Exploring Lexical Stress Perception: An Investigation Across Varied Listener Experience with Prosody

Olivia J. Pride
University of New Hampshire, Durham

Follow this and additional works at: https://scholars.unh.edu/honors

Recommended Citation
https://scholars.unh.edu/honors/792
Exploring Lexical Stress Perception:

An Investigation Across Varied Listener Experience with Prosody

Olivia Pride

Department of Communication Sciences and Disorders

Thesis Advisor: Dr. Jill Thorson

Communication, Acquisition and Translational Studies (CAT) Laboratory
Table of Contents

Abstract .............................................................................................................................................. 3

Introduction ....................................................................................................................................... 4

Methods ........................................................................................................................................... 9

Results & Discussion ......................................................................................................................... 12

Conclusions ..................................................................................................................................... 15

References ....................................................................................................................................... 16
Abstract

Prosody is a characteristic of speech that allows others to understand emotion, emphasis, and meaning of language more deeply. Perceptual accuracy by speech-language pathologists is critical to diagnosing speech disorders. This study examined perceptual abilities of three different groups in identifying lexical stress across groups. Four SLPs, 6 SLP graduate students, and 6 untrained listeners listened to 396 audio files elicited in a prior study using the lexical stress tasks from the Prosodic Profiling Elements of Prosody in Speech-Communication (PEPS-C) and the naturalistic lexical stress (NLS) tasks. Listeners were asked if syllable stress was on the first syllable, second syllable, or ambiguous. Findings show that perceptual accuracy was ~60% for SLPs, who performed significantly better than the SLP grad students and the untrained listeners. This work highlights the importance of training perceptual discrimination in speech language pathology.

Keywords: prosody; lexical stress; perception; speech-language pathology
Introduction

Words themselves are not the only part of language that people use to clearly communicate. To emphasize ideas, show emotion, communicate meaning, connect with others, and, most importantly, be intelligible, prosody must be utilized. Prosody—the melody and rhythm of speech—consists of multiple levels, one of which is lexical stress. Lexical stress refers to the relative emphasis between syllables within a word. Producing appropriate prosody in speech is crucial to being understood by listeners. Identification and treatment of prosodic impairment is essential for individuals to clearly communicate. Thus, it is critical that a speech-language pathologist (SLP) can accurately identify prosodic differences. Many aspects need to be considered when listening to prosody, and being trained to listen for these specific traits is imperative for SLPs.

Syllables within a word vary in terms of their relative stress to one another, creating patterns of stronger and weaker syllables. For example, the word *monkey* contains initial stress or emphasis in the first syllable of the word. Although most people do not consciously consider the lexical stress of the words they are producing, listeners are sensitive to hearing when someone places stress on the wrong syllable. For instance, if *conTRIbute* was produced *CONtribute*, it may be perceived as incorrect, but the meaning would still be understood (Peppé, 2009). In other words, a difference in stress placement can change a verb to a noun and alter the meaning (*INcrease* = noun; *inCREASE* = verb). Initial stress is the most common stress pattern in spoken English with 90% of two-syllable (content) words being of this type (Brown et al., 2015). Listening for irregularities in lexical stress patterns is important for SLPs. When SLPs listen to a client sometimes the variability in their lexical stress may not be as obvious as some of the examples used, therefore their auditory perception must be trained to be sensitive to the slightest
differences. Currently, the field of communication sciences and disorders lacks a reliable measurement of prosody that assesses multiple prosodic domains, specifically lexical stress (Nayak et al., 2022).

Prosodic irregularities must be differentiated from typical prosody for accurate diagnosis and treatment. Certain prosodic features can be distinguishing in diagnosing a motor speech disorder. For example, one of the main characteristics of childhood apraxia of speech (CAS) is disordered prosody, including equal/excess lexical stress and syllable segmentation (McNeil et al., 1977; Macrae et al., 2015). Excess and equal stress is prominent or distinguishing or both (but not necessarily always present) in ataxic dysarthria and monopitch is prominent or distinguishing or both (but not necessarily always present) for hypokinetic dysarthria (Darley et al., 1969). Another study by Shriberg et al. (1997) compared children with both developmental apraxia of speech (DAS) and speech delay (SD). Results found that one of the only characteristics that discriminated between DAS and SD was inappropriate stress.

