (House et al., 1965) were used in the study, consisting of a corpus of 300 words grouped into 50 sets of 6 words each. All were CVC words
- Words in each group rhymed and only differed by one phoneme, making them members from a phonetically similar neighborhood (Luce and Pisoni, 1998)
- Stimuli created by the Institute of Telecommunication Studies (Institute for Telecommunications Sciences, 2015), using 9 different talkers [4 female (F1-F4), 5 male(M1-M5)]
  - Stimuli were in the form of carrier phrase + target word: "Please select the word [target word]."
  - All stimuli were calibrated based on the entire utterance, and the target word was allowed to fall naturally in relation to utterance
- Audacity ver 2.3.0 used to find RMS level in dBFS (dB full scale)
- 48 words were chosen from the corpus. Each talker was analyzed for RMS level of the 1) entire utterance (ALL), 2) carrier phrase and 3) target word.

Behavioral Pilot Study

- M1 and M5 were chosen as the stimuli for a behavioral study due to the large individual differences in average RMS levels for words (4.1 dB)
- The first author listened binaurally with headphones (Sennheiser, HD620) while seated in a double walled, sound isolated booth. The speech level was 65 dB SPL
- 100 words from each talker were presented at four signal-to-noise ratios (SNR). The masking noise was a speech-shaped noise representing the average of the long-term spectra of all the stimuli produced by M1 and M5 (see Figure 3 for their spectra)
- The tasks were programmed using MATLAB® (The MathWorks, Inc.), and a spreadsheet was generated with results as part of the program

III. Acoustic Analysis

Level of Target Words in Relation to the Entire Utterance

- There exist significant inter-talker differences in the RMS level of target words when equating the RMS levels of the entire utterance

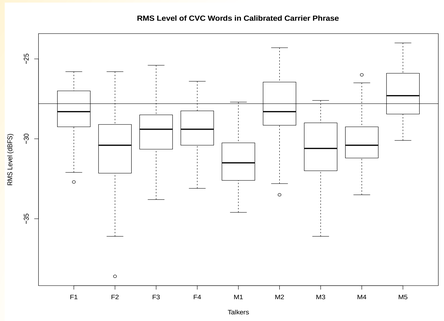
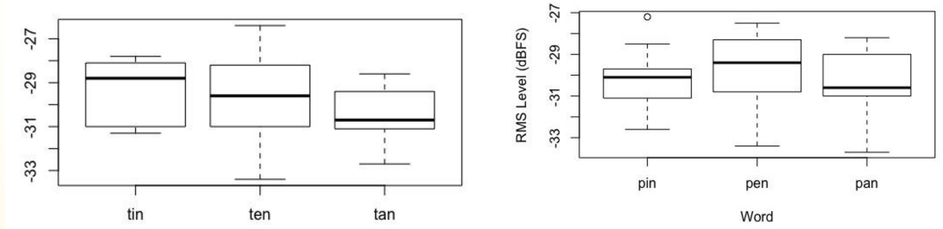


Figure 1.
- The RMS level of a sample of 48 words analyzed from the MRT speaker files (Institute of Telecommunications Sciences, 2015), analyzed for all 9 talkers
- The horizontal line represents the RMS level of the entire utterance, which is -27.8 dBFS.
- Inter-talker variability is large, with extreme cases showing very little overlap in their RMS levels for words (e.g., M1 and M5)

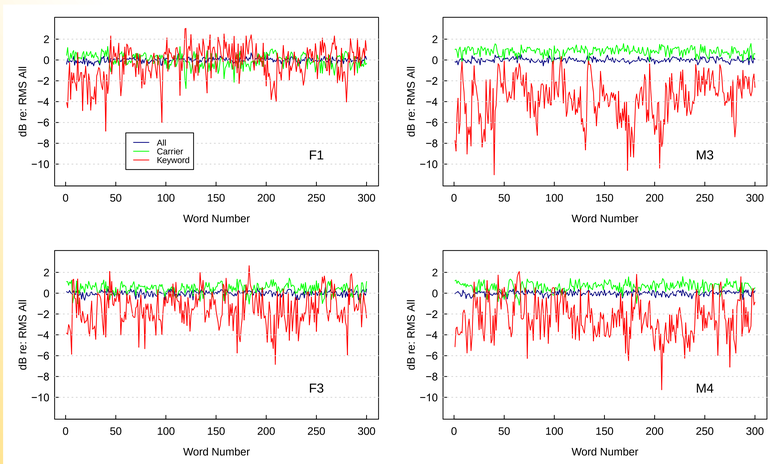
The RMS Level of Words Differing by One Phoneme: The Vowel

Figure 2: RMS levels for the triads of words were evaluated for all 9 talkers to show the relationship between vowel intensity when words only differ by one phoneme. The average differences for these vowel contexts do not follow the pattern in earlier studies, which are based on small samples of talkers (Blood, 1981; Lehiste & Peterson, 1959). What is notable is that the reported differences in those studies (/a/ 1.7 to 4 dB higher than /I/) is about the same magnitude as the inter-subject differences in Fig. 1.



