Gestural control in the English past-tense suffix: an articulatory study using real time MRI

Adam Lammert[1], *Erik Bresch*[3], *Louis Goldstein*[2], *Shrikanth Narayanan*[1,2,3]
Departments of Computer Science[1], Linguistics[2] and Electrical Engineering[3]

USC VITERBI SCHOOL OF ENGINEERING

USC COLLEGE OF LETTERS, ARTS AND SCIENCES

web: sail.usc.edu/span
email: lammert@usc.edu

In certain allomorphs of the English past-tense, a vocoid appears before the final coronal closure.

English past-tense involves a tongue tip closure gesture. When preceded by a voiced coronal stop, an *epenthetic* vocoid appears (e.g., needed). Traditionally, this has been transcribed as a schwa (/ə/) or as a barred-i (/ɨ/) [1].

This vocoid may lack a specific gestural target.

It may simply result from constraints relative to separating two voiced coronal closures. EMA studies have provided evidence consistent with this hypothesis [2].

Epenthetic schwa displays a vocal tract configuration similar to a lexical schwa, on average.

(a) Lexical Mean, MB (b) Epenthetic Mean, MB

(c) Lexical Mean, BP (d) Epenthetic Mean, BP

Figure 1: Midsagittal vocal tract images, showing the mean shapes for two subjects producing lexical and epenthetic schwa.

This may explain the traditional ə-notation. Moreover, some instances show slight tongue raising, which is consistent with the ɨ-notation.

We utilized real time MRI to further investigate the targetless nature of epenthetic schwa.

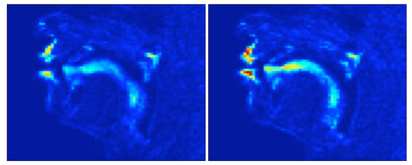
Real time MRI provides a more complete view of the vocal tract than EMA can provide [3]. Two subjects were imaged. Stimuli consisted of paired sentences with embedded schwas.

Sentences were structured according to:

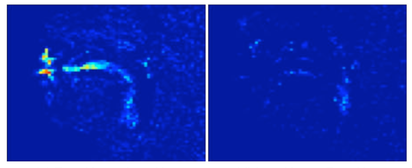
Schwa	Example
V_2	$..../V_1dV_2dV_1/....$
Lexical	"If Ch**eetah'd e**ven known"
Epenthetic	"If ch**eated e**ven once"

where V_1 was from the set {ɑ, ɛ, i, oʊ, u, æ, ɪ, eɪ}.

Since it is untargeted, epenthetic schwa's shape is significantly more variable than lexical schwa's.



(a) Lexical Variance (b) Epenthetic Variance

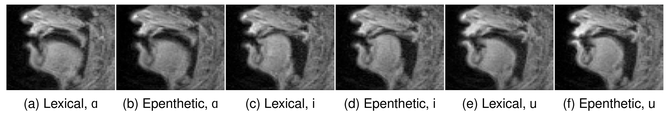


(c) Epe. Var. > Lex. Var. (d) Lex. Var. > Epe. Var.

Figure 2: Difference in pixel-wise variance for both schwa types (subject MB).

Images representing $/dV_2d/$ were segmented, and the center-most frame was chosen to represent schwa. Pixel-wise variances were directly compared to establish a difference. Using Wilcoxon's signed rank test, $p \ll 0.001$ for both subjects. Crucially, this difference could not be attributed to differences in duration. We tested for differences in duration using Wilcoxon's rank-sum test, and found that $p > 0.25$ for both subjects. The median duration of each schwa type was 2 frames.

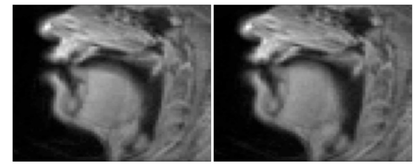
The shape of both lexical and epenthetic schwa is highly influenced by vowel context.



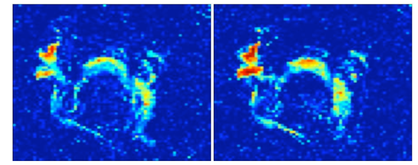
(a) Lexical, ɑ (b) Epenthetic, ɑ (c) Lexical, i (d) Epenthetic, i (e) Lexical, u (f) Epenthetic, u

Figure 3: The mean shape of lexical and epenthetic schwas in different vowel contexts.

Since it is untargeted, epenthetic schwa's shape is significantly more correlated with the vowel context.



(a) Lexical Mean, MB (b) Epenthetic Mean, MB



(c) Lexical COD (d) Epenthetic COD

Figure 4: The Coeff. of Determination between schwas and vowel context for subject MB. In comparing figures *c* and *d*, we can see higher values for: Labial (74% vs. 99%), Tongue Dorsum (51% vs. 83%), Tongue Root (48% vs.72%).

We calculated the coefficient of determination (r^2) between pixel-wise intensity variations in the schwa images and a set of exemplar vowels. Using Wilcoxon's rank sign test, r^2 was found to be significantly higher for epenthetic schwas ($p < 0.05$). The test was performed only on pixels relevant to vocal tract movement, which were isolated by looking at pixels with the highest variance ($>$ 90th percentile).

Phonological Conclusion: Not all articulatory events are controlled or phonologically specified.

Methodological Conclusion: Pixel-wise image analysis affords direct insights into variance and correlation in vocal tract coordination.

References

[1] Flemming, E., & Johnson, S. (2007). Rosa's roses: reduced vowels in American English. JIPA, 37:83.

[2] Smorodinsky, I. (2001). Schwas with and without active gestural control. JASA 109, 2446.

[3] Narayanan, S., Nayak, K., Lee, S., Sethy, A., & Byrd, D. (2004). An approach to real-time magnetic resonance imaging for speech production. JASA 115:1771.

Acknowledgements

This work was supported by NIH Grant DC007124.