Individual Differences in the Formant Dynamics of Vowels at Different Levels of Stress.

Kirsty McDougall
University of Cambridge, United Kingdom
E-mail: kem37@cam.ac.uk

ABSTRACT

This paper examines individual differences in aspects of the formant dynamics of vowels at three levels of prosodic stress. Experiment 1 compares formant contours of /aɪ/ under nuclear and non-nuclear stress produced by five subjects. Experiment 2 examines vowel-to-vowel coarticulation differences in the formant frequencies of /ə/ produced by five subjects. The trajectory of F1, F2 and F3 of /aɪ/ yields distinguishing features for each of the speakers, with differences particularly marked at the non-nuclear stress level. Individual speakers coarticulate full vowels with /ə/ to varying extents in F1 and F3, and to a greater but more uniform extent in F2, however these differences are not as clearly speaker-specific as those identified in the formant dynamics of the stressed vowels in Experiment 1.

1. INTRODUCTION

This paper is concerned with why speakers, even members of the same sex and accent group, sound different from one another. Different sounding voices result from differences between speakers in their vocal morphology combined with differences in their articulatory movements. Individual differences in the operation of the vocal apparatus can be understood in terms of a task-dynamic model of speech production in which speakers can use functionally equivalent combinations of articulatory gestures to achieve the same phonetic goal [1, cf. 6]. If speakers do use different motor control strategies to approach phonetic goals, one would expect evidence of this in their formant frequency dynamics. This paper investigates how individual differences are manifest in the formant dynamics of vowels at different levels of stress, and in vowel-to-vowel coarticulatory effects on the formant dynamics of unstressed vowels.

Various studies have identified speaker-specific variation in formant frequency contours [e.g. 4, 5, 11], in coarticulation [e.g. 10, 13], and in vowel-to-vowel coarticulation in particular [2, 7]. Initial attempts to utilise information from coarticulation for speaker discrimination have been made by van Heuvel et al. [13]; an object of the present study is to ascertain whether speaker-specific variation in vowel-to-vowel coarticulation can be thus employed.

Speaker-variability in the production of /ə/ has not been investigated extensively. In /ə/ produced by four American English speakers, Gick observed notable differences between /ə/ in lexical versus functional words for one speaker, a distinction not present for at least two subjects [3]. Van Bergem found considerable between-speaker variation in /ə/ produced by three Dutch speakers, with significant but less marked effects of neighbouring consonants and flanking vowels [12]. Between-speaker variation was attributed to differences such as those in vocal tract length; coarticulatory behaviour of individual speakers in terms of interactions between the factors Speaker and Neighbouring Segment is not analysed. These studies provide preliminary evidence of speaker-specific behaviour in /ə/ production; more detailed research is needed in this area.

In the present study, Experiment 1 aims to determine the extent of variation present in F1, F2 and F3 contours of /aɪ/ under nuclear and non-nuclear stress when produced by different speakers. (This work is reported in greater detail in [8, 9].) A possible tendency for formant dynamics of /aɪ/ under a lower degree of stress to reveal clearer differentiation between speakers prompted interest in formant dynamics related to the unstressed vowel /ə/. Experiment 2 examines individual differences in the formant dynamics of vowel-/ə/-vowel sequences. It investigates whether a speaker can be characterised by the coarticulatory effects of full vowels on F1, F2 and F3 of /ə/. If the results show a relationship between different levels of stress and differences in individuals' speech behaviour, this will have implications for forensic speaker identification and speaker recognition, as well as speech production theory.

2. METHOD

2.1 SUBJECTS

Experiment 1: Subjects were 5 male Australian English speakers from Brisbane or the Gold Coast, Queensland, aged 22-28 years (denoted A, B, C, D, E).

Experiment 2: Subjects were 5 male Standard Southern British English speakers, aged 24-32 (denoted F, G, H, I, J).

2.2 MATERIALS

Experiment 1: The test sentences elicited productions of the syllable /aɪk/ in *bike*, *hike*, *like*, *mike* and *spike*. Each /aɪk/ word took nuclear stress in one sentence, and non-nuclear stress in a second, with the phoneme sequences before and after the word in the two sentences kept identical as far as possible, as in Table 1. Five instances of each sentence were randomly arranged along with five of each of four foil sentences. Subjects were asked to read the sentences firstly at a normal, then at a fast but still natural speed. Thus the target sentences were produced five times at two speaking rates, i.e. 10 x 5 x 2 = 100 utterances for each speaker.

