

A lexical locus for the integration of asynchronous cues to voicing: An investigation with natural stimuli



Joseph C. Toscano and Bob McMurray
University of Iowa Dept. of Psychology

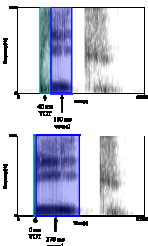


1pSC29

Experimental Design

- Previous studies: differences in identification functions to identify contribution of VL.
 - Problem: May not be sensitive to small VL effects.
- Solution: **Visual world paradigm** is more sensitive.
 - Sensitive to **sub-phonemic variation**. (McMurray et al. 2002)
 - Sensitive to **small effects**: could detect VL effects that are invisible to other methods.
 - Sensitive to **short-lived, temporal effects**.
 - Allows us to evaluate contribution of cues as they are heard: allows us to ask *how* cues are integrated.

- Stimuli**
 - Stimuli consisted of recorded speech in which we manipulated voice onset time (VOT) and vowel length (VL)
 - 7 minimal pairs:
 - Back/Peak
 - Beach/Peach
 - Be/Pet
 - Bike/Pike
 - Bath/Path
 - Back/Pack
 - 9 step VOT continua
 - 2 vowel lengths
 - 40% longer or shorter than original vowel length
 - 2 unrelated items per stimulus pair
- Participants**
 - 30 undergraduates at the University of Iowa participated in the experiment.



Multiple acoustic cues in speech

- Multiple cues contribute to phonetic categorization.
 - temporally asynchronous
 - vary in usefulness
 - different units (time vs. frequency).
- Cues must be **integrated** during phonetic categorization.
- e.g.: Perception of syllable-initial stop consonants depends on later-occurring context information, such as vowel length (VL). (Summerfield, 1981)

Context effects in natural and synthetic speech

- VL contributes to both voicing and manner categorization.
- VL effects found with certain stimuli, but not others:
 - VL effects:
 - Synthetic speech (Miller and Liberman, 1979; Summerfield, 1981)
 - More natural synthetic speech in multi-talker babble background noise (Miller and Wayland, 1993)
 - No (or reduced) VL effect:
 - Natural-sounding synthesized speech (Shin et al., 1985)
 - Natural speech (for voiceless stops) (Utman, 1998)
- This suggests that synthetic and natural speech may be processed differently from each other.

Issues

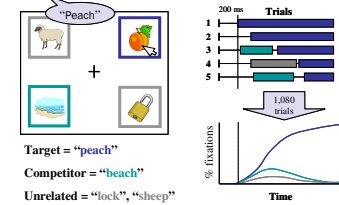
- Are synthetic and natural stimuli perceived differently?
- Can we see effects of multiple cues in both natural and synthetic speech?
 - Are multiple cues used in natural speech?
- Are cues integrated in the same way in synthetic and natural speech?
 - When multiple cues are used, **how** are they integrated?

Questions

- Under what circumstances are multiple cues used?
 - Synthetic vs. natural speech
 - Background noise
- Does eye-tracking reveal effects of cues that are not detectable in identification responses?
 - When multiple cues are used, how are they integrated?
- When multiple cues are used, how are they integrated?
- How are temporally asynchronous cues combined?
 - Integration at a pre-lexical (e.g. cue) level
 - Independent integration at the level of the lexicon

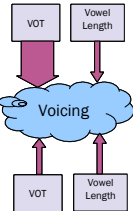
- Procedure
 - On each trial, listeners hear the target word and click on its picture with a mouse.
 - Eye movements are recorded via a head-mounted eye tracker (SR Eyelink II) at 250 Hz.
 - Eye-movements reveal unfolding of lexical activation during recognition.
- Why use the visual world paradigm
 - Natural task
 - Subjects are unaware of eye movements
 - Can be used without breaking up speech
 - High temporal sensitivity
 - Reflects activation of lexical items (Allopena et al., 1998)
 - Eye movements reflect which referents are considered during online word recognition. (Tanenhaus et al., 1995)

