Eye movements reflect acoustic cue informativity and statistical noise

Jessie S. Nixon[1, 2, 3], Jacolien van Rij[4], Peggy Mok[5], Harald Baayen[4], Yiya Chen[1,2]
[1]Leiden Institute for Brain and Cognition (LIBC), Leiden University, Netherlands
[2]Leiden University Centre for Linguistics, Leiden University, Netherlands
[3]MARCS Institute, University of Western Sydney, Australia
[4]Department of Linguistics, University of Tübingen, Germany
[5]Department of Linguistics, Chinese University of Hong Kong, Hong Kong
https://doi.org/10.36505/ExLing-2015/06/0013/000250

Abstract

Listeners rely on highly variable, non-discrete acoustic information to understand spoken messages. The present 'visual world' eye tracking study investigated whether the *amount of acoustic cue variation* affected Cantonese listeners' perception of speech contrasts. Participants saw pictures of word pairs which were identical except for initial consonants (unaspirated versus aspirated). Auditory stimuli were continua of increasing VOT presented in bimodal distributions. The amount of acoustic variation varied between conditions: high-variance versus low-variance. *Generalised Additive Modelling* analyses showed, in the low-variance condition, eye movements reflected cue values: there was differential fixation behaviour for category means, boundaries and peripheries. In contrast, in the high-variance condition, the acoustic cue had little effect: fixation behaviour was similar across the different acoustic cue values. This demonstrates listeners' high sensitivity to the discriminative value of acoustic cues. How much cue dimensions are utilised depends on their variance.

Key words: speech perception, acoustic variation, word discrimination, Cantonese

Introduction

The acoustic information that listeners use to discriminate between possible spoken messages is highly variable and non-discrete. Speech consists of a complex collection of cues, which listeners can potentially use for discrimination. Which factors affect whether and the extent to which listeners use a particular cue? There is substantial evidence that infants and adults detect and respond to the *number of peaks* in a distribution, i.e. unimodal versus bimodal distributions (e.g. Maye et al.,2002). However, little is known about how *acoustic variation* affects speech perception. In one innovative recent study to address this question, Clayards et al. (2008), showed that increased within-category variability led to a greater proportion of looks to competitor objects in English stop discrimination. However, only cue values near distribution peaks were analysed and fixation proportions were averaged over trials, so there was no analysis of the time course.

The present study investigates how acoustic cue variability affects native Cantonese listeners' discrimination during speech perception. This informs the

question of how listeners are able to utilise informative cues and ignore irrelevant ones. Importantly, the present study also investigates the *time course* of effects. We analysed fixations over the course of the trial. In addition, we also analysed the whole *cue continuum*, to examine the effects on cue values near the category boundary and perimeters. We use the term *variance* to describe the amount of *within-category* acoustic variation. This refers to the degree to which acoustic values spread out from the category mean. A variance of zero means that all values are identical.

Method

Participants

Thirty-seven native Cantonese-speaking students from the Chinese University of Hong Kong participated in the experiment for payment.

Experiment design and stimuli

The experiment design and stimuli were based on Clayards, et al. (2008). Visual stimuli were picture pairs, identical except for initial consonants, which were *unaspirated* (bou3, `cloth'; jun1 `brick') or *aspirated* (pou3, `shop'; chun1, `village'). Auditory stimuli were recorded by a native Cantonese speaker and resynthesised into a 12-step VOT continuum. All participants heard a bimodal distribution of auditory stimuli. Only token presentation frequency varied between conditions: high-variance vs. low-variance.

Procedure

Participants wore an SR Eyelink II eye-tracker. The session started with familiarisation, then a practice. Trials consisted of a brief (1000 ms) preview of four pictures, followed by a gaze-contingent fixation cross, then auditory stimuli were played simultaneously as the pictures reappeared. Eye movements were monitored until participants clicked on the picture they heard.

Analysis and results

Eye movement data were analysed using *Generalized Additive Modelling* (GAM; Wood, 2006, 2011) using the *mgcv* package in R. GAMs area type of regression analysis that drop the linearity assumption. The degree of non-linearity is determined from the data itself: the optimal linear or non-linear equation to avoid model over-fitting and over-generalizing (Wood, 2006).

