Multitier Annotation of Urdu Speech Corpus

Benazir Mumtaz, Amen Hussain, Sarmad Hussain, Afia Mahmood, Rashida Bhatti, Mahwish Farooq, Sahar Rauf

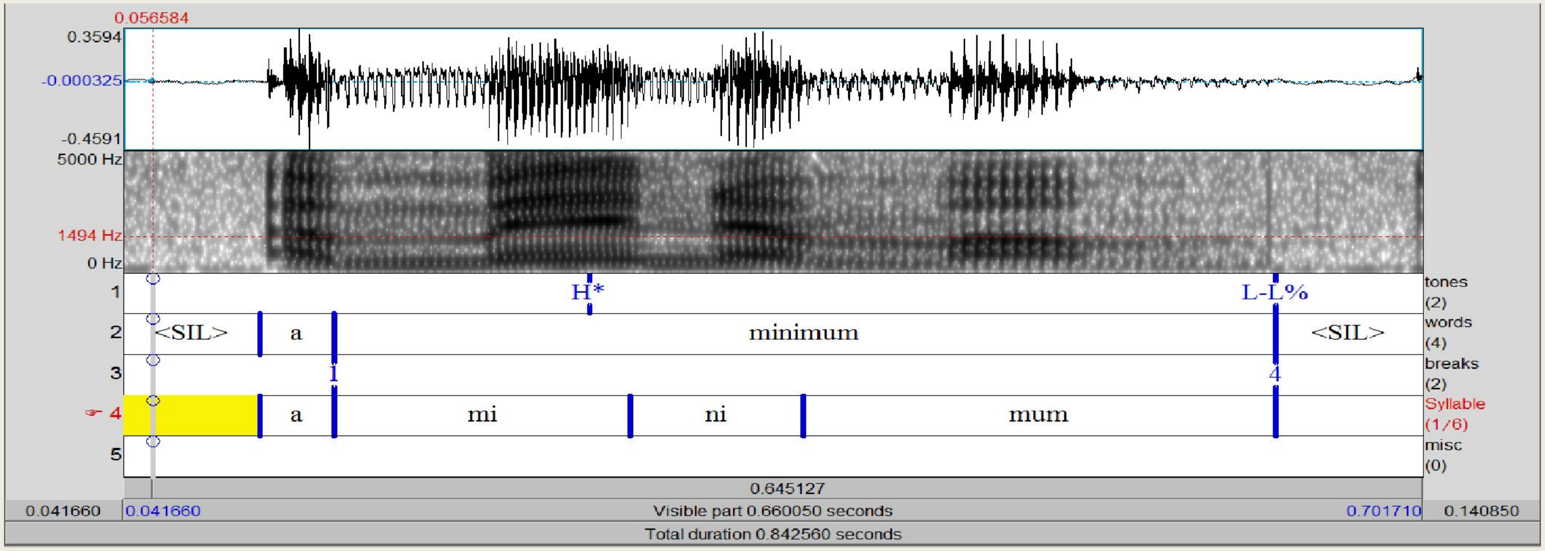
Centre for Language Engineering
Al-Khawarizmi Institute of Computer Science
University of Engineering and Technology Lahore, Pakistan

Contents

- What is an annotated speech corpus?
- Why is an annotated speech corpus useful?
- The process of annotating Urdu speech corpus at multiple tiers
- Speech annotation quality assessment process
- Results and Discussion

What is an Annotated Speech Corpus?

- Annotated or tagged speech corpus is an electronic corpus [1] which contains information about the language at phoneme, word, syllable, stress, phrase/ break index and intonation levels.



English Speech Corpus: OpenCourseWare (2006)

Why is an Annotated Speech Corpus Useful?

- To acquire acoustic-phonetic knowledge for phonetic recognition
- To provide speech for training recognizers
- To provide a common test base for the evaluation of recognizers

Description of Urdu Speech Corpus

- Speech Corpus Size: Thirty minutes
- Recording Sampling Rate: 8 kHz
- Software: PRAAT
- Phonetic Character Set: Case Insensitive Speech Assessment Method Phonetic (CISAMPA)

Multitier Annotation of Urdu Speech Corpus

1. Segment/Phoneme Level Annotation
2. Word Level Annotation
3. Syllable Level Annotation
4. Break Index/Phrase Level Annotation

1. Segment/Phoneme Level Annotation

- The process for segment marking layer describes how, when and where to split following combination of vowel and consonant:
  - Splitting of vowel and consonant sounds