Experiment 2: The test sentences elicited production of /ə/ in the nonsense context /bV1bəbV2b/, with V1 taking

ISBN 1-876346-48-5

Nuclear target word	Non-nuclear target word
I don't want the scooter, I want the *bike* now.	Later won't do, I want the bike *now*.
We're not driving, he said we might *hike* there.	It's not definite, he said we *might* hike there.
She's never going to *like* you.	She only likes basketballers, so she's never going to like *you*.
Not Kate, I think *Mike* sings well.	He's hopeless at dancing, but I think Mike *sings* well.
Careful of that cactus, it might *spike* you.	I'm a good distance from the cactus, but it might spike *you*.

Table 1: Test sentences for Experiment 1. Words taking nuclear stress are in italics.

nuclear stress. V1 and V2 were all possible combinations of the 'corner vowels' /i, æ, a, u/. The corner vowels were represented in English orthography as /i/ 'ee', /æ/ 'a', /a/ 'ar' and /u/ 'oo'; /ə/ was represented using the indefinite article. The sentences were of the following form, with the second /bV1bəbV2b/ sequence (hereafter 'test sequence') in each sentence being analysed only:

Not BEEB a barb, it sounded like BARB a barb.

Nuclear stress was shown by capitalising the two words containing a V1 in each sentence. The two V1 vowels were always contrasting in each sentence, and the two V2 vowels always matched. The materials were designed to elicit six repetitions of each of the sixteen V1-V2 combinations appearing in the test sequence, such that each speaker produced 16 x 6 = 96 utterances. The target sentences were arranged in random order, with an extra item of the same form included at the end of each page to carry any end-of-list intonation effects. Each subject was asked to read the sentences in a normal relaxed speaking style.

2.3 RECORDING

Recordings were made in the sound-treated booth in the Phonetics Laboratory in the Department of Linguistics, University of Cambridge. Each subject was seated with a Sennheiser MKH 40P48 condenser microphone positioned about 20 cm from his mouth. The recordings were made on Digital Audio Tape using a Sony DTC-60ES DAT recorder. The subjects practised reading the sentences aloud before the recording was made.

2.4 MEASUREMENT

Recordings were digitised at 16 kHz on a Silicon Graphics workstation, and analysed using *xwaves+*. A wide-band spectrogram was produced for each utterance and vertical markers placed by hand to segment each /aɪ/ token for Experiment 1, and for Experiment 2, V1, /ə/, and V2 from the test sequence. For both experiments the total duration of each vowel, and the time-points at which formant frequencies were measured were computed automatically. Centre frequencies of F1, F2 and F3 were measured at intervals 10% of the duration of each /aɪ/ segment for Experiment 1, and at the midpoints of V1, /ə/ and V2 for Experiment 2. Formant frequencies were measured manually using 18-pole auto-correlation LPC spectra with a 50ms Hanning window, with continuity of the formant contours on the spectrogram guiding the measurements. In cases where formants were unclear, the auto-correlation spectrum was supplemented by a DFT spectrum with a 50ms Hanning window.

3. RESULTS AND ANALYSIS

3.1 EXPERIMENT 1

The five speakers' mean trajectories of /aɪ/ differed in both shape and relative frequency, as can be seen in Figure 1.

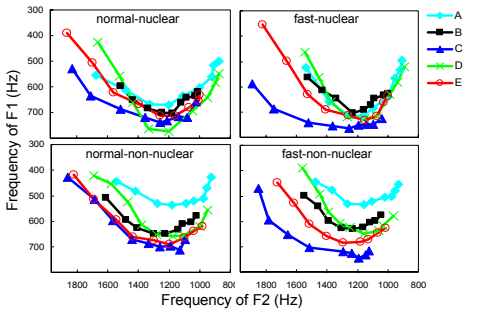

Figure 1: Mean F1 and F2 frequencies plotted at each +10% step of /aɪ/, for each speaker, with a separate panel for each rate-stress condition. The 10% point is the rightmost of each curve, and the 90% point leftmost.

