Investigating holistic measures of speech prosody through ratings of filtered speech

Aliel Cunningham

Department of English Language and Linguistics, Purdue University, USA

dacunnin@purdue.edu

Abstract

Prosody has been demonstrated to be a key component in second language acquisition and assessment. This study analyzes raters' ability to distinguish different levels of L1 and L2 prosody when listening to filtered speech. The goal of this study is to investigate how many levels of prosodic proficiency can be made when it is isolated in a filtered speech condition. An experiment was conducted with 45 L2 speech sound clips that had been recorded as a part of the OEPT (Oral English Proficiency Test). The test items used for this project included describing a graph in English and reading aloud an academic passage. These L2 English sound clips were produced by native Mandarin speakers at the novice, intermediate, and advanced levels of English proficiency. For control purposes, another 30 sound clips were recorded of L1 Mandarin and L1 English speakers completing the same task - each in their native language. All sound clips were filtered with a low pass filter of a 500 Hz cutoff in Praat 5.1.18. The results indicated a high correlation between the raters' scores in the filtered condition (listening primarily to prosodic components) and the original scores assigned by raters in an unfiltered condition. This reveals a strong link between the stages of second language prosody acquisition and overall L2 proficiency ratings. The results demonstrate that trained raters' can distinguish typologically distinct prosodic systems when listening to filtered speech, as well as make reliable judgments concerning the L2 prosodic proficiency level.
Index Terms: filtered speech, trained raters, second language prosody

1. Introduction

Speech prosody has become an increasingly crucial factor that is being considered in the field of second language acquisition – this is true both of stakeholders in second language assessment and pedagogy. Studies have shown that prosody plays an undeniable role in L2 English comprehensibility. Pickering (2001) found that prosody plays a pivotal role in effective communication for ITAs in a classroom setting. Another study that looked at the use of intonation and ITA proficiency judgments was conducted by Wennerstrom (1998). She found there was a strong relationship between the holistic score of speaking proficiency and intonation patterns – those students whose intonation pattern closely paralleled native English speakers also had higher proficiency scores. While a few of the ITAs (Mandarin speakers) had acquired these aspects of English intonation very well, many of the ITAs appeared to be transferring their L1 intonation system to varying degrees in their oral English communication. This attempt to transfer prosodic patterns found in L1 and to apply them in L2 utterances has shown up in other cases as well. Cutler, Dahan & Donselaar (1997) observe that "…given the opportunity listeners will apply their native language-specific procedures to foreign language input, even in cases where the procedures may not operate efficiently at all. The French listeners apply syllabic segmentation to English words…(Cutler et al., 1986);…and Japanese listeners apply moraic segmentation where possible to English input (Cutler & Otake, 1994). Nguyen, Ingram & Pensalfini (2008) found consistent patterns of transfer–both at the phonetic and phonological level of L2 learners whose native language was Vietnamese. The results of their study suggested that the L2 learners maintained a "paradigmatic tonal pattern where a lexical tone is preserved for each syllable". This tendency is not exclusive to learning English. In fact, L1 English speakers are guilty of the same types of errors when learning other languages. Beckman (1996) showed that L2 Japanese learners whose native language is English tend to "misinterpret the long closure of Japanese geminate stops as pauses, and consequently postulate strong syntactic boundaries occurring at what are actually word-internal consonant sequences." One possible reason why prosody tends to be transferred so readily from L1 may be because it has been shown to be one of the first linguistic elements to be acquired in our first language (Moon, Cooper, R., & Fifer, W. 1993). Shukla, White, & Aslin (2011) demonstrated that 6 month old infants could utilize prosodic phrases to identify word boundaries. This early acquisition of prosodic patterns is both a phonetic and phonological reality. Nazzi, Bertoncini, & Mehler (1998) demonstrated that infants were also able to distinguish between language-specific rhythmic typology – such as syllable-based versus stress-based prosodic systems. How ingrained the L1 prosodic system remains when learning a second language can be related to the age of the learner when they are exposed to a competing prosodic system. Guion

(2005) found that the age of L2 exposure and acquisition affected her Korean participants' ability to discern stress patterns in English. She hypothesized that this was due to the fact that they were using their L1 prosodic parameters of tone patterns mapped onto the Korean accentual phrase and had not acquired the English phonological categorization of stress accent.

