Neural encoding of talker-specific phonetic variation

Emily Myers[1,2], Rachel M. Theodore[1,2], Sahil Luthra[3]
[1]University of Connecticut, [2]Haskins Laboratories, [3]Brown University

Background

The same acoustic cues of the speech signal carry information about phonetic identity (/g/ or /k/) and talker identity (Joanne or Sheila)

- The acoustic parameter voice-onset-time (VOT) cues the voicing distinction (/g/ or /k/) in word-initial English stop consonants
- Talkers show systematic variation in their production of some speech sounds, including VOT variants for voiceless stops (e.g. /k/, Theodore et al., 2009)
- Listeners are sensitive to talker-specific phonetic variation (Theodore & Miller, 2010; Goldinger, 1996)
- Listeners use talker-specific phonetic variation to guide processing:
  - Ambiguous tokens during exposure will produce shifts in the **category boundary** (e.g. Norris, et al., 2003; Kraljic & Samuel, 2008) and unambiguous tokens will produce shifts in **internal category structure** (Theodore et al., 2015)
  - In accented speech, exposure to ambiguous or shifted tokens produces speeded lexical decisions to consistent words, shifts in the category boundary, and shifts in 'goodness' judgments for non-standard tokens (e.g. Eisner et al., 2013; Xie et al., under review)

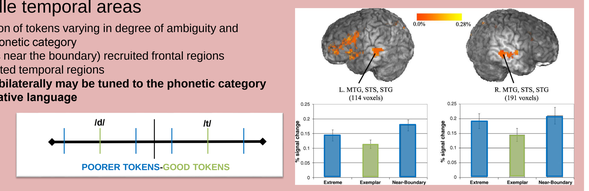
Does the neural system separate processing of VOT for phonetic identity and talker identity?

- Left and right temporal regions tune to the phonetic category structure (e.g. Myers, 2007, see inset, above)
- Right middle temporal & frontal regions respond to the shifted category boundary that results from exposure to an ambiguous token (see inset, right)

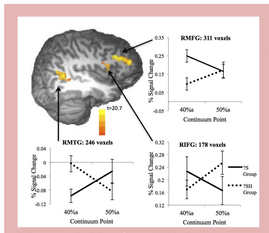
→ Will listeners recruit the same right hemisphere regions for perception of VOT variants that are unambiguous (that is, do not require or result in a shift in phonetic category boundary)?
→ Is there a core neural system for processing talker-specific phonetic variation?

Sensitivity to phonetic category variability (i.e. "goodness of fit") in bilateral superior temporal / middle temporal areas

- Phonetic categorization of tokens varying in degree of ambiguity and 'goodness of fit' to phonetic category
- Ambiguity (i.e. tokens near the boundary) recruited frontal regions
- Goodness of fit recruited temporal regions
- **Posterior MTG/STS bilaterally may be tuned to the phonetic category structure of one's native language**

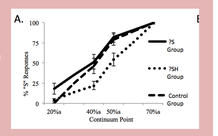


Myers, 2007



Sensitivity to ambiguous talker-specific variants encoded in right hemisphere temporal/frontal areas

- Perceptual learning for speech (e.g. Norris, et al., 2003, Kraljic & Samuel, 2008)
- Listeners hear an ambiguous s/sh blend inserted in either an s-biasing (e.g. 'epi?ode') or sh-biasing (e.g. 'flouri?ing') context
- Later interpret that ambiguous token according to the exposure (i.e. shift in the phonetic category boundary)
- **Modulation in right-hemisphere regions also involved in processing talker identity** (e.g. von Kriegstein, et al., 2003)



Myers & Mesite, 2014

Methods

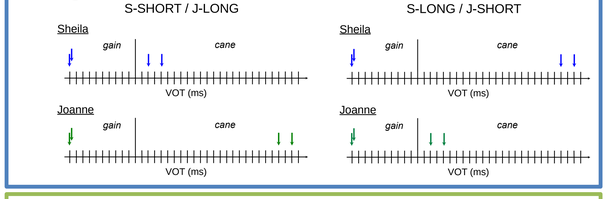
Participants

- 15 Monolingual English listeners, assigned to one of two training groups:
  - S-SHORT/J-LONG (n = 7)
  - S-LONG/J-SHORT (n = 8)

