SAY AGAIN? INDIVIDUAL ACOUSTIC STRATEGIES FOR PRODUCING A CLEARLY-SPOKEN MINIMAL PAIR WORDLIST

James M. Scobbie and Joan Ma

Queen Margaret University, Edinburgh
jscobbie@qmu.ac.uk and jma@qmu.ac.uk

ABSTRACT

People make their speech clearer in difficult conversational contexts using global mechanisms (e.g. "Lombard Speech") and by targeted enhancements of linguistic constituents ("hyperspeech"). We describe production changes observed in four speakers of Scottish English who produced three repetitions of twelve CVC words: V was one of six monophthongs and C_C was either /p_p/ or /m_m/. Thus each word differed (near-) minimally from six others. In a "neutral" condition each participant read aloud from a randomised wordlist. A "clear" condition was an interactive task in which an interlocutor had to repeat back every word correctly, despite their hearing being impaired by headphone-delivered noise. If the speaker was mis-perceived by the interlocutor, the speaker tried again, until the word was correctly repeated. We describe the surprisingly speaker-specific acoustic hyperspeech effects (in vowel F1, vowel space area, and acoustic segment durations) in the clear speech. A companion paper describes the associated articulatory changes.

Keywords: Lombard speech, hyperspeech, acoustics, intelligibility, vowels.

1. 1. INTRODUCTION

Spoken words vary in response to a range of factors, such as the desire or need to speak clearly. Poor communicative conditions may trigger an increase in vocal effort, perhaps as a universal (reflex) Lombard effect. Greater speaker effort boosts intensity, pitch, duration, and other global factors [1] [4] [5] [6]. Clarity can also expand phonemic dispersion and enhance cues to contrast [2] [3] to maintain sufficient discriminability [7], perhaps with quantal effects [10]. But sociolinguistic [11] and affective [8] changes also interact with clarity. We thus expect (a) dialect-specific and (b) task-specific influences on clear speech, though we are not led to expect idiosyncratic yet systematic variation within dialect.

An independent area of interest is the production of single words. Though single full lexical words like "rabbit", "flower", "red" and "jumping" are unusual in real-world conversations (as opposed to discourse items and fillers), single-word utterances are not uncommon in a range of important if "artificial" contexts (e.g. psycholinguistic reaction-time experiments, speech acquisition studies, phonetics experiments, quizzes and educational tasks). Elicitation may be by picture naming, reading aloud, repetition, delayed naming, or cloze tasks.

Both these topics are relevant to the production of single words in the speech therapy clinic, where many normalised assessments and ad-hoc therapeutic activities involve single word production. (Hearing assessments and research into listening often use pre-recorded single word speech samples.) Moreover, a speaker is often explicitly asked to utter a single word *as clearly and accurately as possible*. In the (paediatric) clinical context, the client may be expected to produce their clearest possible versions of diagnostic wordlists for assessment. They may contain phonological minimal pairs or sets. Contrast enhancement may be part of the therapeutic process, intended to alter a speaker's productions permanently. Clinical meta-linguistic discourse involves therapist and client estimating the functional intelligibility and social acceptability of the client's production of phonemic contrasts.

We are therefore interested in the social-cum-interpersonal, linguistic-cum-dialectal, task-specific and universal factors that can be used to pronounce a single content word more clearly. What changes might a speaker make? Here, we explore a small set of phonemic distinctions in single words (for the reasons above). Specifically, we ask how each speaker produces the words within-dialect to a physically-present, sighted interlocutor whose hearing is at first normal, (in which case intelligibility is 100%), then temporarily impaired, modelled experimentally by wearing headphones delivering loud aperiodic noise.

Our study provides a baseline for research into changes in segment production which speakers (choose to) make to enhance intelligibility. In the longer term we want to elicit variation in a wider range of materials, with alternative tasks, and using dialectally-varied or cross-linguistic interlocutors. Here, we consider various measures including vowel formant space related to segmental dispersion as well as some general reflexes of clarity.