1.3. Holistic Ratings vs. Acoustic Measurements

While the acquisition of L2 prosody has been an area of in depth investigation over the past decade, it has been mostly targeted indirectly through measurements of speech rate, mean length of run (MLR) or frequency of pauses. While these measurements have been shown to have strong correlations with fluency scores, they do not capture prosody as such and may not be the best way to assess L2 prosodic proficiency. Derwing, Rossiter, Munro, & Thomson (2004) looked at the ability of untrained raters to rate fluency in relation to temporal measures such as speech rate, MLR, and number and duration of pauses. While these individual components have been highly correlated with holistic fluency ratings it is not clear how exactly they relate to a system of prosody. These two terms can be distinguished by their relation to language -- Fluency is not a component of language itself, but rather a function of how well all linguistic elements are being composed together in a fluid flow of speech. Prosody, on the other hand, is a component of language and differs in its parameters and functionality from language to language. We should not mistake one for the other or assume we are assessing the one because we are able to isolate indirect correlates of the other. While fluency is a composite picture of several components working in tandem, prosody is a complex and interrelated system through which linguistic phrases are segmented, colored, organized, and highlighted. A purely quantitative description of this process is inadequate and does not give us a complete picture of an integrated system at work. In an attempt to capture a more complete picture of prosody and its essential elements, some studies have explored using a low pass filter on speech samples (Munro, 1995; Nazzi, et al. 1998). The low pass filter is set at a frequency cutoff (often between 300 Hz and 600 Hz) and all speech signals that are filtered by this method will have most frequencies above that cutoff significantly attenuated – thus isolating the lower frequencies for examination. In theory this separates those linguistic components that are more segmental in nature (usually fully formed with the addition of higher frequencies) from those components that are more continuous – highlighting the rhythm, speed, and alternations in vowel quality. The resulting effect is sound that has been muffled – as though listening to a conversation through a wall. Once in this filtered condition, we have a unique sound quality that is mostly made up of prosodic features rather than segmental elements. Some would argue that prosodic information without content is of little use for assessment purposes. However, raters who listen to this filtered sound have demonstrated the ability to assess different levels of "accentedness" when given speech sound clips in the filtered condition (e.g. Munro 1995) In the past this filtered stimuli has been looked at to see if certain distinctions could be made with regard to accentedness or language type, but no study has made use of this filtered signal to investigate the different stages of L2 prosodic proficiency. In this study we are investigating two main questions with relation to the filtered stimuli:

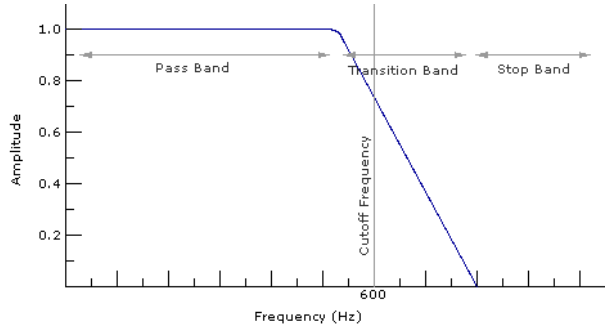

Figure 1: ***Low pass filter with a cutoff of 600 Hz***

Can trained raters make distinctions between typologically distinct prosodic systems when listening to filtered speech?

Can trained raters reliably distinguish among various levels of L2 prosodic proficiency when listening to filtered speech?

A study conducted by Iwashita, Brown, McNamara, and O'Hagan (2008) looked at a similar issue when investigating how many distinct levels of speaking proficiency could reliably be made. In this study, they compared holistic scores given by raters and temporal measures. By contrast, in this study, we are mainly focused on raters giving holistic scores – in filtered and unfiltered conditions.

2. Method

To assess the above research questions, we employed raters who have been trained to rate unfiltered speech on a holistic scale for the OEPT (Oral English Proficiency Test). This assessment is given to international graduate students coming into Purdue University for the first time whose native language is not English. The OEPT is a unique, multi-faceted exam with 12 different items requiring examinees to respond to various prompts such as a newspaper article, a bar chart, a graph, a voicemail, a video of a conversation, an academic passage, etc. (for more information on the OEPT layout see – www.oepttutorial.org). The responses to these 12 items are recorded as sound clips and stored online so they can be assigned to 2 trained raters who have been trained to use a holistic rubric ranging from a score of 35 (novice) to 60 (very

