Journal of Memory and Language 90 (2016) 103–125

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

The temporal dynamics of perceptual uncertainty: eye movement evidence from Cantonese segment and tone perception

Jessie S. Nixon [a,*], Jacolien van Rij [b], Peggy Mok [c], R. Harald Baayen [b], Yiya Chen [d]

[a] Department of Pyschology, University of Potsdam, Potsdam, Germany
[b] Quantitative Linguistics Group, University of Tubingen, Germany
[c] Chinese University of Hong Kong, Hong Kong
[d] Leiden University Centre for Linguistics, Leiden, The Netherlands

ARTICLE INFO

Article history:
Received 10 September 2014
revision received 15 March 2016

Keywords:
Discriminative learning
Statistical learning
Speech perception
Cantonese
Lexical tone

ABSTRACT

Two visual world eyetracking experiments investigated how acoustic cue value and statistical variance affect perceptual uncertainty during Cantonese consonant (Experiment 1) and tone perception (Experiment 2). Participants heard low- or high-variance acoustic stimuli. Euclidean distance of fixations from target and competitor pictures over time was analysed using Generalised Additive Mixed Modelling. Distance of fixations from target and competitor pictures varied as a function of acoustic cue, providing evidence for gradient, nonlinear sensitivity to cue values. Moreover, cue value effects significantly interacted with statistical variance, indicating that the cue distribution directly affects perceptual uncertainty. Interestingly, the time course of effects differed between target distance and competitor distance models. The pattern of effects over time suggests a global strategy in response to the level of uncertainty: as uncertainty increases, verification looks increase accordingly. Low variance generally creates less uncertainty, but can lead to greater uncertainty in the face of unexpected speech tokens.

Introduction

Human listeners rely on highly variable, non-discrete acoustic information to discriminate between the different possible messages a speaker might intend to convey in an utterance. The question of how acoustic variation affects perceptual uncertainty during speech processing is an intriguing one. Listeners use variation *between* speech sounds to discriminate between words and messages. For example, in English, voice onset time (VOT) is longer in voiceless sounds (e.g. the /p/ in *pat*) than voiced sounds (e.g. the /b/ in *bat*). VOT is the time between the release burst of the consonant and the onset of voicing in the vowel, and is the most important cue for distinguishing voiced from voiceless sounds in English. However, there is also a considerable amount of variation *within* speech categories. For example, the mean VOT of English /p/ is 58 ms (Lisker & Abramson, 1964), but /p/ can be produced with a range of VOTs. Acoustic variation can even occur in productions of the same word by the same speaker in the same phonetic context under controlled lab settings (Newman, Clouse, & Burnham, 2001) and increases greatly across speakers (Ladefoged & Broadbent, 1957), in different phonetic contexts (Nixon, Chen, & Schiller, 2015) and even depending on word frequency (Gahl, 2008).

The high degree of variation in the acoustic signal means that there is nothing in the speech stream that conclusively points to particular meanings, words or even phonemes. The listener can only use cues to assess the likelihood that a speaker intended one message rather than another,

* Corresponding author.

http://dx.doi.org/10.1016/j.jml.2016.03.005
0749-596X/

meaning that there is always some degree of uncertainty in the process of speech perception. In addition to the issue of within-category acoustic variation, listeners also face the challenge of changes in the whole statistical distribution of acoustic cues in particular contexts, for example, when encountering a new speaker or accent. Recent evidence suggests that both variation in acoustic cues (McMurray, Aslin, Tanenhaus, Spivey, & Subik, 2008; McMurray, Tanenhaus, & Aslin, 2002, 2009) and changes in the statistics of cue distributions affect listeners' level of perceptual uncertainty during speech perception (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Escudero, Benders, & Wanrooij, 2011; Escudero & Williams, 2014; Liu & Kager, 2011; Wanrooij, Boersma, & van Zuijen, 2014; Wanrooij, Escudero, & Raijmakers, 2013). The present study aims to contribute to our understanding of perceptual uncertainty in speech perception by examining the time course of effects of (a) variation in acoustic cues and (b) the degree of variance in statistical distributions of acoustic cues in native Cantonese listeners. In this paper, we use the term *variance* to describe, in a given speech sample, the amount of acoustic variation there is *within* a speech category. This term refers to the degree to which acoustic values spread out from the mean of the distribution of that speech category. A variance of zero means that all values are identical.

