Acoustic and Temporal Predictors of Listening Passage Difficulty

Deirdre J. Derrick

Northern Arizona University

Abstract

Different test purposes (e.g., placement) require different configurations of item difficulty, and knowing which temporal features predict listening passage difficulty can aid test developers with passage selection and test development. Research on listening test difficulty has generally focused on textual features such vocabulary difficulty or grammatical structure. Although some studies have included intonation as a possible predictor variable for item difficulty (e.g., Freedle & Kostlin, 1996), it has been coded from transcripts and not from actual speech files. To explore the relationship between temporal and acoustic variables and listening passage difficulty, this study used listening passages from an Intensive English Program placement/exit test and the corresponding passage *p* values. It analyzed the passages for temporal features (e.g., rate of speech and number of pauses) and acoustic features (e.g., prominence), then used Pearson product-moment correlations to see which features were significant predictors of difficulty. Phonation time ratio emerged as the only significant predictor, with higher phonation times resulting in more challenging passages. Implications of these results are that the proportion of time spent speaking may be a more accurate measure of passage difficulty than words or syllables per minute.

Acoustic and Temporal Predictors of Listening Passage Difficulty

Much research has focused on segmental and suprasegmental features of non-native English speakers (NNESs) and how certain features predict comprehensibility (e.g., Kang, Rubin, & Pickering, 2010). This has implications in speaking assessments, where NNESs are judged on the basis of speech samples. Less work has been done in listening assessments, where NNESs are expected to comprehend the speech of native (and, increasingly, non-native) English speakers. As a result, we have an idea about which features of non-native speech affect comprehensibility, yet we have less of an idea of how features of native speech can affect listening comprehension in an assessment context.

Within the listening assessment context, studies have been conducted to examine predictors of test item difficulty (Freedle & Kostin, 1996; Nissan, DeVincenzi, & Tang, 1996). These studies coded features of the listening stimulus and the items and examined their effect on item difficulty (*p* values). Yet the coding was performed using the scripts rather than the actual recordings and did not include phonological features. As a result of these studies, we know that features such as the presence of negatives in items and infrequent vocabulary increase *p* values. What we do not know is which phonological features may correspond to higher or lower *p* values.

Literature Review

Second language (L2) pronunciation research has tended to focus on what factors influence the constructs of intelligibility, comprehensibility, and perception of accentedness. There are studies that look at these constructs using NS listeners and NNS speakers (e.g., Field, 2005; Kang, 2010; Trofomivich & Baker, 2006). Other studies have looked at NNSs'

understanding of other NNSs. Bent and Bradlow (2003) found a mutual intelligibility benefit between NNSs of different L1s.

A question still arises regarding which features of NS speech are challenging for NNSs' comprehension. Certain features might be inferred, though, from studies that look at perception and production, with the idea NSs may have difficulty producing features that they cannot perceive. For example, Derwing, Thomson, Foote, & Monro (2012) found that perception training for mandarin and Slavic language speakers lead to improvements in perception of sentence stress, intonation, and *–teen*/*-ty* numbers. There were, however, no improvements in perception of word stress and *can*/*can't*. These features, then, are some which may cause listening difficulty for NNSs, particularly in a more stressful assessment situation. Whether this difficulty may be enough to influence item difficulty scores remains to be tested.

Another strand of research is the examination of predictor variables in second language listening assessment research. Second language listening assessment research has tended to focus on which variables influence item difficulty, how different text types influence examinee performance, and how speech rate and accent affect comprehension. Nissan et al. (1996) coded TOEFL dialogues and items to determine which features had the greatest effect on item difficulty. They used the written scripts, however, so acoustical features were not coded for. Word frequency, utterance pattern (question-answer, etc.), presence of negative in stimulus, explicit-implicit information, and role of the speaker were all significant predictors of item *p* values.

Research has also shown that speech rate influences listening comprehension for NNS examinees, with comprehension declining as speech rates increase (Griffiths, 1990; 1992; Zhao, 1997). Brindley & Slatyer (2002), explored listening test variations, including speech rate and

repetition of stimulus. They included a faster speech rate of 200+ wpm and a normal speech rate of 188 wpm. They found that increasing the speed of the stimulus increased difficulty, but only for some of the items associated with the listening passage. Two (out of ten) items were easier in the faster version than in other versions. When examining the items and the stimulus, the researchers found that one item had high lexical overlap between the key and the stimulus. They concluded that there is a "complex interaction" of various task features (p. 387).

Research Questions

The current study seeks to determine which acoustical and temporal features may predict listening item difficulty. The research question that will be answered through this analysis is: Which acoustical and temporal features in listening test stimuli most strongly predict passage difficulty?

Methods

Listening Passages

This study used 15 listening passages. These passages have been used in multiple placement/exit exam administrations in an Intensive English Program (IEP) from spring 2011 to fall 2015.

Acoustical and Temporal Features

This section describes the temporal and acoustic variables included as predictor variables (Table 1). Variables for analysis were also chosen to reflect the nature of the speech samples as scripted speech recorded by native speakers. Speech rate has been shown to affect item difficulty (Griffiths, 1990; 1992; Zhao, 1997); accordingly, variables that would affect speech rate (including pauses and filled pauses) were selected for analysis. Variables related to prominence were also included, as it may be related to difficulty speaking and listening (Vanderplank, 1993).