proficient) with 6 distinctions in all (35, 40, 45, 50, 55, 60). For this study, 5 raters whose L1 was English and had been trained using the OEPT listened to filtered stimuli to rate each sound clip for its level of prosodic proficiency. The stimuli used in this project were of two varieties – a reading item (utilizing a passage of academic English) and a spontaneous description of a bar chart (the graph represented the salaries of various job positions at an American university). The rationale for selecting the academic reading passage was to have one item in which the same text was utilized by all the speakers. This would eliminate prosody differences that could be attributed to genre, topic, or other external factors related to the prompt. The second item (bar chart) was selected because it would generate spontaneous speech with a common prompt. Since we are more interested in the prosodic patterns in spontaneous speech as opposed to that produced in a more formal reading style, this bar chart item was the more crucial of the two prompts utilized. We selected our speech samples from a database of OEPT exams that have been made available for academic research. We selected both male and female speakers who had Mandarin as their L1 maintaining about a 50/50 ratio. Then we choose an even distribution of exams from the beginner level (35-40), intermediate level (45-50), and advanced level (55-60). To establish control groups at either end of the scale we recruited L1 English and L1 Mandarin participants to respond to the same academic reading passage and bar chart (translated in the case of the L1 Mandarin speakers). All of these sounds clips were filtered at 500 Hz with a pass Hann band filter in Praat 5.1.18 (if x < 500 then self else 0 fi) with a smoothing rate of 50 Hz. Most of the raters had been part of a pilot study that introduced them to the sound of filtered speech so there would not be a significant learning effect during the actual study. The raters were given a form to assign a score to each unique item ID (each ID had a reading and a bar chart item).The raters were given a scale that ranged from 1-9 with these specifications: (1) L1 Mandarin Prosody, (3) Novice L2 English Prosody, (5) Intermediate L2 Prosody, (7) Advanced English Prosody, (9) L1 English Prosody. Before they began rating, each rater was asked to listen to a benchmark recording of L1 Mandarin Prosody and L1 English Prosody. Each rater also recorded (in a separate document) what they heard that caused them to assign that particular score. In total the trained raters listened to 75 sound clips in 4 sets (which had been grouped to contain an even distribution from the 5 different categories).

3. Results

In order to answer the first of our research questions the results were compared first with the holistic scores given in their unfiltered condition. In analyzing the results, we did both a mean comparison of the trained raters' scores listening to the filtered stimuli versus the original scores given by raters in the unfiltered condition as well as looking at the Spearman and Pearson correlation coefficient range as it relates to the holistic scores.

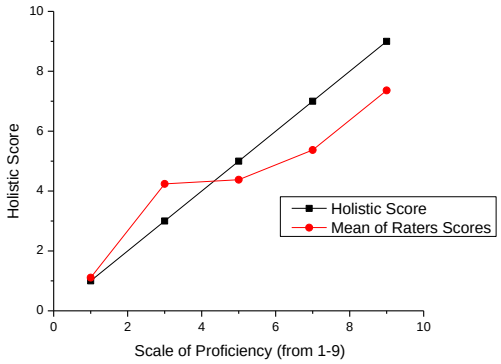

Figure 2: ***Trained raters versus holistic scores***

The statistical analysis showed that there was a significant correlation between the 5 raters' scores and the original holistic scores ranging between coefficients of 0.68 to 0.88. The mean of these scores tracked fairly closely with the holistic scores as can be seen in the graph above. Another interesting finding was the almost unanimous agreement among raters concerning what constituted "L1 Mandarin Prosody" (which was ranked on the scale as a "1"). The results revealed only 1 sound clip did not have unanimous agreement across all raters resulting in a mean rating score of 1.1. The clear distinction that raters made between L1 Mandarin prosody and English prosody was confirmed by the fact that there was only one instance in 75 clips where a (1) score was assigned to a speech clip that was actually not Mandarin. Interestingly, this did not happen on the opposite end of the scale when evaluating L1 English prosody. The raters tended to avoid assigning the highest score of (9) "L1 English Prosody" -- only assigning it 8% of the time in their composite ratings.

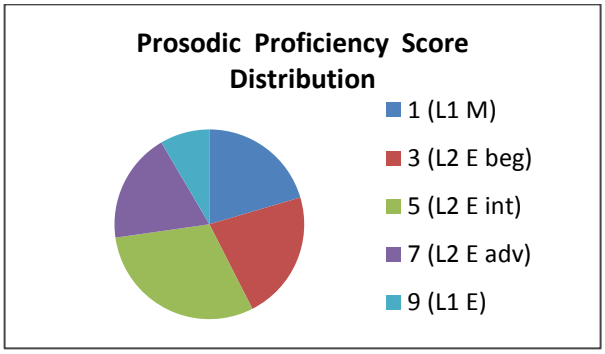

Figure 3: ***Distribution of scores by trained raters***

One possible reason for this could be that some of the L1 English speakers were older participants and this may have interacted with the quality of their prosodic production. Despite this avoidance at the top end of the scale, more than

85% of the scores assigned by the raters in this filtered condition were either exactly the same or had adjacent scores [e.g. a sound clip could be rated as a (5) or a (7), but not also as a (1) or (9)]. Overall the raters tended toward the center with (5) "Intermediate L2 English Prosody" being the most frequent score assigned during the rating task. This finding was interesting to note in light of the fact that there originally was an even distribution of items taken from each category when the sound clips were assigned ratings in the unfiltered holistic condition.