Although prosody is an important diagnostic criteria, many SLPs report avoiding prosodic testing due to limited sources of assessment of typical versus atypical prosody (Peppé, 2009). Hawthorne and Fischer (2020) found that 88% of the 245 SLPs who responded to their questionnaire thought that the assessment and treatment of prosody were within the scope of practice of an SLP, although many did not assess prosody even when they thought the patient may have a prosodic assessment. This leaves a gap in what SLPs should be able to identify and treat and what they are doing. A majority of participants responded with ‘occasionally’ or ‘rarely’ for their likelihood to target prosody in therapy with clients that may have prosodic impairments. Most of the SLPs who responded felt they should have had more time learning about prosody (Hawthorne & Fischer, 2020). Another 89% of the SLPs who completed the
survey did not feel equipped to assess or treat prosody. Most SLPs responded that they would assess prosody if there were tools available that were easy to use and more accurate. The conclusion was that there needs to be a greater concern for prosody, a focus on how to accurately diagnose and treat prosodic impairments, as well as proper training for SLPs. An effective prosodic assessment could help address these concerns, but SLPs must also be properly trained in prosody.

Pernon et al. (2022) conducted a study that examined how accurate SLPs’ perception was at classifying motor speech disorders. The two French-speaking groups of listeners were professional SLPs as well as student SLPs who were at the end of their first year of their master’s degree. The speech the SLPs and student SLPs listened to and made judgments on were elicited in a variety of settings including, spontaneous speech, from the speaker reading and diadochokinetic (DDK) tasks, or alternating motion rates (AMRs). After listening to the stimuli, they were asked to answer a series of questions about the speaker and a follow-up question about why they made the decisions they made. The first question determined if they received follow-up questions or not. Question 1 asked the listener if the speaker was neurotypical or pathological. Listeners were aware that there were neurotypical samples included in the study. The accuracy for question 1 across both groups was 72, with SLP accuracy slightly higher than student accuracy. Question 2 asked listeners which perceptual features they used to make their decision and Question 3 examined the listeners’ accuracy at classifying dysarthria versus apraxia of speech (AoS). Accuracy for Question 3 for both groups was 74%. Expert SLPs were again more accurate than student SLPs. Question 4 was a follow-up about the type of dysarthria and had an overall accuracy of 68%, with SLPs scoring higher than student SLPs. Accuracy overall was
72%, with continuous speech samples yielding the most accurate responses. The level of expertise of the listener as well as the elicitation method impacted perceptual accuracy.

Overall, there is a lack of prosodic assessments, with one of the most comprehensive being the Prosodic Profiling Elements of Prosody in Speech-Communication (PEPS-C) (Peppé & McCann, 2003). The PEPS-C was developed to fill the need for an assessment of prosody and intonation. It examines individuals’ receptive and expressive prosodic skills through 14 tasks (Peppé & McCann, 2003). The PEPS-C is a more controlled assessment, for example, for the expressive function task a single word is presented, and the individual is to say it in isolation based on what they see (DIScard vs disCARD). The PEPS-C is a comprehensive assessment of prosodic form and function (Peppé, 2015). Potential issues have been reported though with the lexical stress tasks, and if they are truly assessing an individual’s ability (Courter & Thorson, 2022; Patel et al., 2023).

A novel lexical stress assessment is the naturalistic lexical stress (NLS) expressive and receptive tasks (Dempsey & Thorson, 2023). The NLS expressive task has a naturalistic elicitation method in comparison to the PEPS-C, which is more controlled. The NLS tasks consist of 1) selecting the lexical stress of a disyllabic word after hearing an audio stimulus (receptive) and 2) describing a picture presented (expressive). For the receptive task, the words selected were common to everyday life. The audio stimuli were played, and the word was presented on the screen with the varied lexical stress options. For the expressive task of the NLS, trochaic (strong-weak) and iambic (weak-strong) words that were easily represented through pictures were used and were selected using the Macarthur Communicative Development Inventories (MCDI) (Fenson et al., 2007). A picture description task was used to elicit more natural speech and capture participants’ expressive lexical stress in running speech.
Dempsey and Thorson (2023) developed the NLS and compared the outcomes of the tasks with that of the PEPS-C lexical stress tasks. Results showed that adults were more accurate in both receptive and expressive tasks of NLS in comparison to the PEPS-C task administered to the same individuals when assessed perceptually (Dempsey & Thorson, 2023). The output collected during the expressive tasks in Dempsey and Thorson (2023) were used as the stimuli in the current study (with participant recordings from both the NLS and PEPS-C lexical stress tasks).

