

CROSS-DIALECTAL SPEECH PROCESSING: PERCEPTION OF LEXICAL STRESS BY INDIAN ENGLISH LISTENERS

Olga Maxwell^a and Robert Fuchs^b

^aUniversity of Melbourne, Australia, ^bUniversity of Hamburg, Germany

^aomaxwell@unimelb.edu.au, ^brobert.fuchs@uni-hamburg.de

ABSTRACT

In English, lexical stress provides essential information guiding lexical activation. However, little is known about the processing of lexical stress in post-colonial Englishes. The present study examines the perception of lexical stress in disyllabic words by adult speakers of Standard Indian English. Results show that in iambic words (second syllable stressed), participants perform at about 54% accuracy, regardless of social background. In trochaic words, participants with private schooling perform significantly better (60% accuracy; $p < 0.05$) than those with a government school background, approaching the level of accuracy reported for Australian English listeners. Our results suggest that processing of the commonly occurring trochaic condition is easier for participants from private schools, while processing of the rare iambic pattern is not eased by such experience. L1 background and onset of learning English show no systematic effect on participants' performance. Variability in Standard Indian English is shaped mainly by schooling and not L1 background.

Keywords: perception, lexical stress, Indian English, speech processing.

1. INTRODUCTION

1.1. Background

Research suggests that adult linguistic cognition does not have the same processing efficiency in a second language (L2) as in a first language (L1). However, speech perception can be highly robust to variation. Once a listener is familiar with a particular source of variation, they can enjoy significant processing benefits [4], e.g. better sensitivity to unfamiliar contrasts [7]. Cross-dialectal speech perception research shows that participants who are exposed to different varieties ("canonical" vs. new) have both (1) processing benefits, as mentioned above, but also (2) processing costs [4]. Familiarity with multiple linguistic systems can result in (1) independent processing benefits as well as (2) competition among variable multi-dialect representations. Most of these conclusions, however, are based on segmental and lexical perception, not prosody. One of the aims of the present paper is to extend this line of research to the realm of prosody.

Prosodic structures differ across languages and language varieties [19] and influence the way listeners use prosodic cues. The subject of the present study is lexical stress, which refers to the lexically specified distinction between strong and weak syllables in a word [2, 19]. In stress languages, such as English, lexical stress is cued by a number of robust acoustic parameters that make stressed syllables more salient to listeners [20]. Similar to other dimensions of prosody, stress is not universal. Not all languages contrast stressed with unstressed syllables, and even within the category 'stressed', languages can have various patterns. Given such differences in the function and physical properties of stress, listeners may use acoustic information in the speech signal differently.

While most varieties of English as Native Language (ENL) have a distinction between stressed and unstressed syllables, it has been suggested that this may not be the case in many postcolonial varieties of English, such as Nigerian English, Singaporean English or Indian English. This could be due to the influence of typologically distinct first languages, which differ in their prosodic systems and the use of prominence [e.g. tonal varieties of English proposed in 15]. Empirical evidence on such varieties, however, is often limited or presents conflicting results [cf. 35], at times suggesting a lack of distinction for stressed/unstressed syllables.

It is well documented that stress patterns guide listeners of British and Australian English (BrE/AusE) in speech segmentation and provide important cues for lexical activation [5, 6, 10, 26]. Primary stress on the initial syllable (trochaic stress) is widespread in English, and listeners rely on this information in perception. English listeners also show a bias for selecting word-initial strong syllables [5, 8], and identify target phonemes in these words faster [26]. However, little is known about how listeners of postcolonial varieties perceive stress, and how their processing strategies compare with those in ENL varieties. The present study will extend our knowledge of cross-dialectal processing of prosody by investigating the perception of lexical stress by listeners of Indian English (IndE).

1.2. The present study

English is used in India mostly as a second language. In this paper, we focus on 'Standard Indian English',

a more prestigious variety used by proficient educated speakers. IndE is an interesting and challenging case to study. The linguistic landscape in the subcontinent is characterised by vast linguistic diversity and multi-lingualism on the one hand, and the presence of areal features shared across South Asian languages on the other [18, 22]. This complex linguistic context frames the ongoing debate in the literature on the nature and ‘uniformity’ of this variety [12, 23, 24, 27, 30].

