Eye movements reveal Cantonese listeners use statistical information to assess category membership of acoustic cues

Jessie Nixon, Jacolien van Rij, Peggy Mok, Harald Baayen, Yiya Chen

Introduction Human listeners rely on highly variable, non-discrete acoustic information to extract a speaker's intended message from the speech signal. Any given cue provides only probabilistic information about the intended message. The present study investigated whether the amount of statistical noise (acoustic cue variability) affected Cantonese native listeners' perception of speech contrasts.

Methodology In a visual world eyetracking paradigm, participants saw four pictures on screen and heard an auditory stimulus. Critical pictures were of word pairs that were identical except for initial consonants, which varied in aspiration (Experiment 1) e.g. unaspirated (寶, bou2, 'treasure') versus aspirated (譜, pou2 'shop') or tone (Experiment 2) e.g. high (冠, gun1, 'crown') versus mid tone (罐, gun3, 'can'). Auditory stimuli (Table 1) consisted of a 12-step continuum of increasing VOT (Exp 1) or pitch (Exp 2). In both conditions, the number of times participants heard each token followed a bimodal distribution. Only width of distribution varied between conditions (between-participants): wide (high variability) versus narrow (low variability).

Results Analysis was conducted using *generalised additive mixed modelling* (Wood, 2006), a type of generalised linear model that uses non-linear smooth functions to model linear predictors. The proportion of fixations on the clicked target over the trial depended on variant (i.e.VOT or pitch) value. There were relatively many fixations on the clicked object compared to the competitor (yellow areas) in the early part of the trial period, particularly for prototypical acoustic cue values (around vot.c -2.5 and 2.5), compared to values near the category boundary (up to 250 ms, both panels, Fig 1).

More importantly, the effect of variant was modulated by distribution condition. A clear shape of the distribution emerged in the narrow (low-variability) condition (left panel, Fig 1), with differential looking behaviour at category means (vot.c -2.5 and 2.5), boundaries and peripheries. In contrast, in the wide (high-variability) condition (right panel, Fig 1), the pattern of eye movements was flatter (green area); in the latter part of the trial, there was a weaker effect of variant, such that after 600 ms the distribution appears quite flat across all variant values.

Discussion The present results show that subtle differences in acoustic cue distributions can affect the way a particular acoustic cue is perceived. The low noise (i.e. high cue predictiveness) in the narrow condition led to greater overall certainty for cues within the expected range, compared to the wide condition. The results suggest that at later stages of processing, acoustic cues are relied on less for verification of the decision when VOT/pitch is less informative (wide condition), than when it is more informative (narrow condition). Interestingly, there seems to also be a trade-off of this relative certainty in the narrow condition: cues near the category boundary incurred a greater cost of rejecting the competitor, compared to the wide condition.

Table 1: Presentation frequency per variant per condition: each variant is one step on continuum

		Number of iterations											
	Variant	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Distribution	**Narrow**	0	6	54	108	54	6	6	54	108	54	6	0
condition	**Wide**	6	24	54	60	54	30	30	54	60	54	54	6

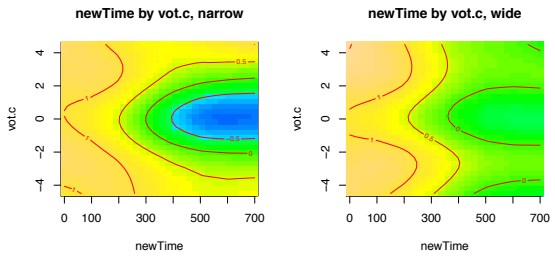

Figure 1: Fixation proportions for VOT over time for the narrow and wide conditions Exp 1. Yellow indicates relatively more looks to clicked target; blue more looks to competitor. VOT is on the x-axis. Category means are at vot.c -2.5 (unaspirated) and 2.5 (aspirated).

References

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