The goal of the current study is to assess the perception of lexical stress by comparing three groups of listeners with varying degrees of exposure to prosody and communication sciences and disorders (CSD). The three groups of interest are untrained listeners, SLP graduate students, and SLPs with at least five years of experience. This study will investigate group performance in identifying lexical stress, with the hypothesis that ability will vary based on the level of experience in CSD and speech-language pathology. One hypothesis is that the practicing clinicians (vs. SLP graduate students and untrained listeners) will be most accurate during the task due to their years of training and experience in listening to speech and lexical stress (Pernon et al., 2022). An alternative hypothesis is that practicing SLPs will not be as accurate as they should be due to a lack of training in prosody (Hawthorne & Fischer, 2020). We predict that accuracy may be higher for the PEPS-C task due to its elicitation technique that results in hyper-articulated prosodic variations (Courter & Thorson, 2022). The use of PEPS-C and NLS stimuli allows listeners to hear both more natural speech (NLS) and over-emphasized stimuli (PEPS-C). This study will explore lexical stress perception across varied listeners. In this study we are also interested in the type of lexical stress that has higher accuracy. We hypothesize that trochaic
responses would have the highest accuracy in comparison to iambic due to being most occurrent in English (Brown et al., 2015).

**Methods**

**Participants**

Participants were recruited by word of mouth. Data were collected from 17 participants, from three groups: 1) individuals with no background in speech, language, or hearing (6), 2) graduate students in SLP master’s program (6), and 3) practicing SLPs who have at least five years of experience (5). Each participant was required to meet the inclusion criteria of being a native speaker of English spoken in the United States, having normal hearing, having no developmental, neurological, or genetic disorders, and being at least 18 years old. All participants were required to pass a pure tone hearing screening in a noise-proof booth (passing at 25 dB for 250, 500, 2000, 4000, 6000, and 8000 Hz in at least one ear). Data from one SLP were excluded due to not passing the hearing screening.

**Procedure**

This study obtained Institutional Review Board (IRB) approval under IRB-FY2023-5 at the University of New Hampshire. Informed consent was obtained first from participants, followed by completion of a demographic form, both of which were collected on tablets. A hearing screening was then administered in a soundproof audiology booth. Once the hearing screening was passed, the experimental portion of the study began in a quiet room. The experiment began with four practice trials where participants were able to listen and respond, ensuring their understanding of the task. Participants were asked to classify lexical stress for a word. They were provided three options after hearing the stimulus item and asked to click one on the computer screen to indicate their response: 1 (first syllable stress), 2 (second syllable stress),
or A (ambiguous). Following the practice stimuli, the study began and contained 396 audio files of a subset of the participant recordings from the PEPS-C (Peppé & McCann, 2003) and the naturalistic lexical stress (NLS) tasks collected from Dempsey and Thorson (2023). The ExperimentMFC function in Praat that was utilized to run the experiment (Boersma & Weenink, 2018).

**Stimuli**

In Dempsey and Thorson (2023), 16 items from PEPS-C and 10 items from NLS were used to elicit participant responses for expressive lexical stress. The current study utilizes those responses to create the stimuli used here. In Dempsey and Thorson (2023), forty native English speakers participated in the study. The PEPS-C tasks were presented with an image on a screen where the stressed syllable was presented with larger font (INdent or inDENT) with a larger circle over the intended emphasized syllable and a smaller circle over the syllable that contained secondary stress. Using the written word on the screen accompanied by the circles, participants were asked to say the word with appropriate lexical stress. In the current study, half the stimuli that the participants listened to are the PEPS-C audio files that were collected during the Dempsey and Thorson (2023), study and decide which syllable contains the emphasis with no visual aid.

There were also 10 NLS stimuli elicited from the participants in Dempsey and Thorson (2023) that were also listened to by participants in this study. During the elicitation of these audio recordings, participants uttered the target words in continuous speech when describing the picture used in the NLS expressive task. These audio recordings are more naturalistic since participants were asked to describe what they saw in an image. There was an even number of trochaic and iambic words chosen (selected MCDI) in the picture used for the NLS expressive
In the current study, the other half the stimuli that the participants listened to were from the NLS audio files that were collected during the Dempsey and Thorson (2023). See Table 1 for a list of stimuli and their attributes.

Table 1: Stimuli list including item, lexical stress position, phonetic transcription, assessment origin, and syntactic class.