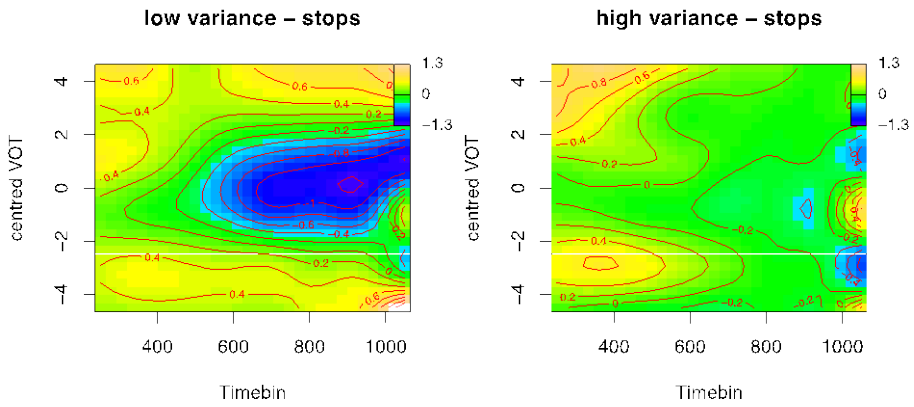

Figure 1. Topographical maps of the proportion of fixations on the clicked target versus competitor for VOT over time in the GAM model for stops in the low-variance (left panel) and high-variance conditions (right panel). Estimated effects are on logit scale. Time (x-axis) is in bins of 100 ms. VOT (y-axis) is centred around 0, the category boundary. Negative VOT values correspond to unaspirated stimuli, positive values to aspirated stimuli. Category means are at VOT -2.5 and 2.5. Fixation proportions (z-axis) are represented by colour codes. Positive values (yellow) indicate relatively more looks to the clicked target; negative values (blue) indicate relatively more looks to the competitor. Random effects are excluded from these plots.

Model comparisons, model summaries and model plots all provide evidence for an effect of distribution condition. A $\chi 2$ test of fREML scores which takes into account model complexity showed thataVOT by condition interaction over time significantly improved model fit ($\chi 2(3)=135.32$, $p<.001$). Since this interaction was significant, lower-level predictors were retained in the model. The model also included manner of articulation (stops versus affricates),but due to space limitations, we will focus on our main predictor of interest, distribution condition. The model summary showed that the variance explained by surfaces ofVOT over time ($\chi 2(32.95)=3637.4$, $p<.001$) and VOT by condition over time ($\chi 2(28.27)=3533.11$, $p<.001$) were significantly different to zero.

The model plots (Figure 1)show differential patterns of looking behaviour between the low-variance (left panel) and high-variance conditions (right panel). The low-variance condition displays divergent patterns between category means, category boundaries and the distribution peripheries. The high-variance condition, in contrast, is quite flat across VOT values after about 500 ms after presentation of the auditory stimulus.

Discussion

The present study demonstrates that subtle differences in acoustic cue variance can have immediate effects on the way a particular cue is perceived. The GAM analysis revealed that fixations on the clicked target over time were affected by VOT value. More interestingly, the effect of VOT significantly interacted with distribution condition. From about 500 ms into the trial, distinct eye movement patterns emerged for different VOT values in the low-variance condition (left panel, Figure 1), with differential fixation behaviour at category means, boundaries and distribution peripheries. In contrast, in the high-variance condition (right panel, Figure 1), VOT had a weaker effect: fixation behaviour was similar for all VOT values.The eye movement patterns seem to reflect different stages of processing: an initial, perceptual stage and a later process of verification. This suggests that, at least at later stages of processing, the acoustic cue is relied on less for discrimination in the high-variance condition, when it is less informative, than in the low-variance condition when it is more informative.

In summary, the present results demonstrate that the discriminative value of acoustic cues has an immediate effect on how they are processed in speech perception. When cues more consistently fall within a small range of values, they are more reliable as discriminators, and consequently used more effectively for discrimination. However, when cues are highly variable over a range of values, their effectiveness for discrimination declines, and the degree to which they are utilised in perception decreases in turn. Interestingly, the analysis of the time course shows that this is a relatively late effect. This may suggest that, following initial perceptual processes, there is increased difficulty at a later verification stage.

Acknowledgements

JSN was responsible for the experiment design, implementation, stimuli creation, data collection and manuscript preparation. JSN and JvR conducted the analysis and interpretation with significant contributions from HB. PM assisted with finding equipment and participants. This work was supported by European Research Council ERC Starting [Grant 206198] to YC.

References

Clayards, M., Tanenhaus, M., Aslin, R., Jacobs, R. A., 2008. Perception of speech reflects optimal use of probabilistic speech cues.Cognition 108, 804–809.

Maye, J., Werker, J.F., Gerken, L., 2002. Infant sensitivity to distributional information can affect phonetic discrimination. Cognition 82 (3).

Wood, S. 2006. Generalized additive models: an introduction with R. CRC press.