Early accounts claimed that speech perception was 'categorical' in that listeners were unable to detect within-category acoustic variation, and only able to detect variation when it occurred across boundaries. Evidence in favour of this claim came from studies showing sharp categorisation functions between speech categories, and chance-level performance in detecting within-category acoustic differences (e.g. Ferrero, Pelamatti, & Vagges, 1982; Liberman, Harris, Hoffman, & Griffith, 1957; Schouten & van Hessen, 1992). However, more recently, abundant evidence has accumulated demonstrating listeners' remarkable sensitivity to fine-grained phonetic information, given the appropriate task (e.g. Andruski et al., 1994; Dahan et al., 2001; Marslen-Wilson and Warren, 1994; McMurray, Aslin, et al., 2008; McMurray et al., 2002, 2009; Utman et al., 2000).

Moreover, not only are listeners sensitive to gradient acoustic variation, they are able to rapidly adapt to context-specific changes in acoustic characteristics of speech, based on the effectiveness of a particular dimension for speech recognition (Idemaru & Holt, 2011, 2014). Relatedly, listeners are also sensitive to *frequency* distributions of acoustic cues. One line of research has investigated how the acoustic distance between speech categories affects categorisation accuracy. For example, several studies have shown that when trained with a unimodal distribution (no distance between categories), participants are less likely to categorise the endpoints of a distribution as different, compared to when they are trained with a bimodal distribution (Escudero & Williams, 2014; Liu & Kager, 2011; Maye & Gerken, 2000; Maye, Weiss, & Aslin, 2008; Maye, Werker, & Gerken, 2002; Wanrooij et al., 2014). Even when trained with a bimodal distribution, a greater distance between categories improves categorisation accuracy, compared to training with a bimodal distribution with a small distance between categories (Escudero et al., 2011; Wanrooij et al., 2013).

Much of the research in adult distributional learning has focused on the acquisition and development of non-native contrasts. For example, a series of recent studies has investigated the effects of statistical distributions on non-native perception of Dutch vowel contrasts (Escudero et al., 2011; Gulian, Escudero, & Boersma, 2007; Wanrooij et al., 2013). Motivated by the observation that infant and foreigner directed speech has a 'stretched' vowel space, Escudero et al. (2011) investigated effects of the acoustic interval between vowel categories in second language acquisition. They used *natural bimodal* (reduced acoustic interval; i.e. vowel categories were similar to each other) versus *enhanced bimodal* distributions (increased acoustic interval) to train Spanish learners to distinguish a Dutch vowel contrast. After two minutes of exposure natural bimodal or enhanced distributions, there was an increase in 'correct' categorisation, compared to the music (control) group. This increase only reached significance in the enhanced group.

Most studies of distributional learning in adults have used offline categorisation responses as the measure of learning. Categorisation measures provide information about the final outcome of the decision process; however, they do not provide information about online processing during perception itself. In discussions of effects on categorisation, it is often implicitly or explicitly assumed that assigning tokens to one category rather than two occurs because the two tokens were not discriminated. This assumption may not necessarily be justified. In a forced-choice categorisation task, regardless of the degree of uncertainty, or any gradient degree of goodness of fit with one category or another, the participant must make a binary choice. While it is interesting that factors such as cue distribution can affect even the final outcome of the decision process, examining the moment by moment online processing can tell us about how subtle differences in statistical distributions can affect the development of perceptual processes over time, prior to the decision process.

One interesting and innovative recent eyetracking study (Clayards et al., 2008) is, to the best of our knowledge, the only other study that has used online measures to investigate statistical processing of acoustic cues during perception of native speech contrasts. This study has examined how the *amount* of within-category acoustic variation affects perceptual certainty. Using the visual world paradigm (VWP; Allopenna, Magnuson, & Tanenhaus, 1998), Clayards et al. (2008) tested the hypothesis that greater variation in the acoustic signal would lead to greater perceptual uncertainty. Native English-speaking participants saw four pictures on screen, heard an auditory stimulus and were instructed to click on the picture of the word they heard. Critical picture stimuli consisted of pairs of words beginning with /b/ and /p/ (e.g. 'beach' and 'peach'). Auditory stimuli consisted of a VOT continuum which spanned the word pair (e.g. from beach to peach). Presentation frequency of the tokens on the continuum always followed a bimodal distribution. However, the amount of within-category acoustic variation was manipulated between participants: participants heard either a high-variance or low-variance distribution of the acoustic stimuli.

In the analysis, the proportion of categorisation responses was calculated per participant per condition