Visual World Paradigm



Cue Weighting

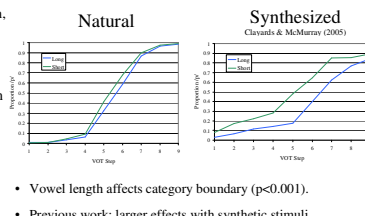
- Previous work: VL may not be used as a cue to voicing in natural speech.
- In contrast: Cues are weighted based on how useful they are.
 - VL effects seen when VOT is less useful.
- Relative weighting determines whether effects of multiple cues are observed.
- Natural speech
 - VOT is usually unambiguous
 - Strong cue to voicing
- Synthetic / noisy speech
 - VOT harder to compute accurately
 - Greater relative weight for vowel length



Predictions

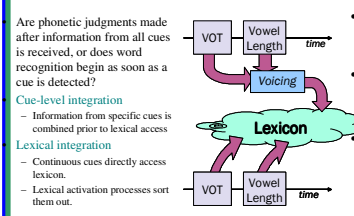
- Eye tracking will reveal influence of VL in natural speech, as differences in fixations to competitors (activation).
 - May be able to see VL effects across entire continuum.
 - Larger effects near the category boundary than at the endpoints.
- Prediction: VL cues are used to determine voicing when VOT is ambiguous.
- VL effects appear as increased looks to competitors when:
 - Short VL for /b/ words
 - Long VL for /p/ words

Identification Responses



- Vowel length affects category boundary (p<0.001).
- Previous work: larger effects with synthetic stimuli.

Locus of cue integration

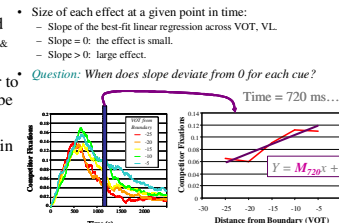


- Are phonetic judgments made after information from all cues is received, or does word recognition begin as soon as a cue is detected?
- Cue-level integration
 - Information from specific cues is combined prior to lexical access
- Lexical integration
 - Continuous cues directly access lexicon.
 - Lexical activation processes sort them out.

Predictions

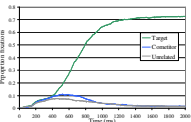
- With synthesized stimuli, cues are integrated independently in time. (McMurray et al., 2004; Claydys & McMurray, 2005)
- If cue integration in natural speech is similar to integration in synthesized speech, cues will be used as they become available.
- Prediction: Effects of VOT observed **earlier** in processing than VL effects.

Measuring timing of cue use



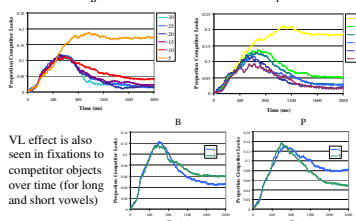
Eye movement results

- Expected pattern of eye movements; similar to previous visual world studies.



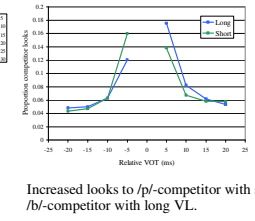
Fixations to target, competitor and unrelated items as a function of time for stimuli with a 0 ms VOT.

Gradient effects of VOT are seen in fixations to competitor objects over time as a function of distance from category boundary.



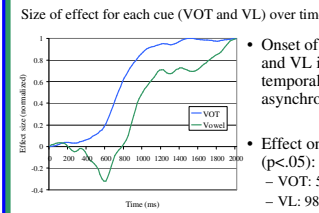
VL effect is also seen in fixations to competitor objects over time (for long and short vowels)

Proportion of looks to competitor



- Target /b/:
 - VOT: p<.001
 - Vowel: p=.059
 - VOT*Vowel: p<.001
- Target /p/:
 - VOT: p<.001
 - Vowel: p=.034
 - VOT*Vowel: p=.018
- Increased looks to /p/-competitor with short VL, and to /b/-competitor with long VL.