<table>
<thead>
<tr>
<th>Item</th>
<th>Lexical stress</th>
<th>Phonetic spelling</th>
<th>Assessment</th>
<th>Syntactic Class &amp; Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>differ</td>
<td>1</td>
<td>/ˈdrfər/</td>
<td>PEPS-C</td>
<td>v: to be different</td>
</tr>
<tr>
<td>defer</td>
<td>2</td>
<td>/dəˈfɜːr/</td>
<td>PEPS-C</td>
<td>v: to put off or delay</td>
</tr>
<tr>
<td>increase</td>
<td>1</td>
<td>/ˈɪnkrɪs/</td>
<td>PEPS-C</td>
<td>n: growth</td>
</tr>
<tr>
<td>increase</td>
<td>2</td>
<td>/ɪnˈkrɪs/</td>
<td>PEPS-C</td>
<td>v: to make bigger, more numerous</td>
</tr>
<tr>
<td>insight</td>
<td>1</td>
<td>/ˈɪnsaʊt/</td>
<td>PEPS-C</td>
<td>v: ability to perceive clearly or deeply</td>
</tr>
<tr>
<td>incite</td>
<td>2</td>
<td>/ɪnˈsaʊt/</td>
<td>PEPS-C</td>
<td>v: to provoke or urge</td>
</tr>
<tr>
<td>insert</td>
<td>1</td>
<td>/ˈɪnsət/</td>
<td>PEPS-C</td>
<td>n: item put in</td>
</tr>
<tr>
<td>insert</td>
<td>2</td>
<td>/ɪnˈsət/</td>
<td>PEPS-C</td>
<td>v: to put in</td>
</tr>
<tr>
<td>indent</td>
<td>1</td>
<td>/ɪnˈdent/</td>
<td>PEPS-C</td>
<td>n: space at start of paragraph</td>
</tr>
<tr>
<td>indent</td>
<td>2</td>
<td>/ɪnˈdent/</td>
<td>PEPS-C</td>
<td>v: to start a line further in than main text</td>
</tr>
<tr>
<td>discard</td>
<td>1</td>
<td>/ˈdɪskɑrd/</td>
<td>PEPS-C</td>
<td>n: item rejected</td>
</tr>
<tr>
<td>discard</td>
<td>2</td>
<td>/dɪsˈkɑrd/</td>
<td>PEPS-C</td>
<td>v: to reject</td>
</tr>
<tr>
<td>import</td>
<td>1</td>
<td>/ˈɪmɑrpt/</td>
<td>PEPS-C</td>
<td>n: item brought in from foreign country</td>
</tr>
<tr>
<td>import</td>
<td>2</td>
<td>/ɪmˈpɔrt/</td>
<td>PEPS-C</td>
<td>v: to bring in from a foreign country</td>
</tr>
<tr>
<td>discount</td>
<td>1</td>
<td>/ˈdɪskəʊnt/</td>
<td>PEPS-C</td>
<td>n: cut in price</td>
</tr>
<tr>
<td>discount</td>
<td>2</td>
<td>/dɪsˈkɑʊnt/</td>
<td>PEPS-C</td>
<td>v: to drop the price</td>
</tr>
<tr>
<td>insult</td>
<td>1</td>
<td>/ˈɪnsɔlt/</td>
<td>PEPS-C</td>
<td>n: rude remark</td>
</tr>
<tr>
<td>insult</td>
<td>2</td>
<td>/ɪnˈsɔlt/</td>
<td>PEPS-C</td>
<td>v: to make a rude remark to someone</td>
</tr>
<tr>
<td>balloon</td>
<td>2</td>
<td>/bəˈlun/</td>
<td>NLS</td>
<td>noun</td>
</tr>
<tr>
<td>bubbles</td>
<td>1</td>
<td>/ˈbʌbəlz/</td>
<td>NLS</td>
<td>noun</td>
</tr>
<tr>
<td>canoes</td>
<td>2</td>
<td>/kəˈnu/</td>
<td>NLS</td>
<td>noun</td>
</tr>
<tr>
<td>chicken</td>
<td>1</td>
<td>/ˈʧɪkən/</td>
<td>NLS</td>
<td>noun</td>
</tr>
<tr>
<td>cookie</td>
<td>1</td>
<td>/ˈkʊki/</td>
<td>NLS</td>
<td>noun</td>
</tr>
<tr>
<td>giraffe</td>
<td>2</td>
<td>/ʤɪˈrɑf/</td>
<td>NLS</td>
<td>noun</td>
</tr>
</tbody>
</table>