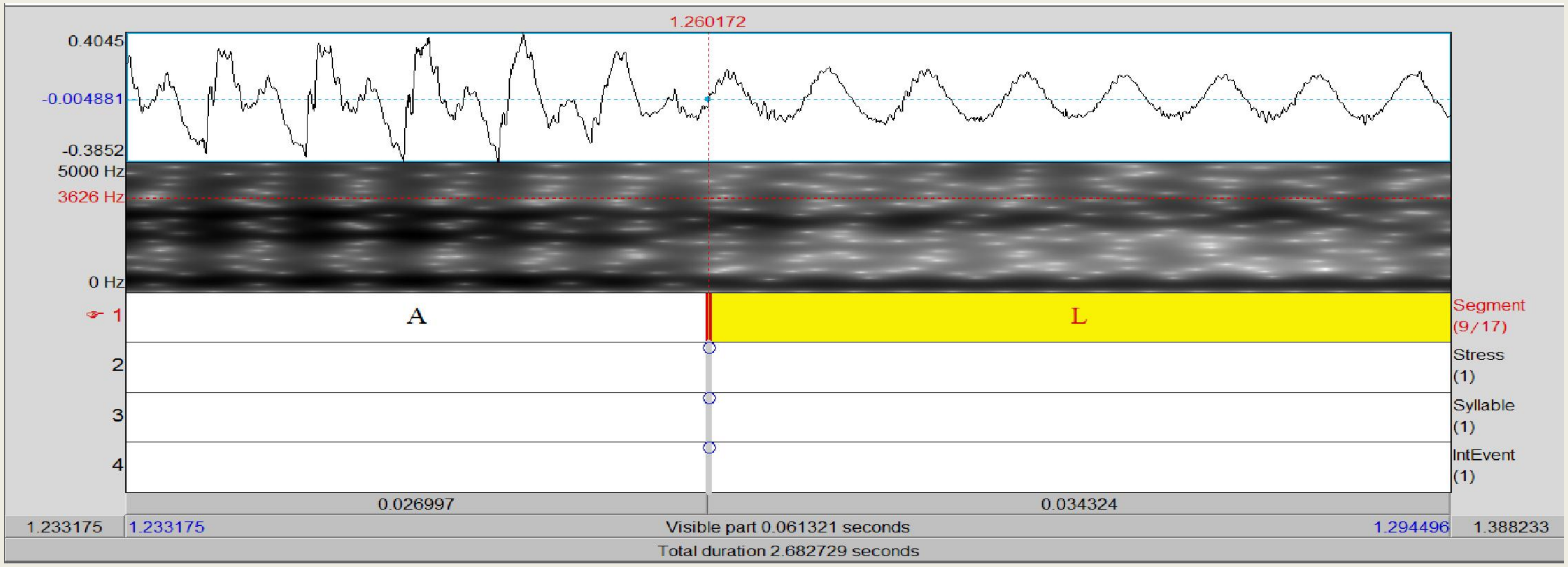


Fig: Splitting the vowel consonant Junction

Cont…

- Splitting the consonant cluster

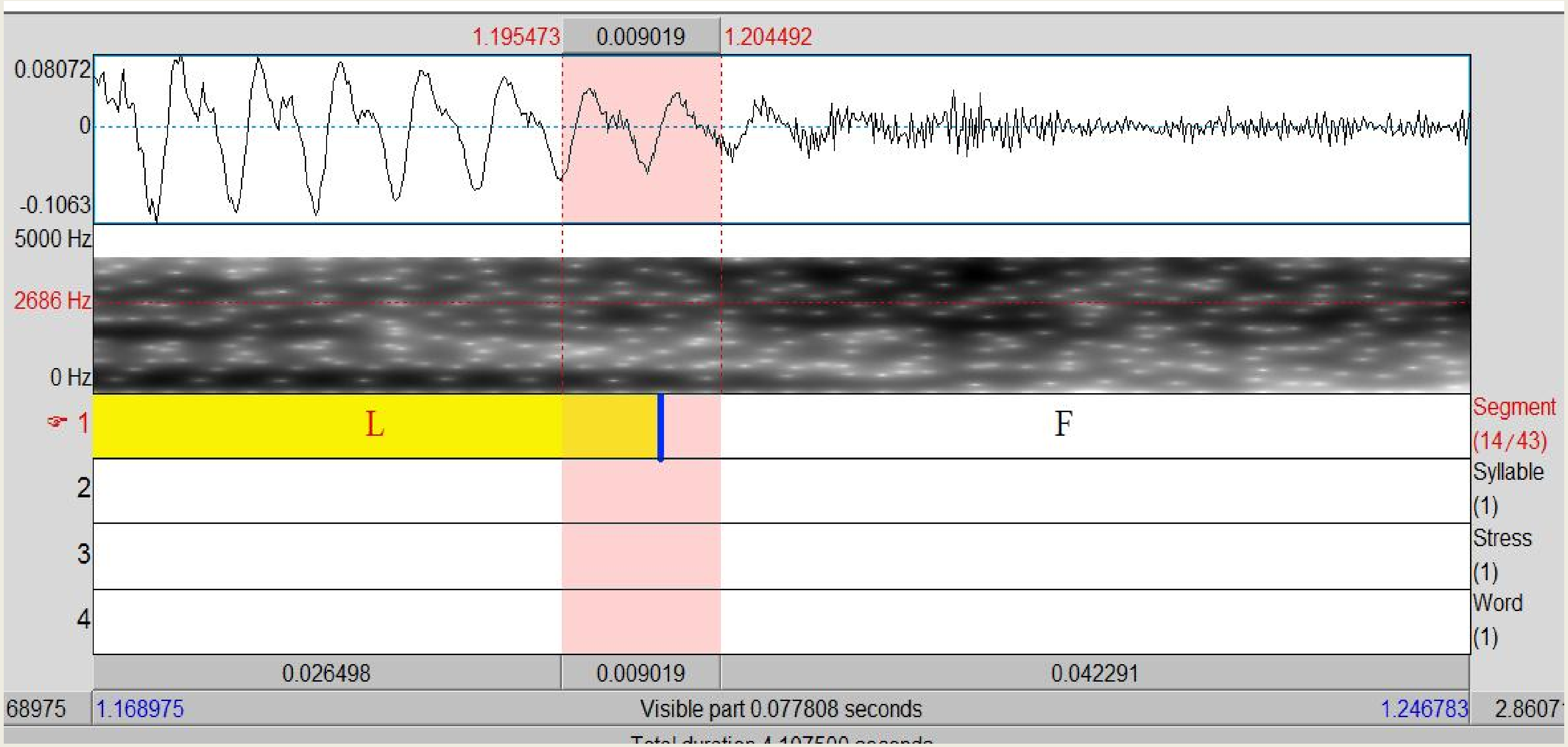


Fig: Splitting the Consonant cluster

Cont…

– Gemination across the words or within the word

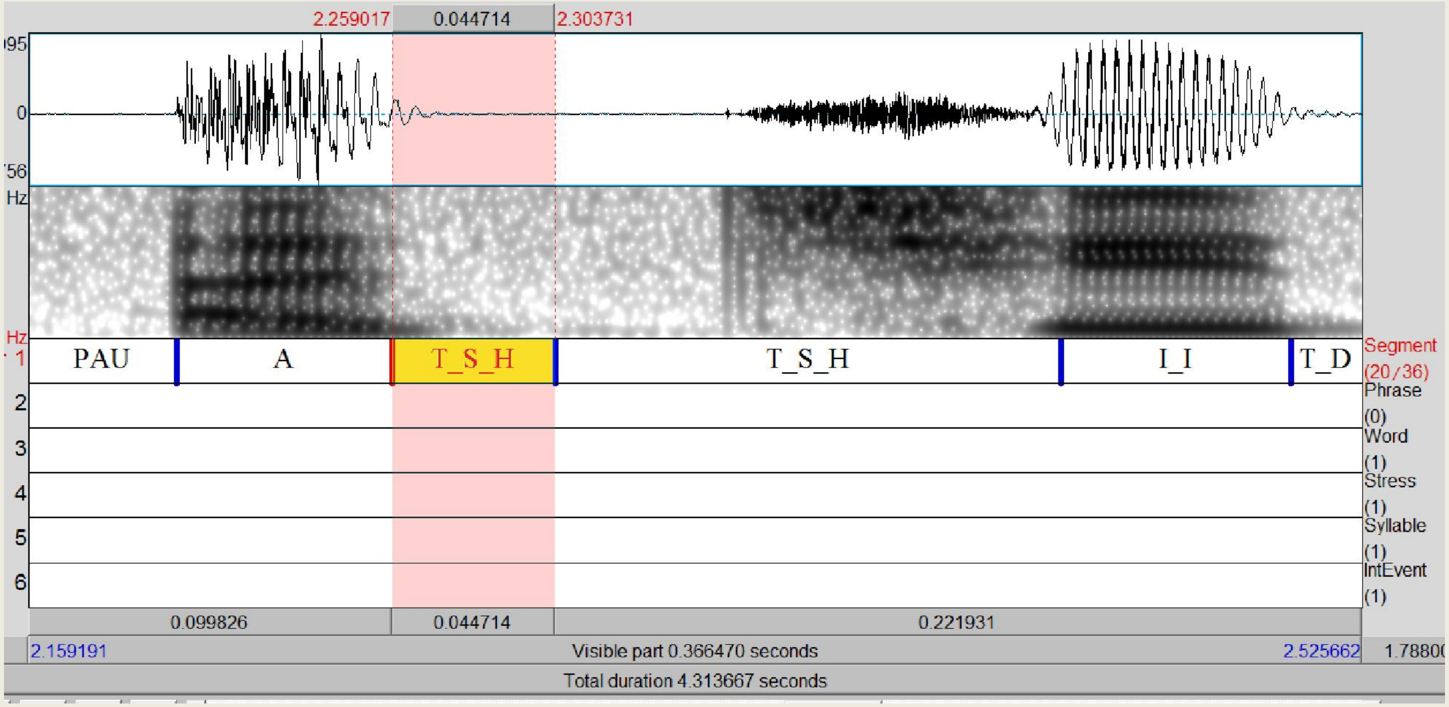


Fig: Splitting the Geminated Consonants

