The forensic implications of comparing soft, neutral and loud speech: A quantitative analysis of inter- vs intra- speaker vowel acoustics and speech rate in Italian.

Massimiliano Canzi, The University of Manchester
`massimiliano.canzi@postgrad.manchester.ac.uk`

The current research focuses on the variation of acoustic and phonetic parameters including mean F0, F1, F2, speech rate and vowel duration in soft, neutral and loud speech throughout the whole Italian vowel monophthong inventory. Considerable intra–speaker variation from soft to loud speech is expected. Furthermore, we expect considerable inter-speaker variation to be conditioned by speech mode. Both inter- vs –intra-speaker differences are considered extremely relevant in the forensic context, as they are known to play a role in speaker identification and recognition (Remez et al. 1997). The study presents a bigger picture of the effect of speech-mode change throughout a whole vowel inventory, compared to existing studies on forensic phonetic analysis of loud speech which mainly focused on one or few vowels (e.g. Elliott et al. 2000).

8 speakers with L1 Italian (5 males) read a word list of 168 items and a sentence list in each of the three speech modes. Results of linear mixed effects model regressions (lmerTest; Kuznetsova et al. 2015) show significant difference (20 Hz) for mean F0 ($t(7) = 7.53$, $p < 0.001$) and F2 from neutral to loud speech only ($t(7) = 3.46$, $p < 0.01$). No stable trend can be discerned for a change in F2, in contrast, F1 is found to increase of 25.5 Hz from soft to neutral ($t(7) = 3.45$, $p < 0.004$) and of 41.1 Hz from neutral to loud speech ($t(7) = 7.77$, $p < 0.001$). Speech rate, measured in syllables per second, decreases around 10% from neutral to loud speech (8 SpS in neutral speech, 7.1 in loud) but it does not show significant difference between soft and neutral speech. Vowel length was found to increase between neutral and loud speech ($t(7) = 8.41$, $p < 0.0001$), but it remains unchanged between soft and neutral. There was no significant interaction between phonological vowel length and speech mode in conditioning pitch and formant increase.

Differences in production and acoustics between speech modes, from soft to loud speech in this particular case, have both been attributed to production demands and perceptual constraints (Schulman 1989). Inter-speaker's standard deviation for mean formant frequencies, for each of the seven vowel monophthongs, was found to be higher in loud and soft speech rather than in neutral, which suggests that the first two formants can be better used for differentiating speakers, both in acoustic analysis and in human perception and identification, when the speech mode is altered from neutral.

References

Elliott, J et al. 2000. Comparing the acoustic properties of normal and shouted speech: a study in forensic phonetics. In *Proc. sst-2000: 8th int. conf. speech sci. & tech*, 154–159.

Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen Christensen. 2015. Package: lmer test. *R package version 2.*

Remez, Robert E, Jennifer M Fellowes & Philip E Rubin. 1997. Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance* 23(3). 651.

Schulman, Richard. 1989. Articulatory dynamics of loud and normal speech. *The Journal of the Acoustical Society of America* 85(1). 295–312.