In both F1 and F2, individuals vary in onset and offset positions and in the dynamics of their trajectories between these two positions. Similar individual differences are also present in the contours of F3. Pairwise comparisons from ANOVAs (Speaker (5) x Speaking Rate (2) x Stress (2)) run at each +10% step of /aɪ/ for each of F1, F2 and F3 confirmed these findings. Differences between mean F1, F2 and F3 frequencies were significant among most pairs of speakers at the majority of +10% steps of /aɪ/ ($p < 0.01$). Speaker-distinguishing features of the /aɪ/ contours were generally consistent across the different speaking rates and levels of stress, but more marked in some conditions. The five speakers' trajectories were spread furthest apart in fast-non-nuclear /aɪ/, and least distinct in normal-nuclear /aɪ/. For normal-non-nuclear /aɪ/, the trajectories were well spread during the early (steady state) part of the diphthong, and for fast-nuclear /aɪ/, the trajectories were more distinctive during the glide. Combining these observations, it could be interpreted that normal-nuclear speech furnishes fewer differences between speakers as here they most deliberately aim for a 'canonical' form, whereas faster speech under non-nuclear stress may exhibit more speaker-specific patterns of speech motor control.

For further analysis of the speaker-characterising properties of the formant dynamics of /aɪ/, including discriminant analyses, the reader is referred to [9]. The possibility that the formant dynamics of a full vowel under a lower degree of stress might exhibit greater differences between speakers raised the issue of whether formant dynamics of the unstressed vowel /ə/ might demonstrate such differences, as is investigated in Experiment 2.

3.2 EXPERIMENT 2

Coarticulatory effects of preceding and following vowels were observed in the first three formant frequencies of /ə/ (henceforth əF1, əF2 and əF3) for all five speakers, as shown by the vowel quadrilateral plots in Figure 2.

For all speakers, the range of əF1 is much smaller than the range for əF2, whether the data is grouped by V1 (Figure 2a)

ISBN 1-876346-48-5

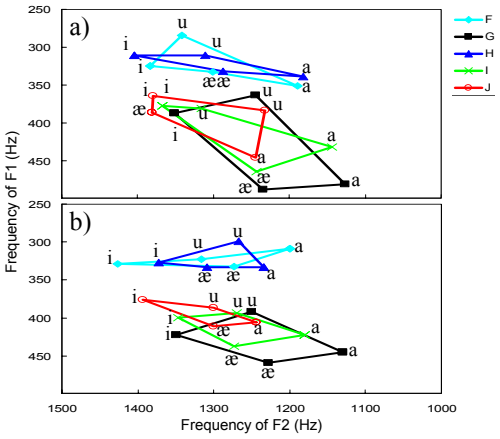

Figure 2: Mean F1 and F2 frequencies of /ə/ for each speaker, a) after each V1, with the data pooled over V2, b) before each V2, with the data pooled over V1. Points are labelled by the identity of the preceding vowel (V1) in a, and by the following vowel (V2) in b**.**

or V2 (Figure 2b). Although the extent of coarticulation in F2 is greater than in F1, coarticulatory behaviour is relatively uniform in F2 and between-speaker differences are more distinctive in F1. It seems that F and H realise /ə/ with a lower F1 and with a lesser degree of coarticulation with V1 than the other three speakers. F and H do not exhibit strikingly lower F1s than other speakers in their full vowels, so their low əF1 means appear attributable to different /ə/ targets rather than differences in vocal tract length. Greater coarticulatory effects of V1 were exhibited in əF1 produced by G, I and J, with G exhibiting the greatest coarticulation among the speakers. Coarticulatory effects on əF3 also exhibited between-speaker variation, with some speakers' əF3 more influenced by V1 or V2 than others.

The statistical reliability of these observations was confirmed by univariate ANOVAs with the factors Speaker (5) x V1 (4) x V2 (4), run separately on each of əF1, əF2 and əF3, as shown in Table 2. A significant effect of V1 or V2 was considered evidence of vowel-to-vowel coarticulation on /ə/, and an interaction between Speaker and V1, V2 or V1 x V2 was considered evidence that the coarticulatory behaviour of individual speakers affected the formant frequencies of /ə/. Here, Speaker interacted with V1 for each of əF1, əF2 and əF3, and with V2 for əF2 and əF3. For each significant interaction, Tukey post hoc comparisons were carried out to determine whether effects of speaker were significant across all levels of V1 or V2.

Significant differences were present among some but not all pairs of speakers in their realisation of əF1 for each level of V1. The speakers appear to form two groups, F and H versus G, I and J, such that əF1 of a given speaker from one group is significantly different from that of all speakers in the other group, but generally not from those of his own group. Speaker-distinguishing patterns were fewer and sporadic in əF2 with respect to both V1 and V2 qualities, indeed for V1 = /æ/ or /i/ and V2 = /i/ or /u/ there were no significant differences between speakers. əF3 exhibited some significant differences between speakers with respect to both V1 and V2, but with no speaker behaving consistently differently from all others.