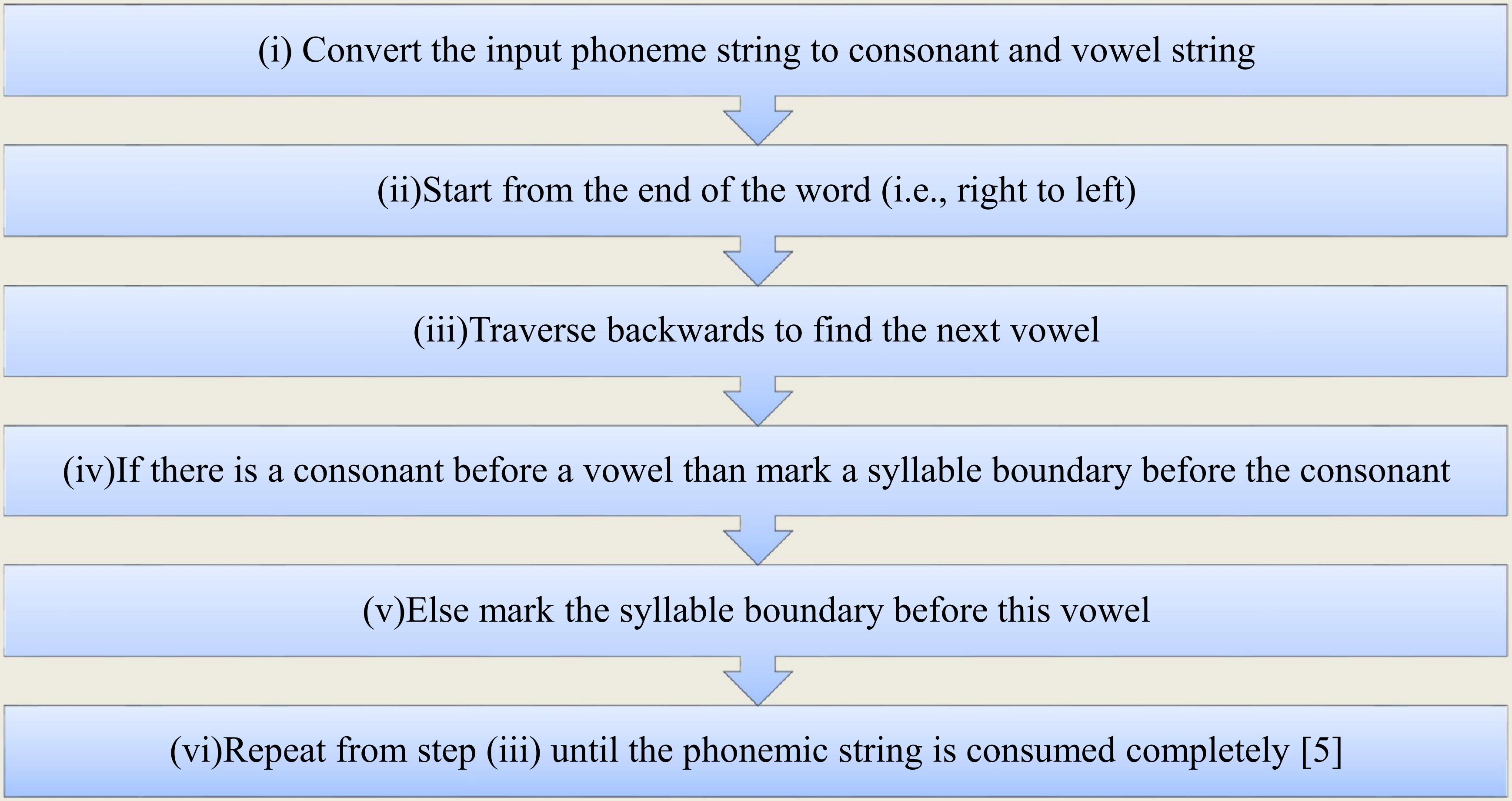
2. Word Level Annotation

- Two stages of Word Level Annotation
  - Listen to the wave file
    - Mispronunciation/misreading,
    - Insertion of extra phoneme in a word
    - Deletion of required phoneme from the word
  - Manual marking of the word boundaries

Principles Used to Mark the Boundaries Between Compound Words

