Gradient sensitivity to acoustic detail and temporal integration of phonetic cues

Bob McMurray, Meghan A. Clayards, Richard N. Aslin, and Michael K. Tanenhaus
Department of Brain and Cognitive Sciences, University of Rochester

Introduction

During online word recognition, the acoustic information underlying phonological segments rarely arrives at the same time. In many cases, significant information concerning segment identity arrives before or after the point in time at which the segment would be traditionally labeled.

It has long been known that voicing (e.g. b/p), and manner (e.g. b/w) judgments are affected by the length of the *subsequent* vowel (e.g. Summerfield, 1981; Miller, Aibel & Green, 1984). Moreover, coarticulatory cues to nasals and rhotics can be found in the vowels that *precede* such consonants (Clark & Hillenbrand, 2003; Lahiri & Marslen-Wilson, 1991).

The computational processes by which such subphonemic acoustic cues are *integrated over time* are thus of fundamental importance to speech perception.

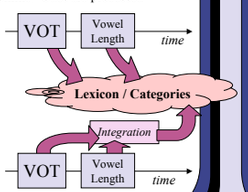
Recent work (McMurray, Tanenhaus & Aslin, 2002; McMurray, Tanenhaus, Aslin & Spivey, 2003) has shown that eye-movements to pictured lexical competitors are *systematically sensitive* to within-category variation in Voice Onset Time (VOT). Moreover, this task is sensitive to fine-grained temporal detail (e.g. Salverda, Dahan & McQueen, 2003; Dahan & Tanenhaus, 2004; Dahan, Magnuson, Tanenhaus & Hogan, 2001).

Purpose: Use head-mounted eye-tracking to examine the relative timing of listeners' sensitivity to asynchronous cues like VOT and vowel length as a way to understand the problem of *temporal integration*.

Cue Integration

Speech cues that show temporal dependency could be integrated in either of two ways:

1) **Immediate cue integration:** each cue is evaluated *as it arrives,* and the partial probabilities of potential lexical candidates are updated online. Cues are processed *independently.*
2) **Contextual cue integration:** cues are evaluated *with respect to each other* first and the result influences the partial probabilities of lexical candidates. Cues are processed *non-independently.*



Methods

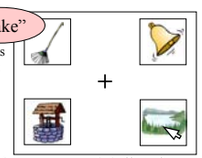
Materials

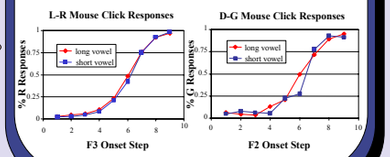
- synthesized speech (Klatt, 1980; McMurray, in prep)
- 4 continua x 3 items each x 9 steps x 2 vowel lengths

Item	Cue 1 *9 step continua*	Cue 2 *vowel length*	
Beach – Peach Bees – Peas Beak – Peak	Word initial VOT	Short: 100 ms Long: 300 ms	Predicted effect
Bell – Well Belt – Welt Bench - Wench	Slope of formant frequency	Short: 100 ms Long: 300 ms	
Lake – Rake Lace – Race Lei - Ray	Frequency of F3 onset	Short: 100 ms Long: 300 ms	No predicted effect / Control conditions
Dew – Goo Dune – Goon Duce – Goose	Frequency of F2 onset	Short: 100 ms Long: 300 ms	

Task

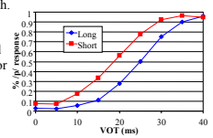
- Listeners hear stimulus and select the referent with a mouse.
- Eye movements recorded via head mounted eye tracker
- Analysis of mouse click responses revealed effect of step but not vowel for *control* continua



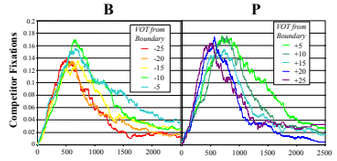


Voicing (B/P)

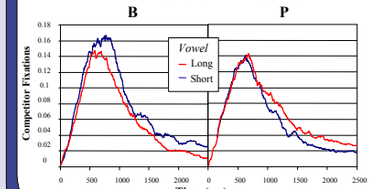
- Results from mouse-clicks yielded strong effects of both VOT and vowel length.
- Analysis of eye-movements examined looks to competitor for "correct" trials. VOT normalized to each subjects' category boundaries.



- Systematic effects of VOT on competitor looks for both /b/ and /p/ sides of continuum (B: p< .001; P: p=.002).

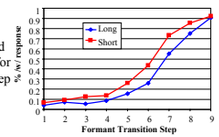


- Late effect of vowel length (B: p=.006, P: p=.061)

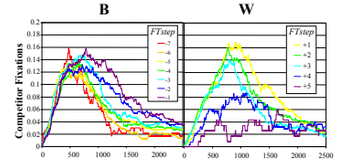


Manner (B/W)

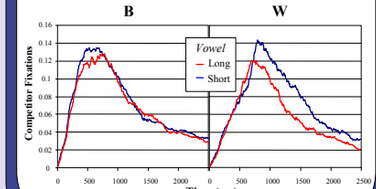
- Results from mouse-clicks yielded strong effects of both formant transition and vowel length.
- Analysis of eye-movements examined looks to competitor for "correct" trials. FTstep normalized to each subjects' category boundaries.



