

Speaker size information: properties and processing – Roy Patterson

Speech sounds contain information about speaker size inasmuch as speech sounds contain detailed information about vocal tract length (VTL), and VTL is highly correlated with speaker size. The size information is available locally in every voiced syllable and, unlike pitch, size does not vary from moment to moment or from day to day. We believe that size is a cognitive parameter that the brain evaluates and uses as part of its model of a speaker. We also believe that there is probably a size processing centre in auditory cortex that extract the information prior to speech specific and language specific processing. We have been developing tools, analyses and experiments in a range of domains to provide evidence for the importance of size in speech perception. Some of the projects are listed below as they pertain to two of the three key areas identified in the “Grand Challenge”:

a. multi-modal imaging and neurocognitive dynamics

With four groups in Britain and Germany, we have used PET, fMRI and MEG imaging to locate, a pitch processing centre in auditory cortex (see papers on our web page). It is on Heschl’s gyrus next to Primary Auditory Cortex (PAC) and it is probably the centre that extracts voice pitch.

Size information is like pitch information inasmuch as it is contained within the vowel and so we would expect to find the size processing centre near PAC in auditory cortex. Accordingly, we have begun fMRI and MEG experiments with several groups to search for a size processing centre in auditory cortex.

We are also participating in fMRI experiments with the CBU that appear to show that speech specific processing (phonological processing) starts, not in auditory cortex but in regions adjacent to it.

c. computational modelling of human language function.

We have developed a simple physical model of formant frequencies and vocal tract length to explain the traditional data of Peterson and Barney and relate it to the fMRI measurements of VTL provided recently by Fitch. Size is the second most important variable in formant data after vowel type. We have developed tools to extract the size information and to summarize it in a single parameter, VTL (in cm). There would appear to be several important uses for a reliable measure of VTL above and beyond vowel normalization. For example, it should be a consistent feature for speaker identification, and a stable parameter for tracking and segregating speakers in noise and reverberant rooms.

We are performing a series of experiments on the perception of size in speech and music. With regard to speech: We have prepared data sets with controlled variation of pitch and VTL using a high-quality vocoder (STRAIGHT). The stimuli are then used to perform psychophysical experiments where the stimuli have natural speech variation of pitch and level, but which are controlled to avoid confounding pitch information about size with

VTL information about size. The experiments reveal that listeners are very good at detecting changes in VTL and that VTL dominates over pitch in the size judgement (contrary to common belief).