- Meaningless Prefix + Meaningful Word (بہ معنی)
- Meaningful Words+ Meaningless Suffix (خیال آرائی)
- Meaningful + Meaningful Words combined with a Conjunction Vao “و" (غور و فکر)
- Compounds combined with یائے اضافت (دریائے راوی)
- Compounds combined with Zair (مخلوقِ خدا)

3. Syllable Level Annotation



(i) Convert the input phoneme string to consonant and vowel string

(ii)Start from the end of the word (i.e., right to left)

(iii)Traverse backwards to find the next vowel

(iv)If there is a consonant before a vowel than mark a syllable boundary before the consonant

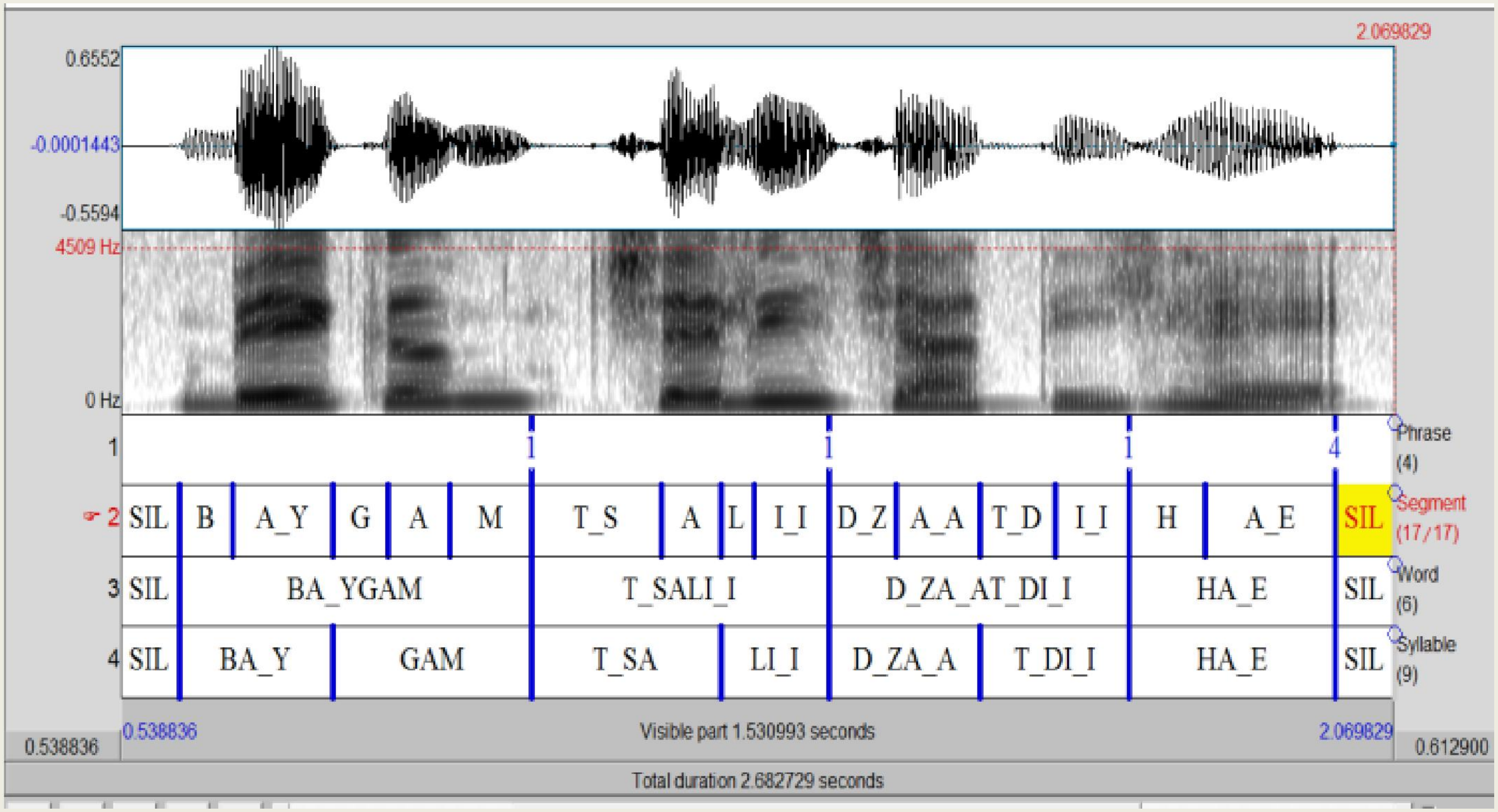
(v)Else mark the syllable boundary before this vowel

