Using phase to recognize phonemes in the brain

Rui Wang

A Symposium on the Occasion of Patrick Suppes' 90th birthday

Background – Why phoneme?

- The smallest segmental unit of speech that differentiates meaningful words.
- Most languages or dialects have very limited number of phonemes. (20 to 60)
- Most commonly used unit to model the acoustic properties of speech, speech recognition being a good example.

Experimental data

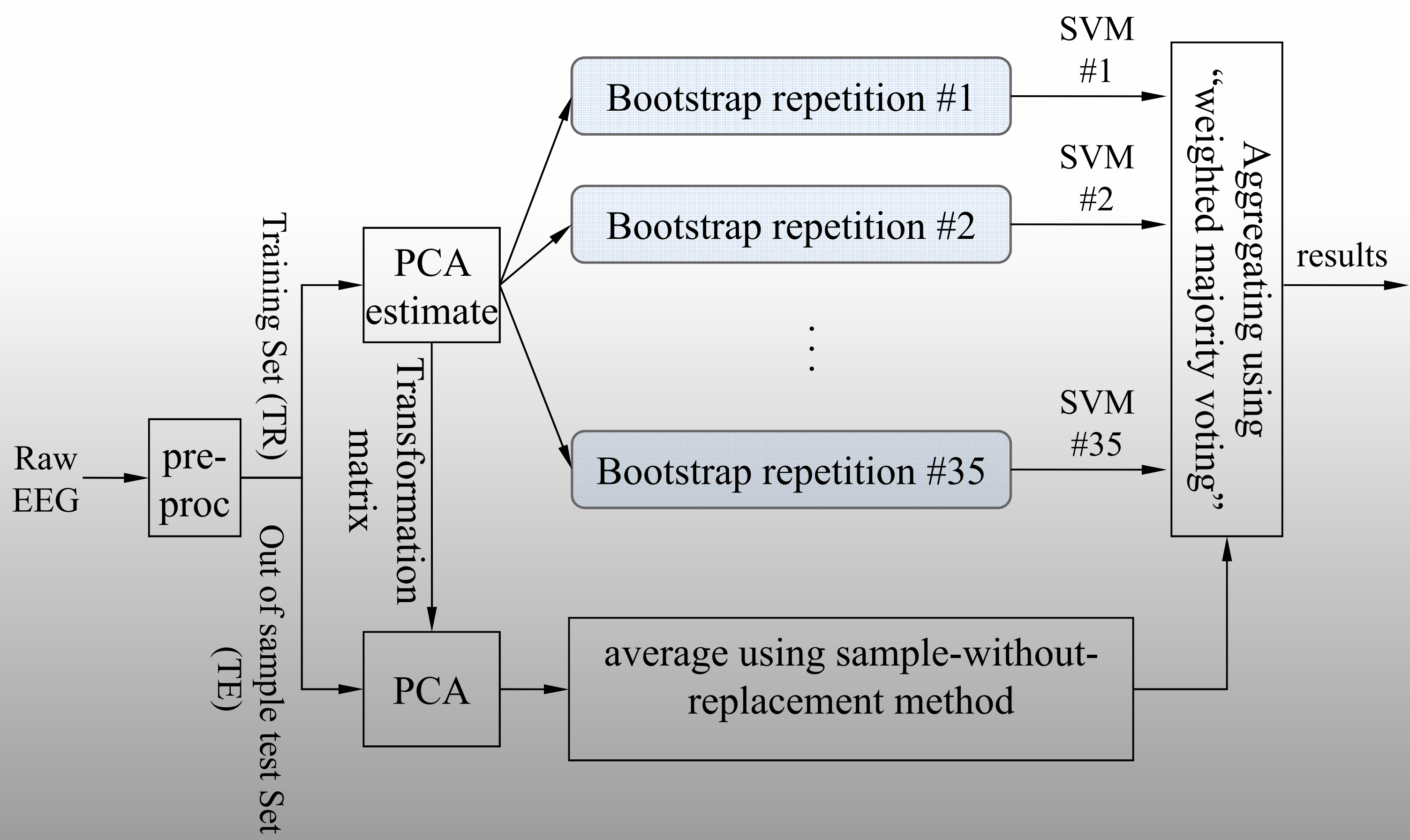
- Syllables
  - 32 consonant-vowel (CV) format syllables, made up by 8 consonants /p/, /t/, /b/, /g/, /f/, /s/, /v/ and /z/, and 4 vowels /i/, /æ/, /u/ and /ɑ/.
  - 128 sensors, 1000Hz, 124 monopolar, 2 reference channels of eye movements
  - 4 subjects. 21,540 trials in total.
- Isolated vowels
  - 4 vowels /i/, /æ/, /u/ and /ɑ/.
  - 1 subject, 7,168 trials in total.