For space reasons we report acoustic measures only, but see [9] for a companion paper analysing the same speakers' tongue and lip articulations.

2. METHOD

In Scottish English, six "unchecked" monophthongal vowels /ieaɔoʉ/ can appear in open or closed syllables. /ɔoʉ/ are phonologically rounded. Two C_C contexts were chosen, in which C was labial (either /m/ or /p/). Thus the wordlist mostly included real words (*pope*) but also pseudowords (*moam*).

Three tokens of each word were incorporated into two speaker-specific randomised wordlists (n=36). First, in the neutral condition, the interlocutor was present but did not repeat each word as it was read aloud. In the second condition, intended to elicit clear speech, the interlocutor faced the speaker at about a 2m distance, and repeated what was perceived, out loud. If the response was correct, the speaker moved on. If the response was incorrect, the speaker had to repeat the item in the list. The interlocutor (1st author) was blinded to the randomisation, but not to the 12 possible targets. They listened to speech spectral noise at a 50dB setting, partially masking the speaker's normal conversational volume.

Since the speaker had to repeat the item if the interlocutor mis-heard (and could detect levels of uncertainty even if correct), we assume that on average the second condition elicited clear speech, but it was obviously not shouted or un-natural.

For the acoustic analysis, standard segmentation processes were followed. Closure of initial and final /m/ and final /p/ were analysed for duration, along with VOT of initial /p/ and the vowel duration. Acoustic word duration was the sum of these. Formant analysis was performed in PRAAT with F1 and F2 (and F3, not analysed here) extracted in the first and last 25% of the vowel on the few occasions the medial 50% included clipping as a result of increased intensity in the clear condition, but mostly formant values were averaged throughout the vowel. Formant values were converted from Hz to Bark. The vowel-space area was then estimated as the sum of the area of series of scalene triangles, but is represented below with a curved perimeter. Since there are just four speakers, results are descriptive, and we do not report any pilot inferential statistics.

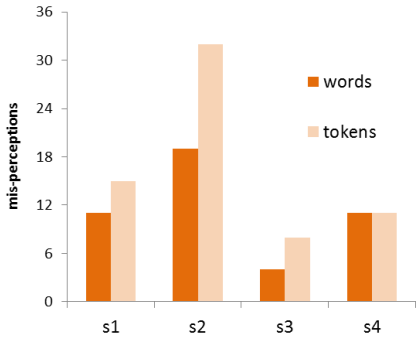
3. RESULTS

3.1 Functional intelligibility

The speakers were 100% perceptible in the neutral condition, though the interlocutor (who was present, but silent) considered S2 to be the least distinct.

The consistency and ease with which each speaker attained 100% functional intelligibility in the clear condition varied (Fig 1). S2 had the highest rate of mis-perceptions, having to repeat 19 target words out of 36, with over 30 repeat attempts. Qualitatively, the interlocutor found S3 easiest to perceive. S1, S2 and S4 were "hard work", requiring careful active listening and lip-reading.

Figure 1: Numbers of mis-perceptions during the process of achieving 100% correct responses.



3.2 Global differences (quasi-Lombard effect)

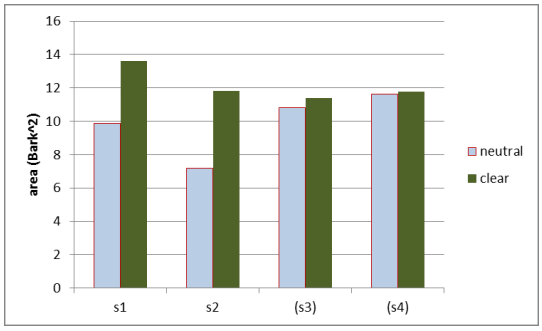
All four speakers increased their global vocal effort in an impressionistic sense. Overall, recordings of the clear condition demonstrated an increase in loudness, and the recorded waveforms had greater intensity, though neither has been quantified.