<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>guitar</td>
<td>2</td>
<td>/ˈɡɪˈtar/</td>
<td>NLS</td>
</tr>
<tr>
<td>monkey</td>
<td>1</td>
<td>/ˈmʌŋki/</td>
<td>NLS</td>
</tr>
<tr>
<td>raccoon</td>
<td>2</td>
<td>/rəˈkun/</td>
<td>NLS</td>
</tr>
<tr>
<td>turtle</td>
<td>1</td>
<td>/ˈtɜːtəl/</td>
<td>NLS</td>
</tr>
</tbody>
</table>

**Results & Discussion**

Using R statistical software, a generalized linear mixed model was conducted with task, group, and lexical stress as fixed effects, and stimulus and ID as random effects to predict the outcome variable of accuracy (v.4.3.0; R Core Team, 2023). Assumptions were checked prior to running the model.

The first hypothesis investigated the lexical stress perception across the participant groups. Consistent with the initial hypotheses, SLPs exhibited the highest overall accuracy in identifying lexical stress patterns (59.85%), followed by SLP graduate students (45.96%), and then untrained listeners (38.30%). The SLPs who participated had been practicing SLPs for a minimum of 15 years and an average of 24.5 years. SLPs had statistically significant higher accuracy than the untrained listeners ($β = .96, SE = .42, p = 0.021$). Graduate SLPs were not statistically significantly more accurate than the untrained listeners ($β = .32, SE = .37, p = 0.385$).

As predicted, the model also showed that accuracy was higher for the PEPS-C productions than the NLS ones ($β = .50, SE = .17, p = 0.003$) most likely due to their over emphasized productions. Contrary to expectations, SLPs demonstrated only marginally higher accuracy (1.26%) on NLS tasks compared to PEPS-C tasks. SLPs did better on these naturally elicited stimuli, due to their experience, yet they still did not perform as high as would be expected. The hypothesis regarding better performance on PEPS-C tasks was confirmed for both untrained listeners and SLP graduate students. The overall accuracy across all participants on both types of tasks was 50.88% for the PEPS-C and 45.19% for the NLS. Figure 1 shows
percentage of lexical stress identification accuracy for each group on the PEPS-C tasks, NLS tasks, and overall.

Figure 1: Clustered bar graph representing the accuracy across listeners for the PEPS-C task, NLS task and overall accuracy of the group.

Looking more closely at performance by lexical stress type (iambi or trochaic), the effect of lexical stress type was significant ($\beta = .39, SE = .18, p = 0.032$), showing that trochaic productions were more accurately identified, which follows from frequency patterns of English. The hypotheses that trochaic responses would have the highest accuracy was supported. Figure 2 shows how the trochaic stimuli had higher accuracy than the iambic ones across the groups. It also illustrates the large the number of stimuli that untrained listeners responded to as ambiguous, which was higher than the accurate number of trochaic or iambic stimuli.
Figure 2: Clustered bar graph that represents the count of correct trochaic/iambic/ambiguous responses across the groups (Note: 6 untrained, 6 graduate, 4 SLPs participated).

Limitations

The main limitation of this study was sample size. A limited sample size impacts the generalizability of the findings. With only 16 participants, there may not have been enough statistical power to be find significant effects, so continued data collection would benefit the analyses and potential implications of the work. Though there were fewer participants, there were still significant findings, showing that this work may be revealing robust findings. This study was also taken from each participant at only one point in time and it was not able to track participant accuracy over time. Tracking the progress of a SLP/SLP graduate student over time on their lexical stress ability could be revealing if coupled with if prosodic training. Despite the limitations, this study supports the need for increased prosodic training for SLPs.
Conclusions

In conclusion, this study aimed to explore lexical stress perception skills across varied listeners, examining their accuracy on stimuli from PEPS-C tasks versus NLS tasks and iambic versus trochaic productions. There is limited research in lexical stress perception of SLPs, thus this study fills a gap in the literature in how both experienced SLPs and graduate students are able to perceptually rate lexical stress differences in neurotypical adult speakers. Findings revealed that while SLPs are more accurate in lexical stress identification, their overall accuracy was below two-thirds (59.85%), highlighting a need for additional clinical training.

This study highlights the need to develop prosodic training programs, with a focus on improving lexical stress perception abilities. This can be accomplished by increasing exposure to prosody in graduate school, and can also be extended practicing SLPs who may require practice for their perceptual prosodic skills. Having more highly trained SLPs in prosody has the potential to positively impact accurate diagnoses and effective treatment of communication skills.
References