While it has been established that IndE has shared features, more recent experimental research finds evidence for specific L1 influence [23, 24, 33, 34]. However, the findings are not always consistent and also indicate that the extent of L1 influence may vary depending on the feature under investigation [23, 24, 31, 33, 34]. In speech production, variation across speakers could be based on various factors or a combination of multiple factors. Further, recent perception studies found no or marginal L1 effects on the perception of rhythm [12] and information structure and focus [25] by IndE listeners, but instead reported that a listener’s schooling experience (private vs. government) influenced results. In summary, previous experimental research has mostly focussed on speech production, and has reported mixed findings in relation to L1 influence. Most importantly, previous phonetic studies have often neglected to take into account a range of sociolinguistic variables.

Consequently, the study addresses the following questions:

1. How do IndE listeners process lexical stress? How accurately can they identify words based on the acoustic information contained in the first syllable?
2. Do the factors L1, age of learning and type of school influence the perception of stress?
3. How do IndE listeners perform compared to the listeners of a “canonical” ENL variety? What kinds of processing costs/benefits can be observed?

1.2.1. L1 Influence

L2 listeners may use specific acoustic cues found in their L1 in lexical stress processing [5, 7, 10]. While the patterns of stress placement in the languages spoken in India vary (e.g. first syllable in Bengali [16, 17], syllable weight and phonological vowel in Telugu [29]), South Asian languages are known to have phonetically weak stress [16, 17, 18] - unlike canonical varieties of English (such as BrE), which have phonetically strong stress with such robust acoustic cues as duration, vowel quality (full vs reduced), intensity and f0.

Further, speakers of South Asian languages often report difficulty in the auditory identification of stress, with disagreement over the presence and type of acoustic cues, and over the location of the stressed

syllable [e.g. 21 for Telugu]. This may explain the lack of any evidence for L1 influence on the acoustic correlates of stress in IndE [13] and justifies the need to further examine the effect of L1 or lack thereof on stress perception in this variety.

1.2.2. School Type

The variable of school type is of particular interest given the educational system in India. All participants in this study attended schools that officially use English as medium of education (as opposed to a local language such as Hindi or Telugu). In private schools, teachers are likely to use English all or the majority of the time, providing an immersion environment to students. Teachers in these schools are also likely to be proficient in English and have an accent that indexes overt prestige locally, i.e. clearly Indian but with limited or no discernible L1 influence that would make them readily identifiable as speakers of a particular Indian language. By contrast, government schools are more likely to have teachers who speak with a local accent and might on occasion have limited proficiency in English. While textbooks and other teaching materials are in English, teachers might use a local Indian language to communicate verbally with students in order to ease their understanding of the subject matter.

Because of the differential impact of private vs. government schooling on participants’ proficiency, we expect a diminished effect of L1 on the perception of stress by IndE listeners but predict potential benefits (greater accuracy in responses) for those participants who went to a private school.

1.2.3. Age of Learning (AoL)

Experience with an L1 is known to influence the perception and production of a speaker’s L2 [e.g. 3, 11]. However, we are not dealing with speakers of a typical L2 variety, who use their L2 in a country where it is the main language. Given that English is taught in India at an earlier age in private schools and speakers receive different input, we predict that AoL may not have a strong effect and will be subject to the participants’ schooling background. More generally, taking into consideration the lack of robust phonetic cues to stress in Indian languages and the exposure to IndE as a variety since early childhood [34], we expect that the participants in this study will show processing costs as compared to AusE listeners (as in [5]).

2. METHOD

2.1. Participants

28 students (aged 22-34; 13 f, 15 m) at the University of Hyderabad, India, took part in the experiment. All were enrolled in a university degree at the time of data

Figure 1: The participants' places of birth in India.



collection, identified as bi- or multilingual, and reported no hearing problems. They had all moved to the city for the purpose of tertiary education and none were born in Hyderabad. Fig. 1 shows the participants' places of birth.

Participants represented four L1 backgrounds (Tamil, Telugu, Hindi and Bengali; 7 for each L1), had never lived in another English-speaking country. They had acquired English at different ages, which we operationalised here as 'early', 'mid' and 'late' onset of learning English (age 3-4, 5-7 and 8-10). Participants had attended government schools (11 participants), private schools (including convent schools; 10), or had a mixed schooling background (7) where they went to a government school for primary education, and a private school for secondary, or vice versa. All participants were speakers of Standard IndE.

