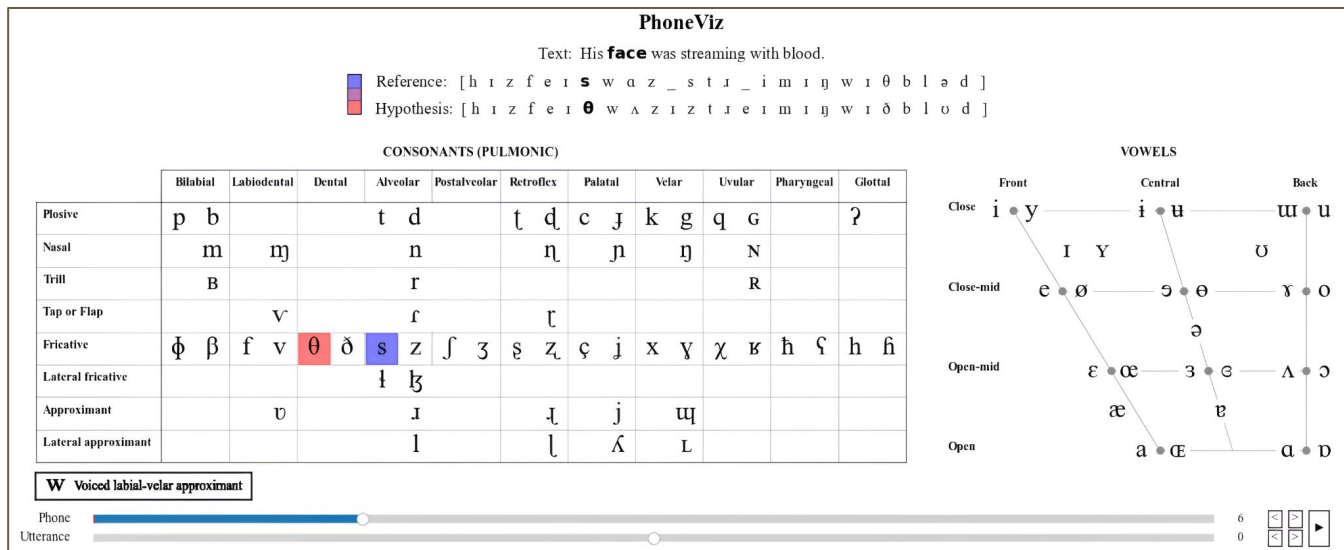

PhoneViz: Exploring Alignments at a Glance

Margot Masson, Erfan A. Shams, Iona Gessinger, Julie Carson-Berndsen
School of Computer Science, University College Dublin, Ireland

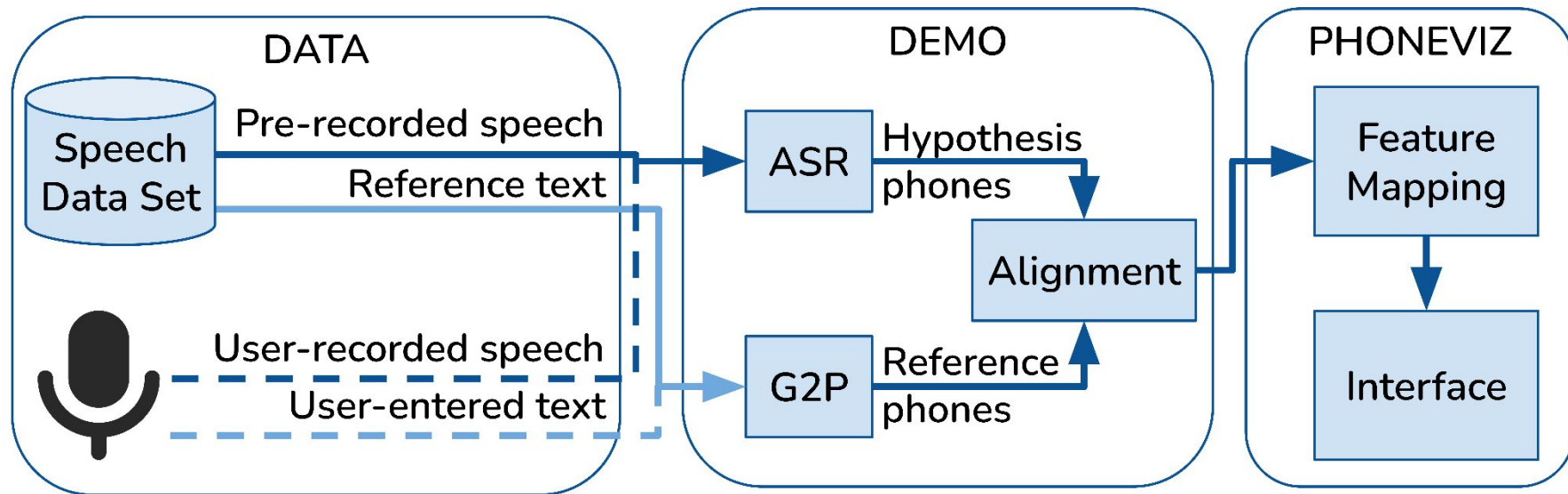
Show and Tell, INTERSPEECH 2024, Kos, Greece