(vi)Repeat from step (iii) until the phonemic string is consumed completely [5]

4. Break Index/Phrase Level Annotation

- **Level 4:** Indicates the full intonational phrase boundary
- **Level 3:** Indicates the intermediate intonational phrase boundary (weak disjuncture, lengthening of the vowel of last syllable and glottalisation)
- **Level 2:** Indicates a disjuncture that is weaker than the intermediate or full intonational phrase boundary
- **Level 1:** indicates most phrase-medial word boundaries
- **Level 0:** indicates the boundary between the words from clitic groups[6]

A Sample of Annotated Speech Wave File


2.069829
0.6552
-0.0001443
-0.5594
4509 Hz
0 Hz
1
1 1 1 4
Phrase
(4)
2
SIL B A_Y G A M T_S A L I_I D_Z A_A T_D I_I H A_E SIL
Segment
(17/17)
3
SIL BA_YGAM T_SALI_I D_ZA_AT_DI_I HA_E SIL
Word
(6)
4
SIL BA_Y GAM T_SA LI_I D_ZA_A T_DI_I HA_E SIL
Syllable
(9)
0.538836
0.538836
Visible part 1.530993 seconds
2.069829
0.612900
Total duration 2.682729 seconds

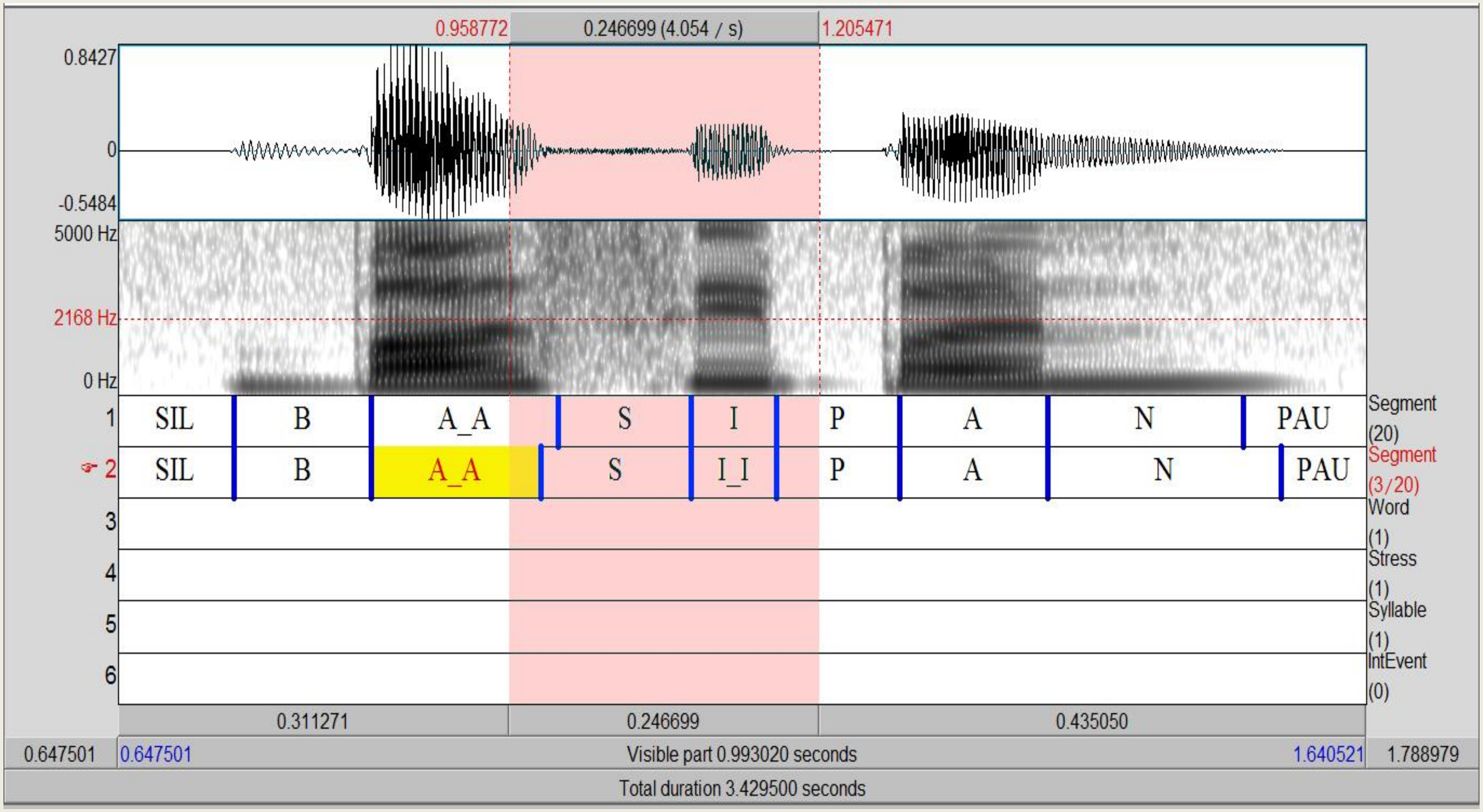
Speech Annotation Quality Assessment

1. Segment/Phoneme Level Assessment
2. Word Level Assessment
3. Break Index/Phrase Level Assessment

1. Phoneme/Segment Level Assessment

- Phoneme labels checking
- Phoneme boundaries checking using maximum string alignment algorithm

Reference File Generation



Phoneme Level Annotation Quality Assessment Results

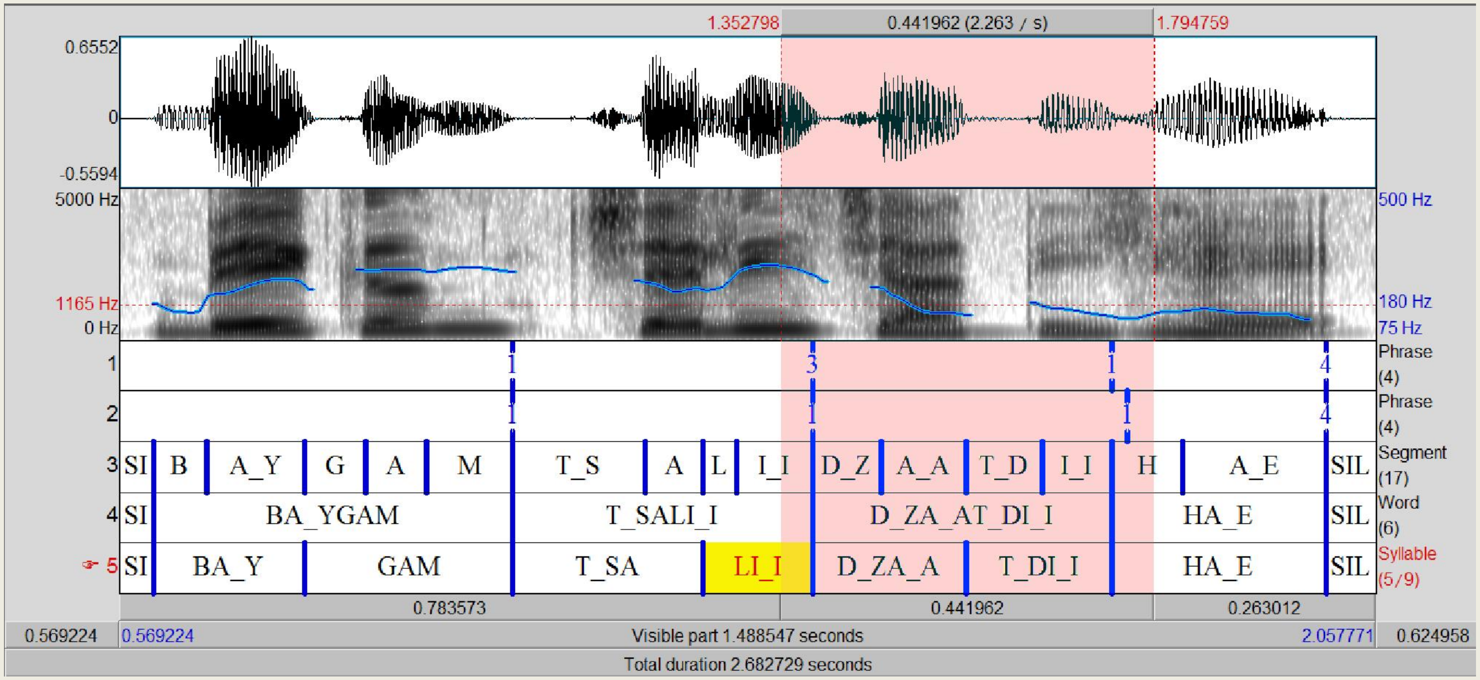
Annotation Quality Assessment Tests	Total Number of Phones	Total Number of Erroneous Phones	Percentage of Accuracy
Phoneme Label Comparison	19600	2083	89.37%
Phoneme Boundary Comparison	38162	11916	68.77%