2.2. Materials and Procedure

Participants were presented with 21 truncated word pairs with segmentally identical first syllable and different lexical stress location (materials developed by and first used in [5]). One member of the pair had primary stress on the first, while the other had stress on the second syllable (e.g. syllable *car-* in *CAR*ton vs *car*TOON). The original set of words was recorded in a carrier sentence [See 5]. The truncated audio files included the first syllable for each word repeated twice, creating a set of 84 fragments in total (pseudo-randomised).

Participants were tested individually in a quiet room on the premises of the University of Hyderabad. The stimuli were presented on an HP laptop (Intel Premium Processor) with the help of PsychoPy (Version 3.0.0.b7) [28] over Audio-Technica ATH-ANC70 noise-cancelling headphones.

On each trial, participants were presented with a word pair on the screen and were given two seconds to view the words before they heard the audio stimulus. They were given the task to judge which of the words in a pair was the source of the current fragment.

For one of the presentations of the fragment, the correct response was the word on the left and for the other presentation the right word. Each trial was separated by a 4.5 second interval. Prior to the main experiment, participants completed an additional short training set consisting of two word pairs.

2.3. Analysis

We determined which factors influence the likelihood that a given participant chooses the correct syllable by fitting logistic mixed effects regression models in R with package `lme4` [1]. We fitted two models, one for trochaic and one for iambic words. All models used CORRECT as dependent variable and included PARTICIPANT and WORD as random factors and a fixed factor accounting for word frequency (FREQ), in order to correct for potentially greater familiarity with frequent words. FREQ was operationalised as the logarithm of the frequency per million words (pmw) in the Indian section of the Global Web-Based Corpus of English [9]. The pmw frequency of a single word that never occurred in the corpus was set to 0.001 in order to avoid an infinite value after logarithmic transformation.

In order to select a model from the remaining factors (SCHOOLING, AGE-OF-LEARNING, L1), we used random forests (package `party` [32]) to determine variable importance. We added influential variables to the models in order of variable importance, as long as the added factor was significant in the regression model. For the model for trochaic stress, this was only true for SCHOOLING. For the model for iambic stress, adding the variable identified as most important by the random forest resulted in a model with a non-significant factor. The final models are:

```
m_right <- glmer(correct ~ freq_log +
  (1|participant) + (1|word),
  data=Stress_right, family=binomial)
m_left <- glmer(correct ~ school +
  freq_log + (1|participant) + (1|word),
  data=Stress_left, family=binomial)
```

3. RESULTS

Overall, the IndE listeners matched the fragments to the correct words at a rate of 55%, i.e. above chance. Their performance was marginally worse than that of AusE listeners, who identified 59.2% of the target fragments correctly [5]. Previous work has shown that a forced choice identification task involving lexical stress is generally difficult for English listeners [5, 8].

Remarkably, the IndE listeners showed a similar performance for fragments from trochaic and iambic words, as shown in Fig. 2 (55.6% correct for left stress fragments and 54.5% for right stress). AusE listeners in [5], by contrast, exhibited a clear bias towards the left-stress condition (62.5% accuracy).

Figure 2: Means and standard deviations for correct responses for words for IndE and AusE listeners ([5]), presented as a factor of stress location (iambic and trochaic).

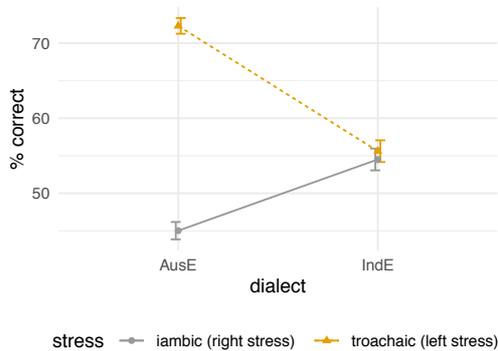
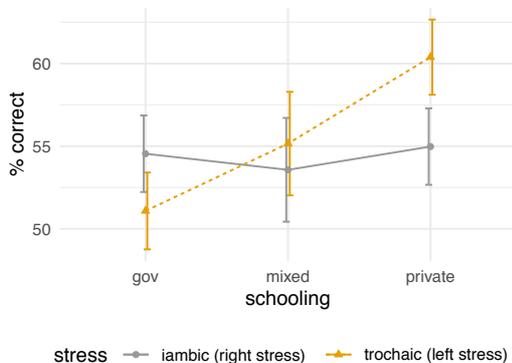


Figure 3: Means and standard deviations for correct responses by IndE listeners, across three types of schooling background (government, private, mixed), presented as a factor of stress location (iambic and trochaic).



