Stress Marking on Urdu Speech Corpus using Acoustic Cues

Presented by: Benazir Mumtaz

Centre for Language Engineering
Al-Khawarizmi Institute of Computer Science
University of Engineering and Technology Lahore, Pakistan

Contents

- Motivation
- Acoustic Impact of Stress
- The Process of Annotating Urdu Speech Corpus at Stress Tier
- The Process of Assessing Stress Tier Annotation
- Results and Discussion

Motivation

- To explore the unpredictability of prominence in speech
- To explore how stress can change the phonetic properties of a segment
- To prioritize the order of acoustic cues for stress marking in Urdu language
- To develop an Urdu text-to-speech system

Acoustic Impact of Stress

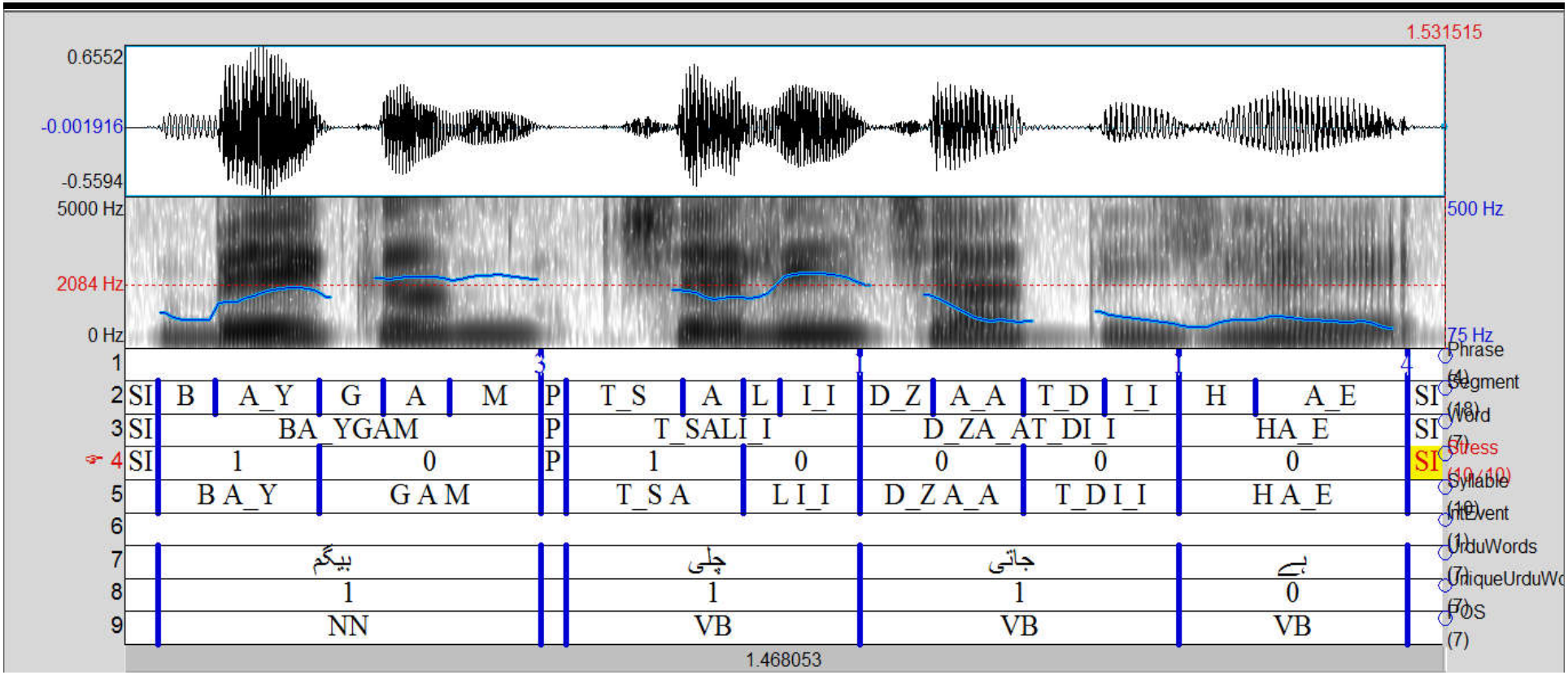
- **Duration**
  - Intrinsic duration of the segment [1]
  - Phonological length [2]
  - Phrase final syllable [3]
- **Fundamental frequency/f0**
  - Intrinsic f0 of the segment
  - Contextual variation [4]
- **Intensity**
  - Intrinsic intensity of the segment
  - Emotional state of the speaker [4]

