UNIVERSITY OF CAMBRIDGE

Centre for the Neural Basis of Hearing

# Is speaker-size estimation automatic and obligatory in streaming ?

*Etienne Gaudrain, Alessandro Binetti, Roy D. Patterson*

Centre for the Neural Basis of Hearing,
Department of Physiology, Development and Neuroscience,
University of Cambridge

## 1. Introduction

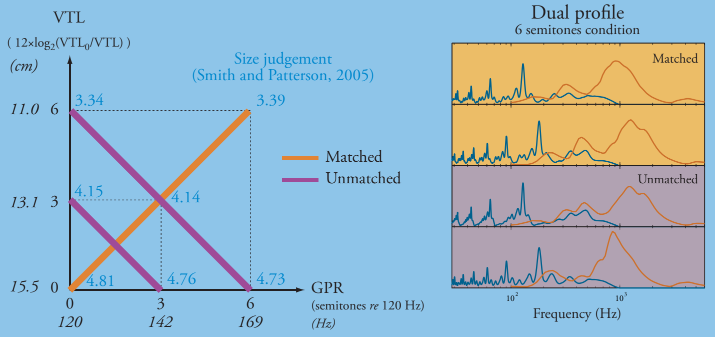
The voice of a speaker contains information that helps to identify the speaker and to segregate their voice from those of others in a multi-speaker environment. Specifically, there is information about the size of the speakers vocal folds in their mean glottal pulse rate (GPR) and information about their vocal-tract length (VTL) in the formant frequencies of their vowels. It seems likely that this size information is used to identify and track a target individual in a multi-speaker environment.

Darwin *et al.* (2003) have reported that GPR and VTL reinforced each other in concurrent sentence reception and raise performance above where it might be expected to be on the basis of either of these components on its own. This observation suggests that for normal combinations of GPR and VTL values, these two factors interact to form a reliable speaker-size estimate that is used to segregate concurrent speakers. The question raised in the current study is whether the mechanism is primitive, or whether it involves higher-level cognitive processing.

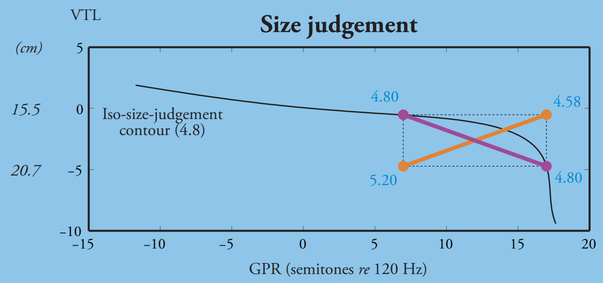
The technique that Darwin *et al.* used does not enable us to answer this question. There is, however, alternative techniques involving obligatory streaming, *i.e.* streaming that cannot be suppressed. This obligatory streaming has been used to observe the effect of GPR on segregation (Gaudrain *et al.*, 2007), and separately, the effect of VTL on segregation (Tsuzaki *et al.*, 2007). In the current study, we  aim to measure the cumulative effect of GPR and VTL on obligatory streaming to determine whether the size estimate used for concurrent voice segregation is a primitive component of perception or whether it requires higher-level processing.