3.3 Acoustic measurements

Where it makes sense, we will present averages of all four speakers, and/or all the vowels. Otherwise, we focus on the descriptive presentation of individual words, speaker by speaker.

S1 and S2 increased the vowel space area in the clear speech condition (Fig 2). Fig 3 shows that the increase was (primarily) due to an increase in F1.

Figure 2: Acoustic vowel space area, neutral (pale bars) vs. clear speech condition (dark).



The speakers used duration in conflicting ways, e.g. in the acoustic duration of the whole word (Figs 5 & 6). Not only did speakers have different patterns

in the neutral condition (e.g. S1 vs. S4), the change in the clear condition varied (and S4 consistently made none). S1 increased word duration for /m/ words in the clear speech condition, but not /p/ words. S2's /m/ words also seemed longer than their /p/ words, but with no clear condition effect. S3's clear speech approach may have been to increase duration generally. Acoustic word duration is a composite of segment effects, of course.

C1 duration cannot be addressed uniformly. For /p/, VOT was measured (Figs 6 & 7). More speakers are needed, but it appears some shortened VOT but some lengthened it. For /m/ (Figs 6 & 7), the consonant was longer in clear speech (S1, S3) or showed no difference (S4). S2's pattern was unclear.

The duration of C2 (Fig 8) was even less clear, and we are reticent to offer a simple descriptive view: more data is needed. One participant (S1), however, seemed to reduce the closure duration of C2 (/p/ and /m/ alike) in the clear speech condition.

Figure 3: Vowel area changes, showing increased F1 in speaker S1 (upper panel) and S2 (lower panel). In this and following figures, the solid line with square markers is for the clear condition.

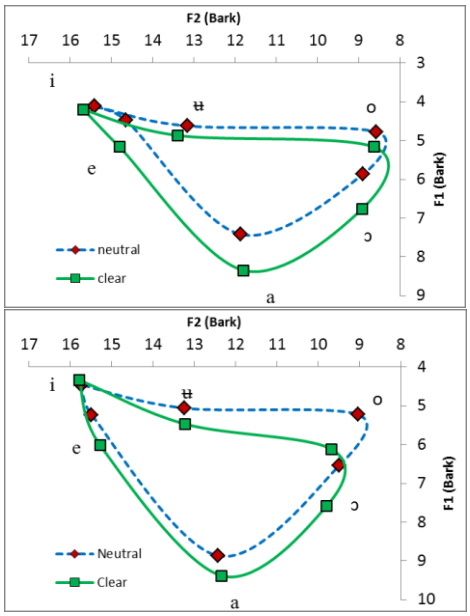
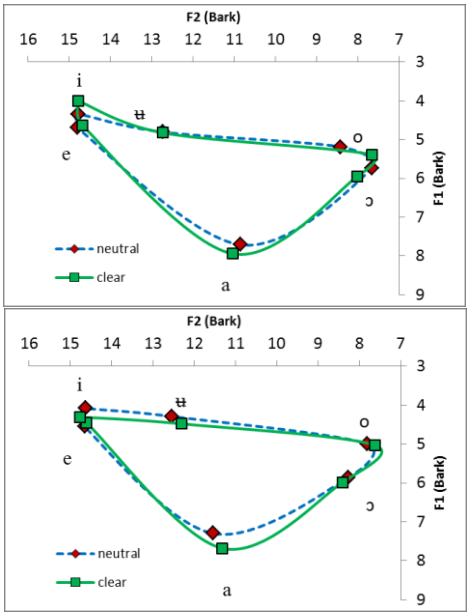


Figure 4: S3 (upper) and S4 (lower).



Finally, vowel duration was complex (Fig 9). S1, S2 and S4 had a very substantial increase in vowel duration in the clear condition, and vowel duration that was similar in /m/-words and /p/-words. S3's non-high vowels were long in /m/-words in both conditions, and shorter in /p/-words in the neutral condition (but the clear condition was variable).