Description of Urdu Speech Corpus

- Speech Corpus Size: 30 minutes
- Recording Sampling Rate: 48 kHz
- Software: PRAAT

Process of Annotating Urdu Speech Corpus at Stress Tier

- While listening to the file for the stress marking, take sub phrases ending in pauses or glottalization
- Assign ‘1’ to a stressed syllable and ‘0’ to an unstressed syllable



Prioritized Order of Acoustic Cues for Urdu Stress Marking

- Duration of a vowel
- Stylized pitch track of a vowel
- Phrase initial glottalization
- Intensity of a vowel

Duration of a Vowel

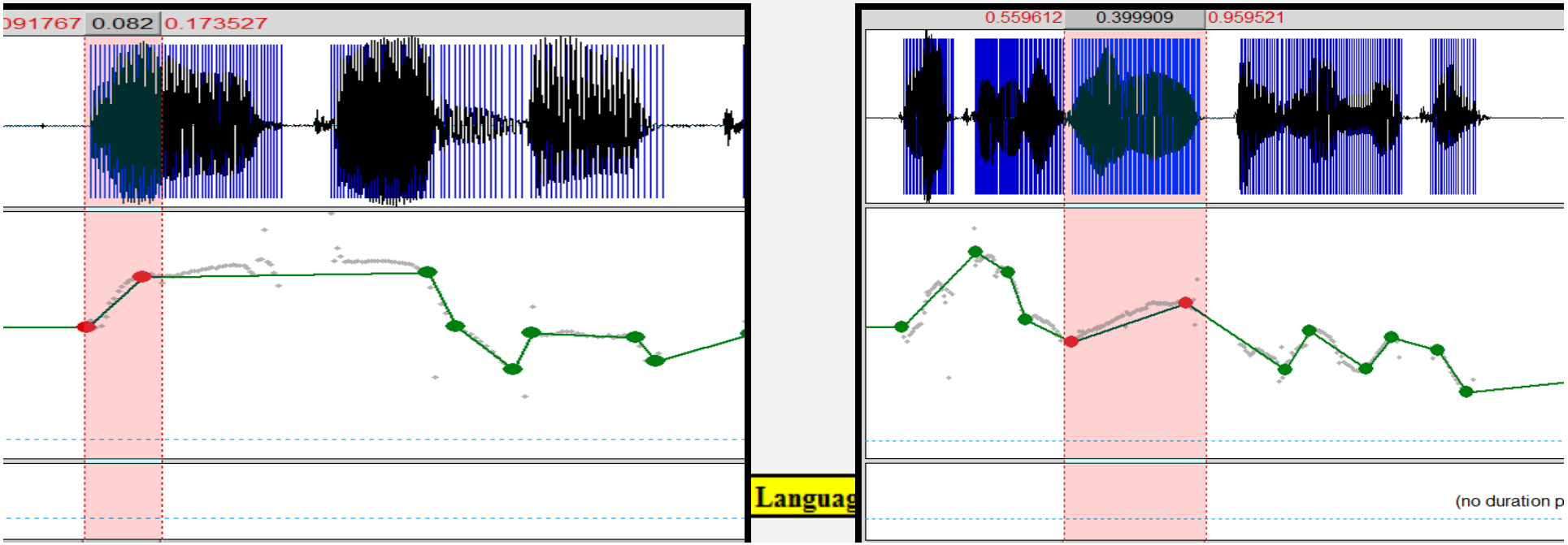
- Categorize the vowel
- Analyze the position of a vowel in a syllable
- Comparison with the same shortest vowel
  - Do not select a vowel which comes at the "final syllable with PAU" position
  - Short vowel duration = less than 57ms
  - Long vowel duration = less than 100ms
- Comparison with the similar shortest vowel

Durational Analysis of Urdu Vowels

VOWEL	Non-Final 0	Non-Final 1	Final 0	Final 1	Final with PAU 0	Final with PAU 1	Increased Duration at Non-final	Increased Duration at final	Increased Duration at final with pause
ə	57	81	61	86	75	107	24	25	32
e:	70	116	81	140	135	188	46	59	53
ɑ̃:	101	155	78	152	148	211	54	74	63
e	57	83	60	96	87	99	26	36	12
əi:	NA	134	113	195	201	245	NA	82	44