Both L1 and onset of learning English did not influence participants’ performance systematically. Despite an observable difference for L1 Tamil in the left-stress condition (possibly compounded by school type), L1 had no significant effect on the accuracy of matching fragments to words.

As predicted, schooling background had a significant effect on the percentage of correct responses but only for one of the conditions (see Fig. 3). In iambic words, participants performed at about 54% accuracy regardless of school type. By contrast, in the trochaic condition, participants with government schooling performed just marginally better than chance (51%), while those who went to a private school performed significantly better (60% accuracy; $z=-2.572$, $p<0.05$).

4. DISCUSSION

Our analysis shows that listeners of IndE are able to distinguish words based on the acoustic cues provided in word fragments, and the overall proportion of correct responses is close to that reported for AusE listeners. However, we find strong evidence for processing costs when looking at trochaic vs iambic

stress words. As mentioned earlier, listeners of ENL varieties are much better at identifying left-stress syllables and perform at chance when listening to words with iambic stress. We find that for listeners of IndE both stress conditions are equally difficult. Processing costs could be explained by the fact that languages spoken in India do not have the same robust cues to lexical stress and this is a convergent feature in South Asian languages, ultimately adopted in IndE. This may also be a likely explanation for the lack of any effect of L1 on participants’ performance. However, further investigation with more participants for each L1 background is needed to get a complete picture.

As predicted, age of learning did not influence the degree of correct performance, and the only variable that was found to be significant was the schooling background. Processing benefits for the participants with a private school education were evident in the trochaic stress condition, where they performed as well as AusE listeners. Most likely, exposure to a more prestigious (and more “canonical”) variety and greater frequency of input provided an advantage in the processing of lexical stress for this group.

The present study raises an important question about the processing of prosody across varieties of English and the effects of multi-dialectal representations on speech perception across dialects. New and emerging dialects do not seem to fully conform to the L1-L2 influence model and may present a context with a complex interplay of sociolinguistic and developmental factors. More broadly, our findings, echoing those in [15], raise the question of how new varieties can be captured within the prosodic typology, in this instance, word-level prosody.

The next steps will be to (1) closely examine the weighting of acoustic cues (i.e. duration, intensity, pitch, and vowel quality) for IndE listeners; (2) further explore the effects of exposure to multiple dialects on speech processing by working with IndE listeners in the diaspora (Australia); and (3) investigate the perception of lexical stress by listeners from other sociolinguistic backgrounds (as lower socio-economic status may potentially show stronger L1 and AoL effects).

5. ACKNOWLEDGEMENTS

This project was supported and funded by the Head of School Investment Fund, School of Languages and Linguistics, University of Melbourne. We are extremely grateful to Professor Anne Cutler and Dr Laurence Bruggeman for sharing the original materials [5, 8]. We also thank the participants, and Akram SK and C. Bhargavi for assistance with data collection.