4. DISCUSSION AND CONCLUSIONS

Speakers of Scottish English produced clearer speech in an interactive task which unusually used single word utterances. We focused our analysis on segmental enhancement rather than prosody or voice quality, and found that speakers seemed to enhance incompatible aspects of their system. A companion paper on lip and tongue articulation [9] shows yet more disparity in the strategies these speakers used to make similar words more clearly distinct. We hypothesise that phonological enhancement can be systematically idiosyncratic.

Figure 5: Acoustic word duration, clear condition (solid) vs. neutral (dashed), /m/-words (dark) vs. /p/-words (light). S1-S4 are shown left-to-right.

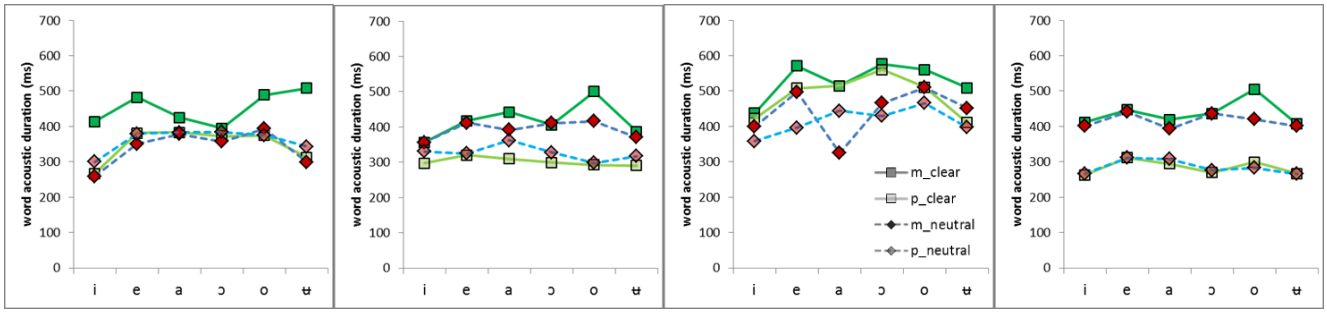


Figure 6: Acoustic word duration (left), /m/-words (dark bars) and /p/-words (pale), mean /p/ VOT (centre) and mean C1 /m/ duration (right), both with neutral (pale bars) vs. clear speech conditions (dark). Whiskers = 1 s.d.

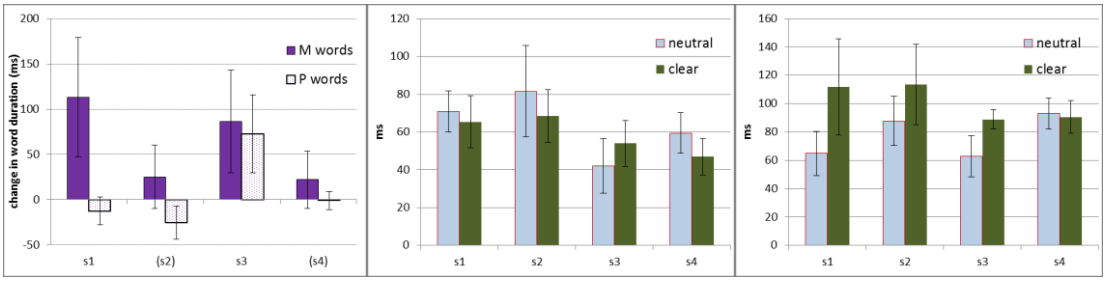


Figure 7: C1 acoustic segment duration, clear condition (solid) vs. neutral (dashed). S1-4 shown left to right, and /p/ VOT duration in the upper panels (pale) and /m/ closure duration in the lower panels (dark).