Onset of effects



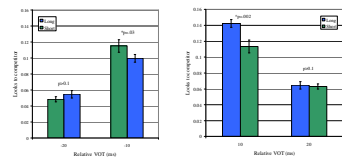
- Size of effect for each cue (VOT and VL) over time
- Onset of VOT and VL is temporally asynchronous
- Effect onset (p<.05):
 - VOT: 520 ms
 - VL: 980 ms

Conclusions

- VOT and VL effects observed as each cue is available.
 - Similar to results obtained with synthesized speech.
- Immediacy: Listeners do not wait for all cue information (for a single feature) to become available before accessing the lexicon.
- This suggests a **lexical locus for the integration of multiple acoustic cues**, rather than a pre-lexical locus of integration.

Effect of VL near category boundary

Significant vowel effect near category boundary (where VOT is ambiguous), but not at endpoints (where VOT is more reliable).



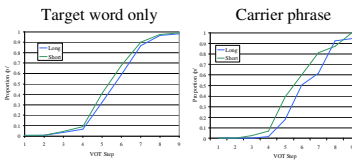
Conclusions

- VL is a cue to voicing in natural speech.
 - VL effects seen near the VOT category boundary.
 - Cues that are normally less reliable (VL) are used when more reliable cues (VOT) are ambiguous.
- Cues may be re-weighted online, as information from different cues is processed.
- When VOT is unambiguous →
 - Vowel length is weighted low
- When VOT is ambiguous →
 - Vowel length is weighted higher
 - Contributes more to voicing decision

Work in progress: Effect of carrier phrase

- Cue weighting may differ in running speech.
 - Insufficient time to compute cues: speed/accuracy tradeoff.
 - Rate compensation more complex: increases importance of temporal cues.
- Cue Weighting Hypothesis: VL effects when the target word is preceded by a carrier phrase.
- Same stimuli spliced onto a series of carrier phrases that instructed the subject to perform a particular task
 - e.g. Please pick the pointer and click on the peach.
- Prediction: larger VL effects with carrier sentence.

Responses with Carrier Phrases



- Carrier phrase creates larger VL effect.
- Suggests that multiple cues are used in more natural contexts when individual cues may be less reliable than in isolation.

General Conclusions

- Gradient lexical activation in natural speech.
 - Cues weighted by their utility.
 - Demands of running speech may change utility.
- Cues are used as they become available, suggesting a lexical locus for cue integration.
- Similar results for natural and synthetic speech, suggesting that the two are not perceived differently.

References

Allopena, P.D., Morrison, J.L., and Tanenhaus, M.K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous updating. *Journal of Memory and Language*, 39(1), 413-430.

Claydys, M., and McMurray, B. (2005). *Visual World Paradigm*. In *Proceedings of the 15th Meeting of the Midwestern Psychological Association*, Ann Arbor, Michigan.

McMurray, B., Tanenhaus, M., and Adin, B. (2002). Gradient effects of within-category variation on lexical access. *Cognition*, 82(1), 83-104.

McMurray, B., Tanenhaus, M., Adin, B., and Papanicolaou, M.P. (2004). High-resolution measures to investigate the timing of phonetic cues. *Proceedings of the 15th Meeting of the Acoustical Society of America*, New York.

Miller, J.L., and Liberman, M.C. (1979). Some effects of pre-processing information on the perception of stop consonants and consonant pairs. *Perception*, 8(2), 41-46.

Miller, J.L., and Liberman, M.C. (1979). Limits on the combination of context-combination effects in the perception of [d] and [t]. *Perception*, 8(2), 201-208.

Shin, P.C., Liberman, M.C., and Liberman, A. Limitations of context-combination effects in the perception of [d] and [t]. *Perception*, 8(2), 209-217.

Summerfield, Q. (1981). Articulatory rate and perceptual continuity in phonetic perception. *Journal of Experimental Psychology: General*, 110(1), 103-119.

Tanenhaus, M.K., Sperry-Kawashita, M.J., Ehrhardt, R.M., and Saddy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632-1634.

Utman, J.A. (1998). Effects of local speaking rate context on the perception of voice-onset time in natural stop consonants. *Journal of the Acoustical Society of America*, 103(5), 2648-2651.

Acknowledgments

We would like to thank Meghan Claydys for theoretical discussions of these issues, Molly Robinson for assistance with stimulus preparation, and the students in the MACLab at the University of Iowa for helping to run subjects in the experiment.