Motivation



Providing articulatory feature information
to facilitate phonetic variation analysis

Demonstration pipeline



Demonstration pipeline

Mode 0

single utterance
w/o alignments

```
----- WELCOME TO PHONE VIZ -----
Type 0 to run single text mode, 1 to run dataframe mode or 2 for dataframe mode with alignments ready.0
Enter the name of the audio file : EBVS_arctic_a0001.wav
Enter reference text : Author of the danger trail, Philip Steels, etc.
Generating alignments |█████| 4 in 17.2s (0.16/s)
```

Mode 1

batch mode
w/o alignments

```
----- WELCOME TO PHONE VIZ -----
Type 0 to run single text mode, 1 to run dataframe mode or 2 for dataframe mode with alignments ready.1
Enter the dataframe name : l2arctic_head_without.csv
Generating alignments |█████| 4 in 16.3s (0.17/s)
Generating alignments |█████| 4 in 12.7s (0.22/s)
Generating alignments |█████| 4 in 12.0s (0.23/s)
Generating alignments |█████| 4 in 12.7s (0.22/s)
Generating alignments |█████| 4 in 12.1s (0.23/s)
```

file_name	reference_text
EBVS_arctic_a0001.wav	Author of the danger trail,
EBVS_arctic_a0032.wav	Since then some mysterious
EBVS_arctic_a0034.wav	Men of Selden's stamp don't
EBVS_arctic_a0035.wav	He stopped, and Philip nodd

Mode 2

batch mode
w/ alignments

```
----- WELCOME TO PHONE VIZ -----
Type 0 to run single text mode, 1 to run dataframe mode or 2 for dataframe mode with alignments ready.2
Enter the dataframe name : l2arctic_head_with.csv
```

file_name	reference_text	ref_phon_align	hyp_phon_align
EBVS_arct	Author of the danger t	['i@', 'i@', 'v', 'θ', 'ə', 'r'	['a', 'v', 'l', 'θ', 'ə', 'j', 'l'
EBVS_arct	Since then some mys	['s', 'l', 'n', 's', 'ð', 'e', 'r'	['s', 'æ', 'n', 's', 'ð', 'e', 'r'
EBVS_arct	Men of Selden's stam	['m', 'e', 'n', 'ə', 'v', 's', 't'	['m', 'e', 'n', 'o', 'v', 't', 'e'

Interaction with the interface

IPA chart with clickable phones that play the standard pronunciation

PhoneViz

Text: His **face** was streaming with blood.

Reference: [h ɪ z f eɪ **s** w a z _ s t r _ i m i ŋ w ɪ θ b l ə d]

Hypothesis: [h ɪ z f eɪ **θ** w ʌ z ɪ z t r eɪ m i ŋ w ɪ ð b l o d]

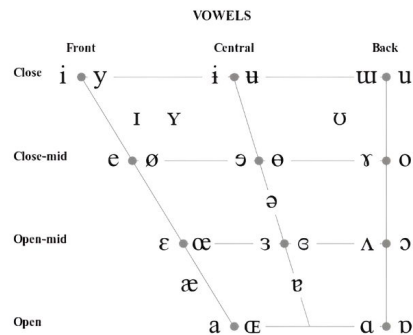
CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroluvial	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill		ʙ							ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

W Voiced labial-velar approximant

Reference text

Aligned reference and hypothesis phones



Navigation through *phones* and *utterances*

Utterance play back

Conclusion

- > PhoneViz is an interactive, customisable and easy-to-use interface
- > ... which offers a nice visualisation of the alignment between a reference and a hypothesis pronunciation
- > ... and helps to phonetically characterise sound substitutions for an in-depth analysis
- > ... making it interesting for both research and education purposes

PhoneViz: Exploring Alignments at a Glance

Margot Masson, Erfan A. Shams, Iona Gessinger, Julie Carson-Berndsen
School of Computer Science, University College Dublin, Ireland

Show and Tell, INTERSPEECH 2024, Kos, Greece