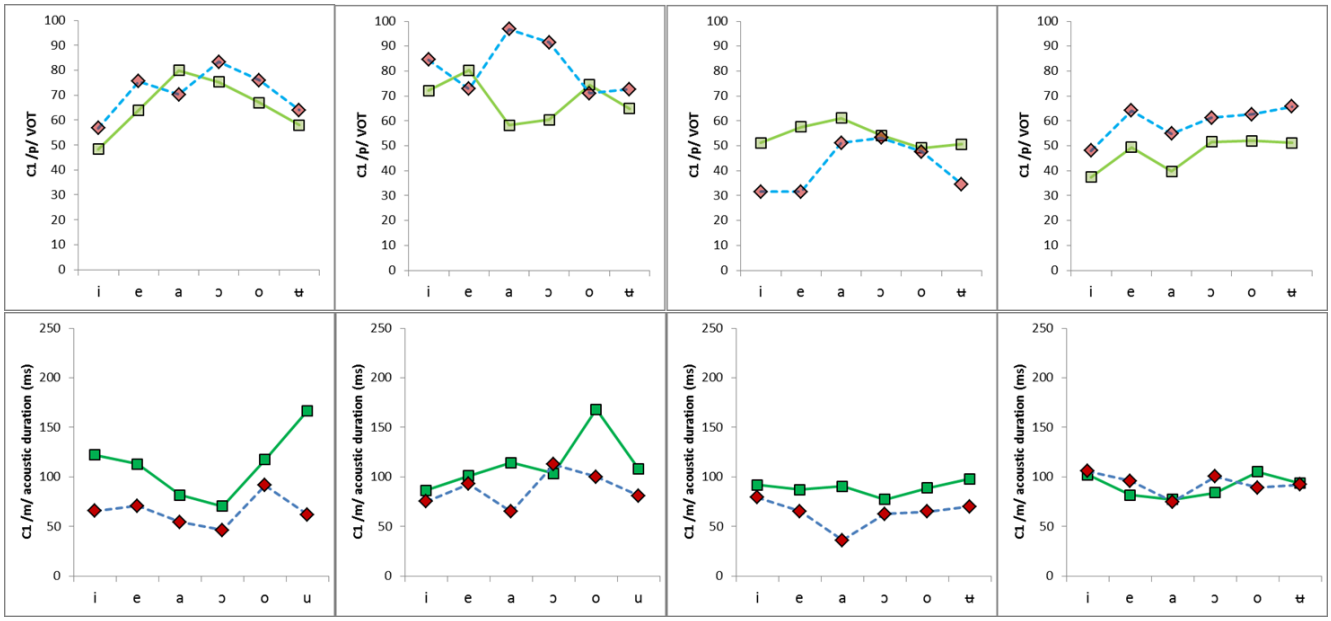


Figure 8: C2 acoustic duration, in clear condition (solid) vs. neutral (dashed); /m/-words (dark) vs. /p/-words (pale).