- Systematic effects of transition slope on competitor looks for both /b/ and /w/ sides of continuum (B: p< .001; W: p<.001).



- Late effect of vowel length (B: n.s. right direction, W: p<.001)



Temporal Integration

To test the temporal integration hypotheses, the timing of sensitivity to each factor was assessed.

For each subject, at *each time step* the following effect sizes were computed:

- **VOT**: effect of VOT on probability of competitor fixations (regression slope).
- **FTstep**: Regression slope as a function of FTstep
- **Vowel**: Difference between competitor fixations for short and long vowels.
- **Unrelated**: Difference between fixations to the target and competitor and the two unrelated items.

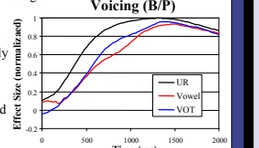
Predictions

- **Immediate cue integration**: VOT/FTstep effects appear before vowel effect.
- **Contextual cue integration**: VOT/FTstep effects occur at a similar time to vowel effect.
- **Sensitivity of measure:** If measure has enough temporal sensitivity, unrelated effect should appear before others.

Results B/P

- Eye-movements show early influence of unrelated/target distinction: our measurements are sensitive to fine grained temporal components of signal.
- Effects of VOT and vowel length appear later at approximately the *same time*.
- ***Contextual cue integration***: VOT and vowel length affect lexical activation at the same time, despite asynchrony in the input.

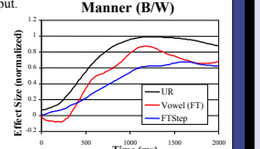
Voicing (B/P)



Results B/W

- Early influence of the unrelated/target distinction.
- Effects of transition step appears *prior to* vowel length.
- Vowel length is integrated with formant transition via ***immediate cue integration*** mechanisms.

Manner (B/W)



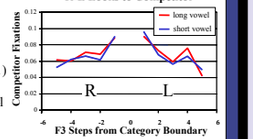
Effect onset (p<.05) in (ms)

	UR	VOT/FTStep	Vowel
B/P	320	< 600 =	600
B/W	400	< 580 <	700

Controls (L/R , D/G)

- Eye-movement analysis examined looks to competitor for "correct" trials. F3 steps are relative to subject's category boundary

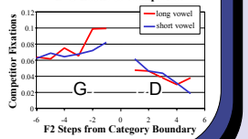
R-L Looks to Competitor



- Systematic within category effects of continuum step (R: p<.01, L: p<.001)
- No significant vowel effect or interaction

- Eye-movement analysis examined looks to competitor for "correct" trials. F2 steps are relative to subject's category boundary.

D-G Looks to Competitor



- Systematic within category effects of continuum step (D: p<.01, G: p<.01)
- No significant vowel effect or interaction

Conclusions

- VOT integrated with vowel length by **contextual cue-integration**—cues are processed together. Consistent with ratio-type models of voicing (e.g. Port & Dalby, 1982).
- Formant transition slope and vowel length are **integrated immediately**—cues affect lexical activation sequentially.
- Replication of gradient effects of subphonemic detail on lexical activation for VOT.
- Extension to new phonetic features: manner of articulation, place of articulation and laterality.
- Head-mounted eye-tracking reveals subjects' sensitivity to acoustic detail within phonetic categories and the time-course of integration across cues.

Acknowledgements

We would like to thank Dana Subik and Joe Toscano for their assistance in data collection and stimulus preparation, and Joanne Miller for suggestion of control items. This work was supported by NIH predoctoral NRSA (F31 DC006537-01) to BM, and NIH research grant (DC005071) to MKT and RNA.

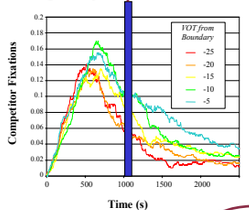
Temporal Integration Analyses

Goal: *Determine when effects of VOT / FTStep and Vowel Length can be seen in the eye-movement record.*

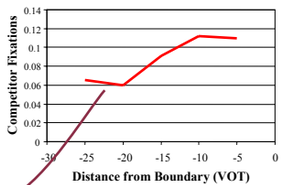
Method: *At each time step, for each subject the following effect sizes were computed:*

VOT Effect

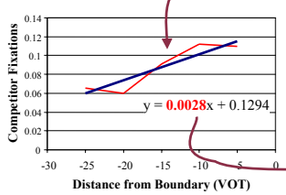
1) Select 20 ms time slice for a single subject.



2) Compute probability of fixating competitor at each VOT during this time.



3) Find regression line



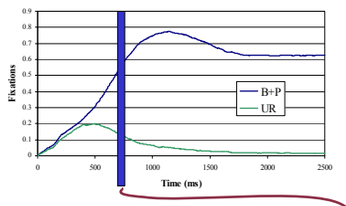
4) Add slope to dataset.