Throughout this study, effects noted were generally similar for both V1 and V2 in corresponding formants, apart from in F1, where Speaker interacted with V1 but not V2. It is worth noting that in the data presented here V1 was always under nuclear stress, but V2 was not. Further work is required to determine whether these patterns also hold when V2 is nuclear-stressed.

To establish the relative extent of vowel-to-vowel coarticulation exhibited by individual speakers, the measurements of əF1, əF2 and əF3 were plotted against the corresponding formant frequencies for each of V1 and V2, for each speaker. A speaker whose /ə/ is strongly influenced in the *i*th formant by adjacent vowels would be expected to produce a scatterplot with a wide spread of measurements of əF*i*. The lowest values of əF*i* should be those corresponding to the adjacent vowel with the lowest *i*th formant, and əF*i* should increase with correspondingly higher values of F*i* in the adjacent vowel, as represented in Figure 3a. A speaker whose /ə/ is little affected by the adjacent vowel would be expected to exhibit a smaller range of measurements of əF*i*, with no obvious relationship between increasing əF*i* and the relative ranking in corresponding adjacent vowels of F*i* (Figure 3b).

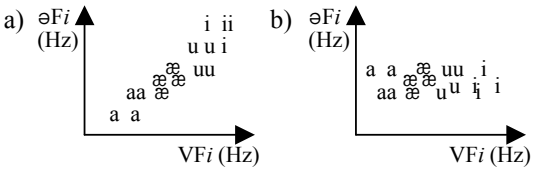

Figure 3: Schematised scatterplots of VF*i* versus əF*i*, for single speakers, with each point representing the identity of V. Speaker exhibiting a) strong evidence, and b) little evidence, of V-to-/ə/ coarticulation in the *i*th formant.

Spearman's rank correlation coefficients were calculated to determine the strength of the relationship between each əF*i* and the corresponding formant frequencies of V1 and V2, as in Table 3.

	əF1	əF2	əF3
Speaker	F(4,400) = 57.658, p = **0.000**	F(4,400) = 19.905, p = **0.000**	F(4,400) = 68.733, p = **0.000**
V1	F(3,400) = 31.785, p = **0.000**	F(3,400) = 213.984, p = **0.000**	F(3,400) = 79.167, p = **0.000**
V2	F(3,400) = 6.710, p = **0.000**	F(3,400) = 157.041, p = **0.000**	F(3,400) = 2.662, p = 0.048
Speaker x V1	F(12,400) = 3.410, p = **0.000**	F(12,400) = 3.827, p = **0.000**	F(12,400) = 8.949, p = **0.000**
Speaker x V2	F(12,400) = 1.032, p = 0.418	F(12,400) = 2.777, p = **0.001**	F(12,400) = 4.016, p = **0.000**
V1 x V2	F(9,400) = 1.446, p = 0.166	F(9,400) = 2.671, p = **0.005**	F(9,400) = 2.823, p = **0.003**
Speaker x V1 x V2	F(36,400) = 0.881, p = 0.668	F(36,400) = 1.302, p = 0.119	F(36,400) =1.502, p = 0.035

Table 2: F-ratios and p-values of ANOVAs (Speaker (5) x V1 (4) x V2 (4)) on əF1, əF2 and əF3 (NS if p > 0.01).

ISBN 1-876346-48-5

	Corr(V1, (V1)ə)			Corr(V2, ə(V2))		
Speaker	F1	F2	F3	F1	F2	F3
F	0.166	0.501	0.327	-0.018	0.588	0.084
G	0.397	0.573	0.328	0.117	0.576	0.078
H	0.070	0.620	0.172	-0.018	0.367	0.238
I	0.387	0.552	0.078	0.264	0.429	0.068
J	0.441	0.485	0.316	0.221	0.432	0.207

Table 3: Correlation between formant frequencies of V1 and (V1)ə, and of V2 and ə(V2) (Spearman's ρ).

Speakers coarticulated to differing extents on each of the formants, with respect to both V1 and V2. In some cases little coarticulation was demonstrated by all speakers (V2F1 ~ əF1), in others the speakers produced comparable levels of coarticulation (V1F2 ~ əF2, V2F2 ~ əF2), while in other cases again, certain speakers coarticulated more strongly than others (V1F1 ~ əF1, V1F3 ~ əF3, V2F3 ~ əF3). The small size of the sample investigated here should be borne in mind in interpreting the results. Some individual differences in coarticulatory behaviour have been highlighted in aspects of the formant frequencies of /ə/, and given the nature of the differences found, it is possible that a larger sample of speakers would vary further in the range of differences yielded. However, speakers also demonstrated their capacity to behave in similar ways in many respects, and the usefulness of properties of their vowel-to-vowel coarticulation for applications requiring fine-grained between-speaker discrimination such as for forensic speaker identification appears less obvious.