2. Word Level Assessment

- Word label should not contain any non speech phoneme label; SIL, PAU
- The number of annotated words in the source file should be equal to the number of words in text file
- All the labeled words can be syllabified according to the Urdu syllabification rules
- The pronunciation of labeled word is compared with the standard Urdu pronunciation available in the pronunciation lexicon

3. Phrase level Annotation Assessment

- The time of break index in the source file is compared with a reference file
- The level of break indices marks are compared



Phrase Level Annotation Quality Assessment Results

Annotation Quality Assessment Tests	Total Number of Break Indices	Total Number of Erroneous Break Indices	Percentage of Accuracy
Break Index Level Comparison	5055	978	80.65%
Break Index Time Mark Comparison	9356	122	98.70%

Discussion

- Issues faced at Segment Level Annotation
  - Co-articulation Factor
  - Diphthongs
- Issues faced at Break Index/Phrase Level Annotation
  - Clitics

Current Status

- Guidelines, Testing process and Annotation completed
- Guidelines and Testing process decided
- Unexplored

	1st Hour	2nd Hour	3rd Hour	4th Hour	5th Hour	6th Hour	7th Hour	8th Hour	9th Hour	10 Hour
Phoneme Level Annotation	🟩	🟩	🟩	🟩						
Word Level Annotation	🟩	🟩	🟩							
Syllable Level Annotation	🟩	🟩								
Break Index Level Annotation	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
Stress Level Annotation	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
Intonation Level Annotation	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥

Acknowledgement

- This work has been conducted through the project,
  - Enabling Information Access for Mobile based Urdu Dialogue Systems and Screen Readers
  - ICTRnD Fund, Pakistan

Thank You

References

1. J. Matoušek, J. Romportl, "Recording and annotation of speech corpus for Czech unit selection speech synthesis". In Text, Speech and Dialogue. 2007. (pp. 326-333). Springer Berlin Heidelberg.
2. S. Kiruthiga, and K. Krishnamoorthy. "Annotating Speech Corpus for Prosody Modeling in Indian Language Text to Speech Systems." International Journal of Computer Science Issues (IJCSI) 9.1
3. H. Sarfraz, S. Hussain, R. Bokhari, A. A. Raza, I. Ullah, Z. Sarfraz, S. Pervez, A. Mustafa, I. Javed, R. Parveen. "Speech Corpus Development for a Speaker Independent Spontaneous Urdu Speech Recognition System." proceeding of OCOCOSDA (2010).

References

4. S. Hussain, "Phonetic Correlates of Lexical Stress in Urdu", PhD, Northwestern University, Illinois, 1997.
5. J. J. Venditti, "The J_ToBI model of Japanese intonation." Prosodic typology: The phonology of intonation and phrasing, 2005, 172-200.
6. S. Hussain, "Phonological Processing for Urdu Text to Speech System", Lahore: Center for Research in Urdu Language Processing, National University of Computer and Emerging Sciences, B Block, Faisal Town, Lahore, Pakistan.