Distinctive features of the phonemes

	voiceless		voicing	
	Labial	Alveolar	Labial	Alveolar/Velar
stop	p	t	b	g
fricative	f	s	v	z

		backness	
		Front (High F2)	Back (Low F2)
height	Close (Low F1)	i	u
	Open (High F1)	æ	ɑ

SVM-with-Bootstrap Aggregating (Bagging) recognizer



Recognizing 8 initial consonants

(384 out-of-sample test samples)

Models		Rate (%)	p-value
Temporal signal (TIME)		37.5	$<10^{-34}$
Spectral features	DFT	38.0	$<10^{-35}$
	Amplitude	11.2	0.801
	Phase	36.5	$<10^{-31}$
Phase (2-9Hz)		50.3	$<10^{-71}$
Phase (2-9Hz) of EFLap		60.7	$<10^{-108}$
Phase (2-9Hz) of EFLap at 5 locations		64.6	$<10^{-124}$

Recognizing vowels

- Recognizing isolated vowels (140 test samples)

$$\mathrm{P}(\hat{V} = V \mid \text{isolated}) = 90.0\%$$

- Recognizing vowels in CV syllables (384 test samples)

$$\mathrm{P}(\hat{V} = V \mid \text{CV context}) = 45.8\%$$

- Recognizing vowels in CV syllables when the initial consonant is known. (384 test samples)

$$\mathrm{P}(\hat{V} = V \mid \text{CV context, C is known}) = 62.8\%$$

Recognizing Syllables

- Use one 32-class SVM-with-Bagging recognizer for 32 syllables. .

  40.6% (p-value$<10^{-100}$)

- Two-pass strategy: Use an 8-class classifier for initial consonants and recognize the vowel using the consonant predictions in the second pass.

  39.1% (p-value$<10^{-100}$)

Similarity differences of phonemes

- Invariant similarities between brain and perceptual representations of phonemes
  - Robust discrimination of *voicing*
  - Robust discrimination of *affrication* in the voiceless consonants.
  - *Vowel-height* is a more robust feature than *vowel-backness.*

Recognizing distinctive features

	feature	grouping		test samples	rate (%)	p-value
consonant	voicing	voiceless	/p/ /t/ /f/ /s/	384	83.5	$<10^{-42}$
		voiced	/b/ /g/ /v/ /z/			
	continuant	stop	/p/ /t/ /b/ /g/	384	85.4	$<10^{-47}$
		fricative	/f/ /s/ /v/ /z/			
	place (labial)	labial	/p/ /b/ /f/ /v/	384	74.4	$<10^{-20}$
		non-labial	/t/ /g/ /s/ /z/			
	place (coronal)	coronal	/t/ /s/ /z/	384	80.7	$<10^{-34}$
		non-coronal	/p/ /b/ /g/ /f/ /v/			
vowel	height	open	/æ/ /ɑ/	140	93.6	$<10^{-28}$
		close	/i/ /u/			
	backness	front	/æ/ /i/	140	85.0	$<10^{-17}$
		back	/ɑ/ /u/			

Recognizing phonemes/syllables using distinctive features

- Parallel structure:

  Use one binary classifier to classify each distinctive feature.

- Hierarchical structure:

  The decision rules of distinctive features are conditional on their position in the hierarchy.

	parallel		**hierarchical**	
	rate (%)	p-value	rate	p-value
8 consonants	49.0	$<10^{-66}$	60.9	$<10^{-100}$
32 syllables	19.0	$<10^{-33}$	34.9	$<10^{-97}$

Conclusion and prospects

- EEG signal components from the primary auditory cortex relevant to phoneme perception in the brain:
  - Phase of 2-9 Hz
  - Electrical field & surface Laplacian vectors.
- Results suggest the importance of distinctive features in the mechanisms that the brain uses to perceive phonemes
- Success in recognizing CV syllables sequentially suggests modeling words using Hidden Markov Models.