6. REFERENCES

- [1] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1), 1-48.
- [2] Beckman, M. 1986. *Stress and non-stress accent*. Dordrecht: Foris.
- [3] Best, C. 1994. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In: Goodman, J.C., Nusbaum, H.C. (eds.), *The development to speech perception: The transition from speech sounds to spoken words*. Cambridge, MA: MIT Press, 167-224.
- [4] Clopper, C. 2014. Sound change in the individual: Effects of exposure on cross-dialectal speech processing. *Laboratory Phonology* 5, 69-90.
- [5] Cooper, N., Cutler, A., Wales, R. 2002. Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language & Speech* 45(3), 207-228.
- [6] Cutler, A. 2009. Greater sensitivity to prosodic goodness in non-native than native listeners. *Journal of the Acous. Society of America* 125, 3522-3525.
- [7] Cutler, A. 2012. *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: MIT Press.
- [8] Cutler, A., Bruggeman, R., Wagner, A. 2016. Use of language-specific speech cues in highly proficient second-language listening. *Journal of the Acous. Society of America* 139(4), 2161.
- [9] Davies, M., Fuchs, R. 2015. Expanding horizons in the study of World Englishes with the 1.9 billion Word Global Web-Based English Corpus (GloWbE). *English World-Wide* 36(1), 1-28.
- [10] Fear, B., Cutler, A., Butterfield, S. 1995. The strong/weak syllable distinction in English. *Journal of the Acous. Society of America* 97, 1893-1904.
- [11] Flege, J.E. 1995. Second language speech learning: Theory, findings, and problems. In: W. Strange (ed.), *Speech perception and linguistic experience: Issues in crosslanguage research* Timonium, MD: York Press, 233-277.
- [12] Fuchs, R. 2016. *Speech rhythm in varieties of English: Evidence from Educated Indian English and British English*. Singapore: Springer.
- [13] Fuchs, R., Maxwell, Olga. 2015. The placement and acoustic realisation of primary and secondary stress in Indian English. *Proceedings of ICPHS 2015*, Glasgow.
- [14] Fuchs, R., Götz, S., Werner, V. 2016. The present perfect in learner Englishes: A corpus-based case study on L1 German intermediate and advanced speech and writing. In: Werner, V., Seoane, E., Suárez-Gómez, C. (eds.) *Re-Assessing the Present Perfect*. Berlin: de Gruyter, 297-337.
- [15] Gussenhoven, C. 2014. On the intonation of tonal varieties of English. In: Filppula, M., Klemola, J., Sharma, D. (eds). *The Oxford Handbook of World Englishes online*. Online Publication.
- [16] Hayes, B., Lahiri, A. 1991. Bengali intonational phonology. *Natural Language & Linguistic Theory* 9(1), 47-96.
- [17] Khan, S. D. 2014. The intonational phonology of Bangladeshi Standard Bengali. In S.-A. Jun (Ed.) *Prosodic typology 2: The phonology of intonation and phrasing*. Oxford, UK: Oxford University Press, pp. 81-117.
- [18] Khan, S. D. 2016. The intonation of South Asian languages. *FASAL-6*, Umass Amherst, 2016.
- [19] Ladd, D. R. 2008. *Intonational phonology*. Cambridge: CUP.
- [20] Liberman, P. 1963. Some effects of semantic and grammatical context on the perception of speech. *Language and Speech* 6, 172-187.
- [21] Lisker, L., Krishnamurti, Bh. 1991. Lexical stress in a 'stressless' language: judgments by Telugu- and English-speaking linguists. *Proceedings of ICPHS1991*, Université de Provence, 90-93.
- [22] Masica, C. 2005. *Defining a linguistic area: South Asia*. New Delhi: Chronicle Books.
- [23] Maxwell, O. 2014. *The intonational phonology of Indian English: An Autosegmental-Metrical analysis based on Bengali and Kannada English*. PhD thesis, University of Melbourne, Melbourne, Australia.
- [24] Maxwell, O., & Payne, E. (2018). Pitch accent types and tonal alignment of the accentual rise in Indian English(es). In *Proceedings of Speech Prosody 2018*, Poznan, Poland.
- [25] Maxwell, O., Loakes, D., Clothier, J., Holt, C., Fletcher, J. 2018. Variability in the perception of prosody (focus structure) by Indian English listeners. *Sociolinguistic Symposium 22*, University of Auckland, New Zealand.
- [26] Mattys, S., Samuel, A. 2000. Implications of stress pattern differences in spoken word recognition. *Journ. of Memory and Language* 42, 571-596.
- [27] Mukherjee, J. 2007. Steady states in the evolution of new Englishes: Present-day Indian English as an equilibrium. *Journal of English Linguistics* 35(2), 157-187.
- [28] Peirce, J.W. 2009. Generating stimuli for neuroscience using PsychoPy. *Front. Neuroinform.* 2, 10.
- [29] Pulipaty, K. 2018. *Metrical structure and stress in Telugu*. Unpublished MA thesis, University of Melbourne, Australia.
- [30] Sailaja, P. 2012. Indian English: Features and sociolinguistic aspects. *Language and Linguistics Compass* 6(6), 359-370.
- [31] Sirsa, H., Redford, M. 2013. The effects of native language on Indian English sounds and timing patterns. *Journal of Phonetics* 41(6), 393-406.
- [32] Strobl, C., Boulesteix, AL., Kneib, T., Augustin, T., Zeileis, A. 2008. Conditional variable importance for random forests. *BMC Bioinformatics*, 9, 307.
- [33] Wiltshire, C. 2005. The "Indian English" of Tibeto-Burman language speakers. *English World-Wide* 26(3), 275-300.
- [34] Wiltshire, C., Harnsberger, J. 2006. The influence of Gujarati and Tamil L1s on Indian English: A preliminary study. *World Englishes* 25(1), 91-104.
- [35] Wiltshire, C., Moon, R. 2003. Phonetics stress in Indian English versus American English. *World Englishes* 22(3), 291-303.