4. CONCLUSION

This study has investigated individual differences in the formant dynamics of vowels under three levels of stress: nuclear-stressed, non-nuclear-stressed, and unstressed. Speaker-specific properties were observed in the first three formant contours of /aɪ/ under both nuclear and non-nuclear stress. A tendency for the formant dynamics of /aɪ/ at the lower stress level to exhibit more marked differences between speakers led to an investigation of speaker-specific differences in the formant dynamics relating to the unstressed vowel /ə/. Coarticulatory effects of full vowels on /ə/ demonstrated some individual differences in speech behaviour, with F1 and F3 of /ə/ exhibiting varying degrees of vowel-to-vowel coarticulation for each speaker, and F2 presenting greater but more uniform coarticulation among speakers. While the formant dynamics of vowel-/ə/-vowel sequences did provide evidence of individual patterns of speech motor control, given their vague and sporadic nature it is not obvious how these tendencies could be usefully implemented for characterising speakers.

The findings described so far emphasise the need for further research in this area. The /aɪk/ data has highlighted the strong speaker-characterising potential of the formant contours of dynamically complex phonetic events, and raised interesting questions about the role of stress in analysing individual differences in speech behaviour. While the /ə/ data did not lend itself directly to a technique for discriminating between speakers, it does provide further evidence of speaker idiosyncrasies in coarticulation, an area requiring more detailed investigation.

ACKNOWLEDGEMENTS

Thanks are due to Francis Nolan and Sarah Hawkins for their advice during this study. This research is funded by the Cambridge Commonwealth Trust, the Clare College Mallinson Fund, and an IFUW Marjorie Shaw Fellowship.

REFERENCES

[1] C.P. Browman and L. Goldstein, "Articulatory gestures as phonological units," *Phonology*, vol. 6, pp. 205-251, 1989.

[2] C.A. Fowler, "Production and perception of coarticulation among stressed and unstressed vowels," *Journal of Speech and Hearing Research*, vol. 46, pp. 127-139. 1981.

[3] B. Gick, "An X-ray investigation of pharyngeal constriction in American English schwa," *Phonetica*, vol. 59, pp. 38-48, 2002.

[4] R. Greisbach, E. Osser and C. Weinstock, "Speaker identification by formant contours," in Studies in Forensic Phonetics, A. Braun and J. Köster, Ed. *Beiträge zur Phonetik und Linguistik*, vol. 64, pp. 49-55, 1995.

[5] J.C.L. Ingram, R. Prandolini and S. Ong, "Formant trajectories as indices of phonetic variation for speaker identification," *Forensic Linguistics*, vol. 3.1, pp. 129-145, 1996.

[6] K. Johnson, P. Ladefoged and M. Lindau, "Individual differences in vowel production," *Journal of the Acoustical Society of America*, vol. 94, pp. 701-714, 1993.

[7] H. Magen, "The extent of vowel-to-vowel coarticulation in English," *Journal of Phonetics*, vol. 25, pp. 187-205, 1997.

[8] K. McDougall, *Individual Differences in Formant Dynamics of* /aɪ/ *at Normal and Fast Tempos*, M.Phil. Dissertation, Department of Linguistics, University of Cambridge, 2001.

[9] K. McDougall, "Speaker-characterising properties of formant dynamics: a case study," in *Proceedings of the 9th Australian International Conference on Speech Science and Technology, Melbourne*, pp. 403-408, 2002.

[10] F. Nolan, *The Phonetic Bases of Speaker Recognition*, Cambridge: Cambridge University Press, 1983.

[11] P. Rose, "Differences and distinguishability in the acoustic characteristics of *Hello* in voices of similar-sounding speakers: a forensic phonetic investigation," *Australian Review of Applied Linguistics*, vol. 22.1, pp. 1-42, 1999.

[12] D.R. van Bergem, "A model of coarticulatory effects on the schwa," *Speech Communication*, vol. 14, pp. 143-162, 1994.

[13] H. van den Heuvel, B. Cranen and T. Rietveld, "Speaker variability in the coarticulation of /a,i,u/," *Speech Communication*, vol. 18, pp. 113-130, 1996.