Subject	Time	VOT	Vowel	UR
1	700	.001		
1	720	.0035		
1	740	.0028		
1	760			
1	780			
1	800			

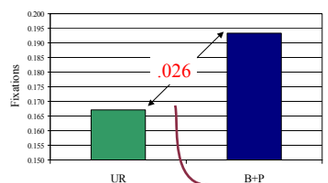
5) Repeat for next time slice/subject.

Unrelated Effect

1) Sum fixations to target and competitor (e.g. B+P) and to two unrelated items (UR 1 + UR 2).



2) Select time slice and compute difference between B+P and unrelated.

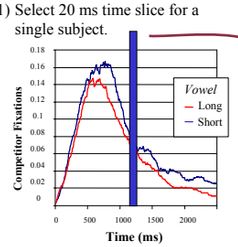


3) Add difference to dataset

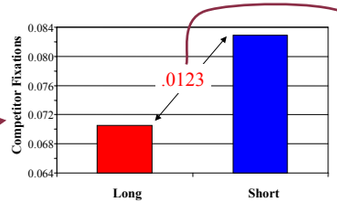
Subject	Time	VOT	Vowel	UR
1	700	.001	.005	.026
1	720	.0035	.015	
1	740	.0028	.0123	

Vowel Effect

1) Select 20 ms time slice for a single subject.



2) Compute difference between long and short.



3) Add difference to dataset

Subject	Time	VOT	Vowel	UR
1	700	.001	.005	
1	720	.0035	.015	
1	740	.0028	.0123	
1	760			
1	780			
1	800			

4) Repeat

All Analyses

Probability of fixating a lexical candidates *at each point in time* approximates temporal unfolding of lexical activation (Allopenna, Magnuson and Tanenhaus, 1998)

Relative effects of different signal components could be seen at different times. To verify that this measure has the required temporal sensitivity, we should minimally see more eye-movements to the target and competitor (b/p) than the unrelated (e.g. l/r/d/g) *early in time* (before any effects of VOT). That is, information in the signal is available to exclude control items from consideration *before* voicing can be determined.

Finding this effect (the UR effect preceding voicing or vowel length), we can then assess cue integration hypotheses.

For the computation of each effect size, the following apply:

- Exclude any eye movement initiated before 300 ms (100 ms of silence + 200 ms oculomotor delay).
- Exclude any subject x time data-point with fewer than 5 trials contributing eye-movements.
- Exclude any time-slice with fewer than 10 subjects.
- Effects on each side of continuum summed: VOT effect = slope for/b/ side + slope /p/ side.
- Repeated one-sample T-tests to determine when effects depart from zero.

References

- Allopenna, P., Magnuson, J.S. and Tanenhaus, M.K. (1998) Tracking the time course of spoken word recognition using eye-movements: evidence for continuous mapping models. *Journal of Memory and Language, 38(4)*, 419-439.
- Clark, M.J. and Hillenbrand, J.M. (2003) Quality Of American English Front Vowels Before /R/. *Journal of the International Phonetic Association, 33(1)*, 1-16.
- Dahan, D., Magnuson, J.S., Tanenhaus, M.K. and Hogan, E. (2001) Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16 (5/6)*, 507-534.
- Dahan, D. and Tanenhaus, M.K. (2004) Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30(2),* 493-513.
- Klatt, D. (1980) Software for a Cascade/Parallel Synthesizer. *Journal of the Acoustical Society of America, 67*, 971-995.
- Lahiri, A. and Marslen-Wilson, W. (1991) The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition, 38(3)*, 245-294.
- McMurray, B. (in preparation). KlattWorks: a [somewhat] new systematic approach to formant-based speech synthesis for empirical research.
- McMurray, B., Tanenhaus, M.K. and Aslin, R.N. (2002) Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86(2)*, B33-B42.
- McMurray, B., Tanenhaus, M., Aslin, R. and Spivey, M. (2003). Probabilistic constraint satisfaction at the lexical/phonetic interface: Evidence for gradient effects of within-category VOT on lexical access. *Journal of Psycholinguistic Research, 32(1)*, 77-97.
- Miller, J.L., Aibel, I.L. and Green, K.P. (1984) On the nature of rate-dependent processing during phonetic perception. *Perception & Psychophysics, 35(1)*, 5-15.
- Salverda, A.P., Dahan, D. and McQueen, J. (2003) The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90(1)*, 51-89.
- Summerfield, Q. (1981) Articulatory rate and perceptual constancy in phonetic perception. *Journal of the Acoustical Society of America, 7(5)*, 1074-1095.

Contact

Bob McMurray
Dept. of Brain and Cognitive Sciences
Melioral Hall
University of Rochester
Rochester, NY 14627

mcmurray@bcs.rochester.edu

147th Meeting of the Acoustical Society of America
New York, NY 5/24/04 – 5/28/04