4. Discussion

This preliminary data analysis suggests several findings. The strong correlation between the trained raters' scores when listening to sound clips in the filtered condition and the original score assigned in the unfiltered condition from the holistic proficiency scale confirms the findings of other studies predicting the crucial role of prosody in holistic fluency scores. Also, the ability of the raters to clearly distinguish between Mandarin prosody and English prosody (even when the English prosody is produced by novice L2 learners whose first language is Mandarin) in this filtered condition adds weight to the claim that L2 prosody can be perceived and learned at various stages of proficiency. From the perspective of the listener, there is evidence for typologically distinguishable prosodic systems that can be isolated through the filtered speech condition. This result points to a potentially useful approach to assessing L2 prosodic proficiency levels as well as more general typological prosodic characterizations. This study also suggests that while the notion of prosody is often linked to holistic fluency, it is a unique proficiency category which can be isolated and tracked in various stages of L2 acquisition.

5. Conclusion

This study reveals a strong relationship between prosodic proficiency and holistic proficiency scores as assigned by trained raters. In future studies, we plan to compare the results of this type of evaluation with acoustic measurements that are more prosody related such as rhythm (i.e. rate of syllables per second and syllable duration) and rate of pitch change to investigate whether these variables were highly correlated with the raters' scores. This may indicate which perceptual prosodic cues were more salient and productive for this type of rating task. However, the results of this study offer evidence that rating filtered speech is a way of assessing prosodic proficiency as a whole system rather than relying solely on these more indirect fluency corollaries such as MLR, pitch duration or frequency of pausing.

6. Acknowledgements

I would like to thank Dr. Mary Niepokuj who generously contributed her time to this project. Also, a special thanks to InKyung Choi and Xun Yan for support in data analysis. The financial support and use of databases from the OEPP program through Dr. April Ginther is likewise gratefully acknowledged. Finally, I give thanks to God who has sustained and guided me throughout this project.

7. References

[1] Pickering, L., "The role of tone choice in improving ITA communication in the classroom, TESOL Quarterly 35(2): 233-255, 2001.

[2] Wennestrom, A., "Intonation as Cohesion in Academic Discourse", SSLA, 20, 1-25, 1998.

[3] Cutler, A., Dahan, D., &Donselaar, W.,"Prosody in the Comprehension of Spoken Language: A Literature review", Language and Speech, 40(2), 141-201, 1997.

[4] Cutler, A., Mehler, J., Norris, D.G., and Sugui, J., "The syllable's differing role in the segmentation of French and English", Journal of Memory and Language, 25, 385-400, 1986.

[5] Cutler, A. and Otake, T., "Mora or phoneme? Further evidence for language –specific listening", Journal of Memory and Language, 33, 824-844, 1994.

[6] Nguyen, T., Ingram, C.L. and Pensalfini, R.,"Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns", Journal of Phonetics, 36, 158-190, 2008.

[7] Beckman, M. "The parsing of prosody", Language and Cognitive Processes, 11, 17-67, 1996.

[8] Nazzi, T., Bertoncini, J., and Mehler, J., "Language discrimination by newborns: Toward an understanding of the role of rhythm", Journal of Experimental Psychology, 24(3), 756-766, 1998.

[9] Moon, C., Cooper, R., and Fifer, W., "Two day olds prefer their native language. Infant Behavior and Development, 16, 495-500, 1993.

[10] Shukla, M., White, K.S., and Aslin, R.N., "Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants", PNAS, 108(15): 60388-6043, 2011.

[11] Guion, S., "Knowledge of English Word Stress Patterns in Early and Late Korean-English Bilingual", SSLA, 27, 503-533, 2005.

[12] Derwing, T.M., Rossiter, M.J., Munro, M., Thomson, R. "Second Language Fluency: Judgments on Different Tasks", Language Learning, 54(4), 655-679, 2004.

[13] Munro, M.J., "Nonsegmental factors in foreign accent: ratings of filtered speech", Studies in Second Language Acquisition, 17 (March), 17-34, 1995.

[14] Iwashita, N., Brown, A., McNamara, T., and O'Hagan, S., "Assessed Levels of Second Language Speaking Proficiency", Applied Linguistics, 29(1), 24-49, 2008.