## 2. Acoustic scale and Size judgement

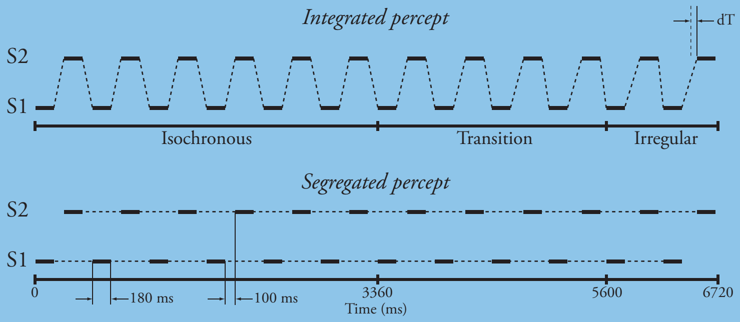
### Acoustic Scale



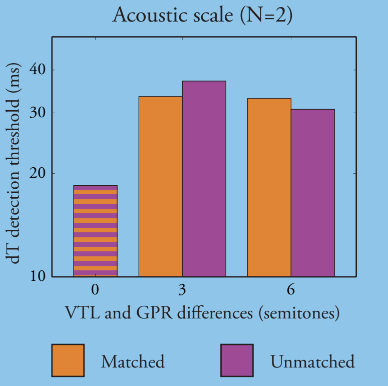
### Size judgement



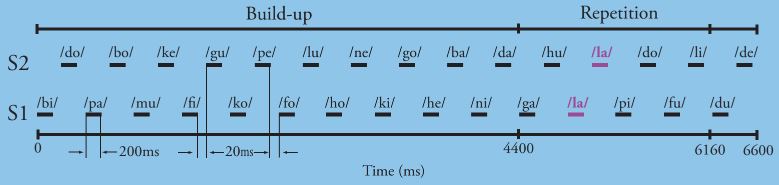
## 3. Delay detection paradigm



This method was described by Roberts, *et al.* (2002). Here, sequences are 24 syllables, randomly chosen from a database of 50 syllables (5 vowels: /a, e, i, o, u/, 10 consonnants : /b, d, f, g, h, k, l, m, n, p/). The stimuli are presented in a 3 down-1 up, 2I2AFC procedure to determine the detection threshold of the delay dT (79%-correct on the psychometric function). More streaming yields larger thresholds.

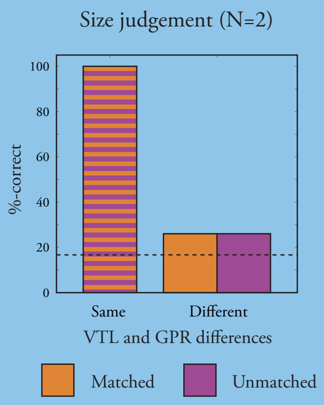


## 4. Repeated syllable paradigm



A sequence of 30 randomly chosen syllables contains only one repetition, across the two speakers. The repetition can occur in any item between the 22nd and the 27th. The subject is asked to provide the repeated syllable after the end of the sequence. They can only do so if the sequence is perceived as sufficiently integrated. In the Size-judgement condition, the Matched condition corresponds to a difference in perceived size, while the Unmatched condition corresponds to no difference in perceived size because the points in the GPR-VTL plan are taken from the iso-size-judgement contour.

The preliminary results show no difference between the Matched and Unmatched conditions neither for changes based on the Acoustic scale, nor for changes based on the Size-judgement. These results support the idea that VTL and GPR are treated independently at the obligatory streaming level.



BSA Short Papers Meeting, University of York, September 2008

Darwin C.J., Brungart D.S. and Simspon B.D. (**2003**). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *J Acoust Soc Am*, **114**, 2913-22.

Gaudrain E., Grimault N., Healy E.W. and Béra J.-C. (**2007**). Effect of spectral smearing on the perceptual segregation of vowel sequences. *Hear Res*, **231**, 32-41.

Roberts, B., Glasberg, B.R. and Moore, B.C.J. (**2002**). Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *J Acoust Soc Am*, **112**, 2074-85.

Tsuzaki M., Takeshima C., Irino T. and Patterson R.D. (**2007**). Auditory stream segregation based on speaker size, and identification of size-modulated vowel sequences. *In:* Kollmeier B., Klump G., Hohmann V., Langemann U., Mauermann M., Uppenkamp S. and Verhey J. (eds.) *Hearing – From Sensory Processing to Perception*. Springer. p. 285.

This work was supported by
UK Medical Research Council
grant G9900369.

Contact: epg22@cam.ac.uk