Stimuli

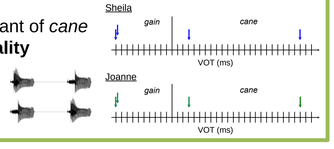
- Two synthesized VOT continua ranging from *gain* to *cane*
- Continua were based on the speech of two female talkers, fictitiously named "Joanne" and "Sheila"
- Continua were created using naturally-produced *gain* tokens; word duration and VOT were equated across talkers
- Subsets of tokens were used during **training** and **test** phases

Training stimuli



Test stimuli

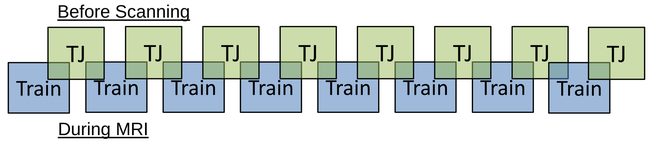
- Short-VOT and long-VOT variant of *cane* used to create pairs for **typicality judgement** task
- All variants used in **phonetic categorization** tests



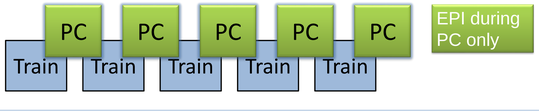
Procedure

- **Talker Training** (out-of-scanner and alternating with scanning):
  - Heard *gain* or *cane* on a given trial
  - Asked to identify the initial sound and the talker
  - Feedback was provided for the talker choice only
- **Typicality Judgement** (**TJ**, out-of-scanner test): Presented with short-VOT and long-VOT pair and asked to choose which is most like each talker (2AFC)
- **Phonetic Categorization** (**PC**, in-scanner task): Presented *gain*, short-VOT *cane*, and long-VOT *cane* for both talkers and asked to categorize as 'GAIN' or 'CANE'

Before Scanning



During MRI

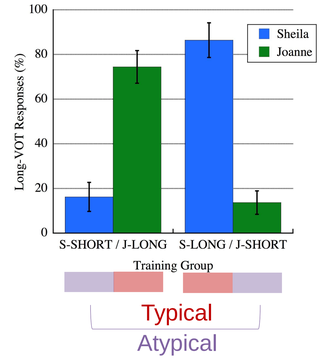


MRI Methods

- EPI and structural images acquired with a Siemens 3T Tim Trio
- EPI: Sparse design, 2x2x2.5 mm voxels, 27 slices
- Functional preprocessed according to standard processing stream
- Beta estimates of each condition submitted to two ANOVAs (Voice X Typicality) and (Voice X VOT)

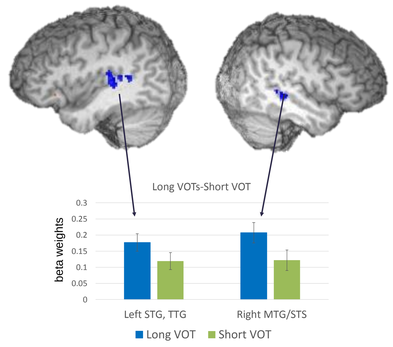
Results

Listeners are behaviorally sensitive to the "typicality" of VOT variants as representative of a talker's voice.

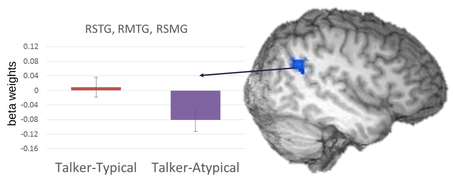


During pre-scanning typicality judgment (TJ): percentage of Long-VOT responses that are judged to be most typical of that talker's voice.