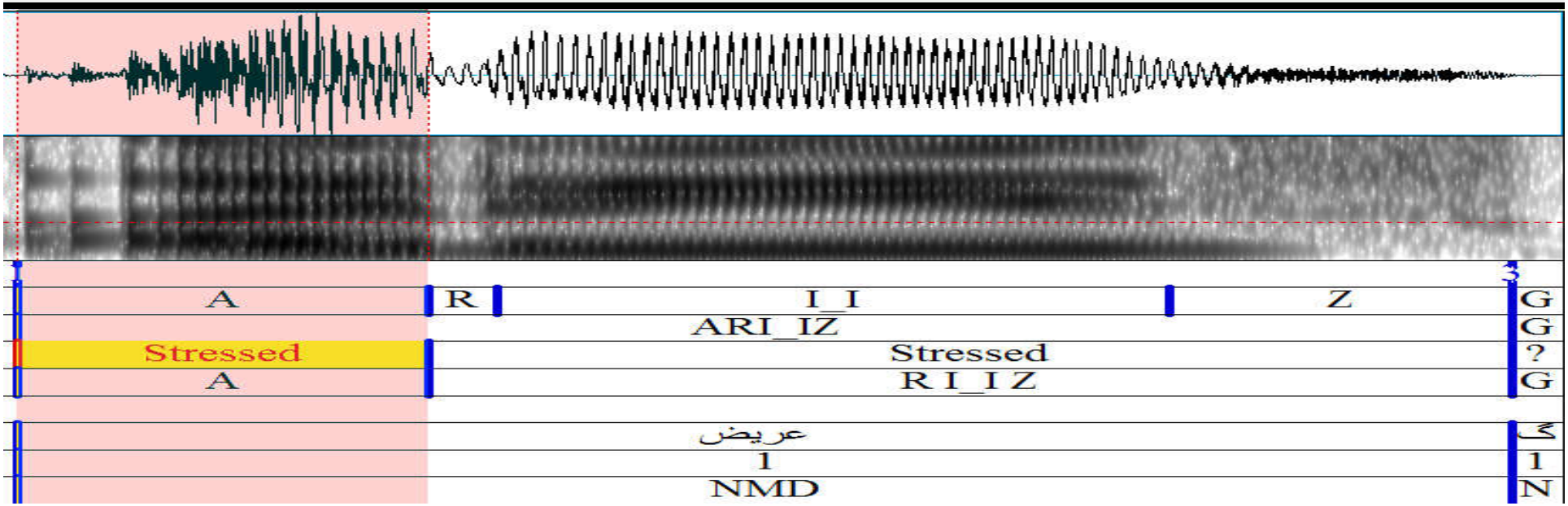
Pitch Contour

- The results indicate that falling or rising slope between L* and H* is abrupt and steep for stressed syllables in Urdu whereas it is gradual and flat for unstressed syllables.



Phrase Initial Glottalization

- **Phrase initial glottalization**
  - Strong glottalization
  - Weak glottalization
- **Phrase final glottalization**
  - Tapering off the vowel



Intensity of a Vowel

- It is observed that intensity of an accented syllable in Urdu is on average 3-5dB more than an unaccented syllable.
- However, the change in intensity with stress is vowel dependent.

Process of Assessing the Stress Tier

- Reference files generation
- Testing utilities to ensure that:
  - All the stress tier labels are from a defined numbering scheme (0, 1)
  - No interval is left unmarked
  - No change has been made at the automatically marked syllabification tier while annotating the stressed tier

Discussion

- Consonant Lengthening
- High intensity of a vowel
- Data scarcity issue in the wave file

Future Work

- Development of an algorithm
- Investigate the unexplored areas i.e., break index, secondary stress, emphatic stress and intonation pattern of Urdu language

Thank You

References

1. Klatt, Dennis H. "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence." *The Journal of the Acoustical Society of America*59.5 (1976): 1208-1221.
2. Laeufer, Christiane. "Patterns of voicing-conditioned vowel duration in French and English." *Journal of Phonetics* 20.4 (1992): 411-440.
3. Berkovits, Rochele. "Utterance-final lengthening and the duration of final-stop closures." *Journal of Phonetics* (1993).
4. Laukkanen, Anne-Maria, et al. "Physical variations related to stress and emotional state: a preliminary study." *Journal of Phonetics* 24.3 (1996): 313-335.