The final list of acoustical and temporal variables that were coded for included speech rate, articulation rate, mean length of run, and phonation time ration. Also included were number of silent and filled pauses, and mean length of silent and filled pauses. Other variables were number of prominent words per run (pace) and the proportion of prominent words (pace), as well as pitch height and overall pitch range. In total, 12 variables were coded from the audio files.

Analysis

The passages were examined using PRAAT, a free computer program used for speech analysis. A one-minute excerpt was taken from each passage, and transcriptions of that except were prepared. Each excerpt was then analyzed to determine variables related to speech rate, pausing, and stress. Data were entered into SPSS and assumptions were checked.

The sample size did not meet requirements for a multiple regression, which are given as $n >= 104 + m$ for prediction, where n is the number of cases required for m number of predictor variables (Tabachnick & Fidel, 2013). Therefore, a series of Pearson correlations were run between the variables to explore the relationships. It was expected that measures of speech rate would be correlated, as would measures of pausing and measures of stress. The alpha value for statistical significance was set at .05.

Results

The first set of correlations were performed using speech rate variables, including syllables per second, articulation rate, mean length of run (MLR), and phonation time ratio, and with passage difficulty. Only one variable, phonation time ratio, had a statistically significant correlation of $r = -.591$, $p = .02$ with passage difficulty, with an R^2 of .349. The R^2 value indicates that nearly 35% of the amount of variance in passage difficulty can be explained by the

phonation time ratio of speech. Other correlations were between syllables per second and articulation rate, and between MLR and phonation time ratio.

The second set of correlations were between pause variables, including number of silent pauses, mean length of silent pauses, number of filled pauses, and mean length of filled pauses, and passage difficulty. The only statistically significant correlations here are those between the dependent variables (number of silent pauses with mean length of silent pauses, and number of filled pauses with mean length of filled pauses). None of the variables related to pauses approached a statistically significant relationship with passage difficulty.

Results of the correlations between stress measures (i.e., pace and space) and passage difficulty were not statistically significant, nor were the results of correlations between pitch measures (i.e., pitch range and overall pitch height).

Relevance to PIE and Second Language Assessment

The results of this study, while interesting, should be taken with caution. Given the limitations (i.e., sample size, use of passage rather than item *p* values) of this study, it would be unadvisable to alter assessment practices without further research.

Data showed a relatively strong relationship between passage difficulty and the phonation time ratio, with nearly 35% of the variation in passage difficulty being accounted for by the phonation time ratio. This indicates that as the percentage of time spent speaking decreases, the *p* value increases. That none of the pause and stress measures showed any relationship with passage difficulty should cause us to question our assumptions about what makes speech difficult for non-native listeners. Indeed, several of these assumptions are based on intuition rather than research. For example, most listening passages that are aimed at lower-level learners tend to not only be recorded more slowly than those intended for more advanced listeners, but they also

make use of exaggerated prominence and emphasis to highlight information. The results of this study, however, showed that neither pace nor space emerged as correlating with passage difficulty.

These findings, taken in conjunction, suggest that our intuitions about what makes passages difficult should be verified before being placed into practice. A wider range of acoustic and temporal variables should be examined. Phonation time ratio is related to speech rate, but it is not the same. Exaggerated prominence may not have the intended effect on examinees. These should be explored further. Answers to these questions can allow for greater refinement in assessments. Developers of placement tests, in particular, would benefit from a greater understanding of how different acoustic and temporal variables affect item and passage difficulty.

References

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *Journal of The Acoustical Society of America, 114*, 1600-1610.

Brindley, G., & Slatyer, H. (2002). Exploring task difficulty in ESL listening assessment. *Language Testing, 19,* 369-394.

Derwing, T. M., Thomson, R. I., Foote, J.A., & Munro, M. J. (2012). A longitudinal study of listening perception in adult learners of English: Implications for teachers. *Canadian Modern Language Review, 68,* 247-266.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39,* 399-423.

Freedle, R., & Kostin, I. (1996). The prediction of TOEFL listening comprehension item difficulty for minitalk passages: Implications for construct validity. ETS RR-96-29. Princeton, NJ: Educational Testing Service.

Griffiths, R. T. (1990). Speech rate and NNS comprehension: A preliminary study in time-benefit analysis. *Language Learning, 40,* 311-336.

Griffiths, R. T. (1992). Speech rate and listening comprehension: Further evidence of the relationship. *TESOL Quarterly, 26,* 385-390.

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System, 38,* 301-315.

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of English language learner proficiency in oral English. Modern Language Journal. Vol 94 (4): 554-566.

Nissan, S., DeVincenzi, F., & Tang, K. L. (1996). An analysis of factors affecting the difficulty of dialogue items in TOEFL listening comprehension. (TOEFL Research Rep. No. 51). Princeton, NJ: ETS.

Plonsky, L. (2015). Statistical power, *p* values, descriptive statistics, and effect sizes: A "back-to-basics" approach to advancing quantitative methods in L2 research. In L. Plonsky (Ed.) *Advancing quantitative methods in second language research*. New York, NY: Routledge.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston: Pearson.

Trofimivich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28,* 1-30.

Vanderplank, R. (1993). 'Pacing' and 'spacing' as predictors of difficulty in speaking and understanding English. *ELT Journal, 47,* 117-125.

Zhao, Y. 1997: The effects of listeners' control of speech rate on second language comprehension. *Applied Linguistics* 18, 49–68.