Right and Left posterior temporal regions sensitive to phonetic category structure (Long>Short variant, see Myers, 2007)



Typicality of a token as a member of a talker's voice modulates activity in right temporo-parietal regions also implicated in adaptation to ambiguous tokens (see Myers & Mesite, 2014)



All clusters significant at whole-brain level, p<0.05, cluster-corrected for multiple comparisons, voxel-level p<0.025, 112 contiguous voxels

Atypical - Typical

Voxels	Peak x	Peak y	Peak z	Region	t-value at maximum
323	-42.9	55.2	24.9	RSTG, RMTG, RSMG	-4.46
192	0.9	44.8	26.6	L Post Cingulate, L Cingulate	-4.39

Joanne-Sheila

Voxels	Peak x	Peak y	Peak z	Region	t-value at maximum
254	-25.4	46.5	31.9	R IPL	-6.57
161	-11.4	55.2	35.4	R Precuneus	-3.65
142	44.6	-6	-10.1	LSTG LIFG	-4.77
136	41.1	-41	-1.4	L MFG	-5
135	49.9	4.5	-13.6	LMTG LSTG	-5.56

Long-Short

Voxels	Peak x	Peak y	Peak z	Region	t-value at Maximum
359	-2.6	-18.2	12.6	Right subcortical	-5.56
268	-56.9	36	-6.6	Right MTG/STG	-3.48
227	39.4	25.5	17.9	Left MTG/STG	-4.78
222	2.6	-34	21.4	Anterior Cingulate	-5.21

Discussion

- A core region for linking talker identity to phonetic variation may occupy the right temporoparietal region:
  - Modulates as a function of typicality of a token as representative of a talker's voice
  - These regions overlap with areas involved in perceiving ambiguous stimuli following lexically-conditioned phonetic category boundary shifts (Myers & Mesite, 2014).
- This system is adjacent to, but does not overlap with, regions responsible for processing phonetic category structure.
  - Left and right superior temporal gyrus/sulcus are sensitive to within-category differences more generally, decoupled from talker information (see also Myers, 2007).
- Suggests that adaptation to talker-specific variation, while resulting in widespread perceptual/processing adjustments, does not fundamentally retune temporal lobe sensitivities, at least over short exposures
- Longer-term adaptation may ultimately result in temporal lobe retuning

References

1. Eisner, F., Melinger, A., & Weber, A. (2013). Constraints on the transfer of perceptual learning in accented speech. *Frontiers in Psychology, 4*,148.
2. Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1166-1183.
3. Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language, 56*, 1-15.
4. Myers, E. B. (2007). Dissociable effects of phonetic competition and category typicality in a phonetic categorization task: An fMRI investigation. *Neuropsychologia, 45*(7), 1463-1473.
5. Myers, E. B., & Mesite, L. M. (2014). Neural systems underlying perceptual adjustment to non-standard speech tokens. *Journal of Memory and Language, 76*, 80-93.
6. Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. Cognitive Psychology, 47(2), 204-238.
7. Theodore, R. M., & Miller, J. L. (2010). Characteristics of listener sensitivity to talker-specific phonetic detailia). *The Journal of the Acoustical Society of America, 128*(4), 2090-2099.
8. Theodore, R. M., Miller, J. L., & DeSteno, D. (2009). Individual talker differences in voice-onset-time: Contextual influences. *The Journal of the Acoustical Society of America, 125*(6), 3974-3982.
9. Theodore, R.M., Myers, E. B., Lomibao, J.A. (2015). Talker-specific influences on phonetic category structure. *Journal of the Acoustical Society of America, 138*(2), 1068-1078.
10. von Kriegstein, K., Eger, E., Kleinschmidt, A., & Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Cognitive Brain Research, 17*(1), 48-55.
11. Xin, X., Theodore, R. M., & Myers, E. B. (Under review). More than a boundary shift: Perceptual adaptation to foreign-accented speech reshapes the internal structure of phonetic categories.

This work was supported by NIH NIDCD R03 DC009395, R01 DC013064, Myers, PI. The content is the responsibility of the authors and does not necessarily represent official views of the NIH or NIDCD.