The RMS Level of Each Target Word in Relation to Carrier Phrase and Entire Phrase for 4 Talkers

Figure 3.
- Data analyzed by Steve Voran from Institute of Telecommunications Sciences
- Four talkers depicted below, RMS level of each target word (keyword) analyzed as well as carrier phrase (Carrier) and the entire phrase (ALL)
- As depicted below, one talker (M3) has large variability in RMS levels for keywords.



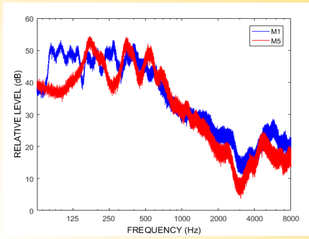


Figure 5.
- ✓ Performance functions (Probit) were fitted for SNR tests done using M1 and M5 as stimuli
- ✓ The large difference in performance between M1 and M5 was made even greater when taking into account the difference in intensity of the target words (middle panel)
- ✓ The RMS level of high pass filtered speech (ALL) at 1500 Hz equalizes performance for these 2 talkers

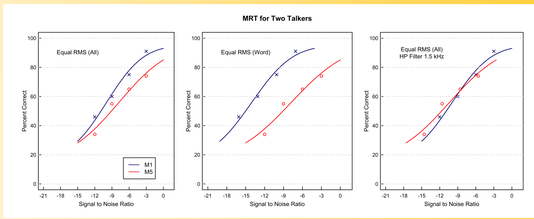


Figure 4.
- The long-term spectra for talkers M1, M5 for carrier phrase and target words (ALL).
- Their overall RMS levels are the same, but M1 has higher intensity when producing high frequency speech sounds. This difference as well as M5's slower speech rate are characteristics of clear speech (Uchanski, 2005).
- A close resemblance to clear speech can explain why M1 yielded higher scores for the same SNR even though M5 produced a higher RMS level for words.

IV. Discussion and Conclusions

- The relation between acoustic factors of speech and intelligibility is complex. More data need to be collected to learn if there are consistent acoustic correlates with behavior.
- We found evidence of large amounts of variability in level within and across talkers.
  - Differences across talkers were greater than predicted by differences across words due to speech acoustics
  - When differences in word level is extreme within a talker, does that affect the precision of the stimulus in detecting differences in listening conditions?
- The results of the behavioral study show that while RMS Level for words does not equate performance for the two talkers in our sample.
  - It may be prudent to examine the stimuli for clear speech characteristics, which significantly affect performance in SNR tests, as shown in the results for the behavioral study
  - Due to the importance for high frequencies for the MRT, a calibration method that examines the level of high-pass filtered speech may lead to more uniform performance across talker.

V. References

- Blood, G. W. (1981, June). The Interactions of Amplitude and Phonetic Quality in Esophageal Speech. *Journal of Speech and Hearing Research*, *24*(2), 308-312.
- Berger, K. (1971). *Audiological Assessment* (pp. 227-228). Englewood Cliffs, NJ: Prentice Hall.
- House, A. S., Williams, C. E., Hecker, M. H., & Kryter, K. D. (1965). Articulation-testing methods: Consonantal differentiation with a closed response set. *The Journal of the Acoustical Society of America*, *37(1)*, 158-166. doi:10.1121/1.1909295
- Institute for Telecommunication Sciences. (2015). Modified Rhyme Test Audio Library. Retrieved May 12, 2015, from https://www.its.bldrdoc.gov/outreach/audio/mrt_library/overview/index.htm
- Lehiste, I., & Peterson, G. E. (1959). Vowel amplitude and phonemic stress in American English. *The Journal of the Acoustical Society of America*, *31*(4), 428-435.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing*, *19(1)*, 1-36.
- Martin, F. N. (1991). *Introduction to Audiology* (4th ed., p. 140). Englewood Cliffs, NJ: Prentice Hall.
- Schlauch, R.S. (2007). "Calibration of speech in the real world: A re quest from industry," *Acoustics Today*, 43-44
- Uchanski, R. M. 2005. Clear speech. *The handbook of speech perception*, ed. by D. B. Pisoni and R. Remez, 207–35. Malden, MA/Oxford, UK: Blackwell.

VI. Acknowledgements

We would like thank Wayzata High School for providing the opportunity to work on this research project through the Honors Mentor Connection program. We would also like to credit Edward Carney for programming assistance and generation of multiple figures used in this presentation.

VII. Contact Information

Emily Chu
Email: chu00018@umn.edu