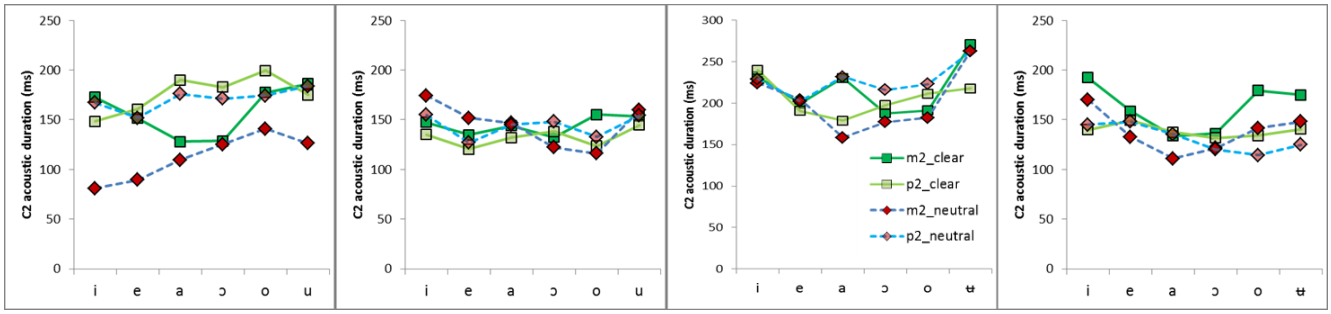
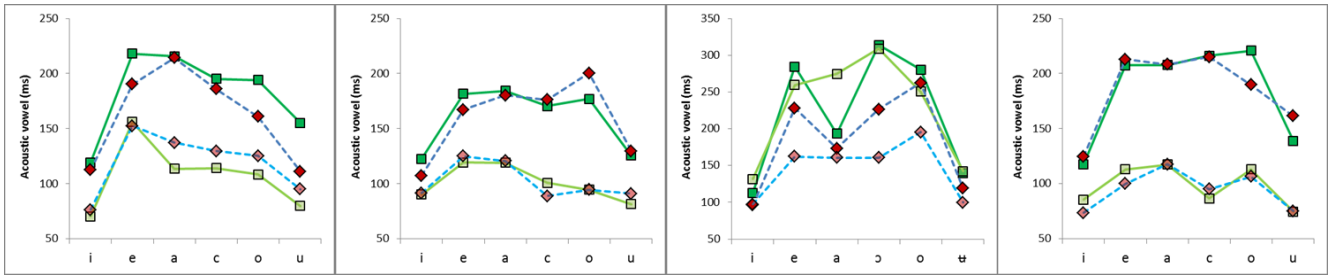


Figure 9: Vowel duration, in clear condition (solid) vs. neutral (dashed); /m/-words (dark) vs. /p/-words (pale).



5. REFERENCES

[1] Castellanos, A., Benedi, J. M., Casacuberta, F. 1996. An analysis of general acoustic-phonetic features for Spanish speech produced with the Lombard effect. *Speech Comm*. 20, 23–35.

[2] Garnier, M., Ménard, L. Alexandre, B. 2017. Hyper-articulation in Lombard speech: An active communicative strategy to enhance visible speech cues? *J. Acoust. Soc. Am*. 144(2), 1059–1074.

[3] Hazan, V., Kim, J. 2013. Acoustic and visual adaptations in speech produced to counter adverse listening conditions. *Proc. AVSP'13* Annecy, 93–98.

[4] Hazan, V., Tuomainen, O., Kim, J. Davis, C. Sheffield, B., Brungart, D. 2018. Clear speech adaptations in spontaneous speech produced by young and older adults. *J. Acoust. Soc. Am*. 144(3), 1331–1346.

[5] Kim, J., Davis, C. 2014. Comparing the consistency and distinctiveness of speech produced in quiet and in noise. *Comp. Speech Lang*. 28, 598–606.

[6] Liénard, J.S., Di Benedetto, M.G. 1999. Effect of vocal effort on spectral properties of vowels. *J. Acoust. Soc. Am.* 106(1), 411–422.

[7] Lindblom 1990. Explaining phonetic variation: a sketch of the H&H theory. In: Hardcastle, W.J., Marchal, A. (eds), *Speech Production and Speech Modelling*. Dordrecht: Kluwer Academic Publishers, 403-439.

[8] Rilliard, A., d'Alessandro, C., Evrard, M. 2018. Paradigmatic variation of vowels in expressive speech: Acoustic description and dimensional analysis. *J. Acoust. Soc. Am*., 143(1), 109–122.

[9] Scobbie, J.M., Ma, J. 2019. Say again? Individual articulatory strategies for producing a clearly-spoken minimal pair wordlist. *Proc. 19th ICPhS* Melbourne.

[10] Stevens, K.N., Keyser, S.J. 2010. Quantal theory, enhancement and overlap. *Jou. Phon*. 38(1), 10–19.

[11] Wassink, A. B., Wright, R. A., Franklin, A. 2007. Intraspeaker variability in vowel production: an investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers. *Jou. Phon*. 35(3), 363–379.