SPEAKER CHARACTERISTICS IN THE COARTICULATION OF THREE DUTCH VOWELS /a,i,u/

H. van den Heuvel, B. Cranen & T. Rietveld
Dept. of Language and Speech, University of Nijmegen, The Netherlands
E-mail: heuvel@let.kun.nl

ABSTRACT

We investigated the coarticulation in the first three formants of the Dutch vowels /a,i,u/, and the speaker variability in this coarticulation. We found the largest amount of coarticulation in the vowel /u/, somewhat less in /a/, and hardly any in /i/. The amount of coarticulation as a function of context turned out to be speaker-dependent for /a/ and /u/. However, as for our data coarticulation proved not to be an important parameter for speaker identification by computer.

INTRODUCTION

Speaker variability in articulation patterns has not been investigated on a large scale, because the invariant aspects of speech production have been considered more interesting from a linguistic point of view. Nonetheless a few studies were published dealing with this topic [1,2]. In the light of these studies it can be hypothesised that coarticulation may exhibit substantial between-speaker variability.

For some phonemes a confirmation for this hypothesis has been found in the field of (automatic) speaker identification. Su, Li & Fu (1974) [3] noted that the amount of coarticulation in nasals (and especially in /m/) varies highly among speakers and can, as a result, effectively be used in automatic speaker identification. Comparable experiments are reported for /l/ and /r/ by Nolan (1983) [4]. In his study the coarticulation in /l/ appeared more speaker-specific than that in /r/. Similar experiments for vowels have not been carried out so far, which is surprising since the coarticulation in vowels as such has been studied extensively.

The main aim of this paper is to ascertain firstly whether coarticulation in vowels is speaker-specific, and secondly to investigate if it can be beneficially used in a speaker identification task. Before we examine this, we will first have a look at the speech data used and at the coarticulation that was observed in /a,i,u/.

SPEECH DATA

In order to keep the experiment within practical limits and to have control over the number of factors that may effect the vowel formants, we opted for a rather restricted data-set. The data set used consisted of 24 /CVCə/ (mainly) pseudo-words spoken in isolation. The three nucleus vowels used were /a,i,u/ and the eight consonants, which appeared once as C_1 and once as C_2 for each vowel /a,i,u/, were /p,t,k,d,s,m,n,r/. See table 1.

The 24 words were printed in a random order on ten 30-word word lists, in a, for Dutch, plausible spelling. All ten word lists were read out by each speaker in one recording session. The initial three words served as fillers, as did the final three. In this way (24 words $\cdot$ 10 repetitions =) 240 words were obtained for every speaker. Since fifteen speakers participated in the ex-

Table 1: The /CVCə/ pseudo-words used in the experiment, in phonemic representation.

$V =$	/a/	/i/	/u/
$C_1 =$ /p/	/pasə/	/pinə/	/pudə/
$C_1 =$ /t/	/tanə/	/tirə/	/tumə/
$C_1 =$ /k/	/kadə/	/kimə/	/kunə/
$C_1 =$ /d/	/dakə/	/disə/	/dupə/
$C_1 =$ /s/	/sapə/	/sidə/	/surə/
$C_1 =$ /m/	/marə/	/mikə/	/mutə/
$C_1 =$ /n/	/natə/	/nipə/	/nukə/
$C_1 =$ /r/	/ramə/	/ritə/	/rusə/

periment a total of 3600 word tokens were collected. The subjects were all male native speakers of Dutch and aged between 20 and 30 years.

The speech data were digitised with a 12-bit AD-converter at a sampling frequency of 16 kHz. Each word was segmented into phoneme-sized units; the nucleus vowel (i.e. /a,i,u/) was additionally split up into a steady-state part flanked by two transitions.

The formants were calculated by means of an LPC-analysis (pitch-asynchronous autocorrelation method; window length 25 ms; frame shift 5 ms; order 20). Root solving was used to obtain the formant values (in Hz). The target formants were extracted from the middle frame of the steady-state of each vowel token (/a,i,u/). Only the first three formants F_{1-3} were used. The formant positions of F_{1-3} in all selected vowel middle frames were checked by hand. They were converted to Barks to prevent that the variations in the higher formants F_2 and F_3 would obtain a dominating weight.

MEASURING THE COARTICULATION IN /a,i,u/

In our score model the coarticulation (COART) in a formant i in a specific context c in replication r as realised by speaker s is given by

$$COART(s,c,r,i) = (f_{scr}(i) - f_{s(ref).}(i))^2,$$

where $f_{scr}(i)$ refers to a raw formant value (obtained from the middle frame of a vowel token) and $f_{s(ref).}(i)$ to the (speaker-dependent) reference value of the vowel formant.

The best reference for coarticulation is the vowel spoken in isolation or in a /hVt/-context [5]. However, our speakers did not produce /a,i,u/ in isolation nor in a /hVt/-context. In order to obtain good estimates of these formant values for our experiment, we used the vowel formant values for /a,i,u/ as published in [5:1094] as initial references. Next we used each context of a vowel as reference and looked which context resulted in COART-values with the closest match to the COART-values resulting from the values given in [5]. These contexts were selected as the ultimate references. It were /pasə/ for /a/; /ritə/ for /i/; /surə/ for /u/.

An ANOVA on the (3 vowels · 8 contexts =) 24 COART-values of the three vowels, followed by a Tukey HSD post-hoc comparison ($\alpha = .05$), showed that the COART-values for /u/ were significantly higher than for /a,i/. Other ANOVAs made clear that significant between-context differences in coarticulation were present only in /a/ and /u/, but most in /u/. The smallest COART-values and between-context differences were observed for /i/.

SPEAKER VARIABILITY IN COART

In the previous section we encountered significant differences in coarticulation between vowel contexts of /a/ and, especially, /u/. We may ask whether the same pattern of between-context differences is observed for all speakers, or whether the pattern is speaker-dependent. If the latter is the case, we may conclude that coarticulation is speaker-specific.

A set of ANOVAs was carried out to answer this question. The ANOVAs were performed on the COART-values of each vowel /a,i,u/, separately. COART was computed for individual speakers, contexts and replications, averaging over formants:

$$COART(s,c,r) = \frac{1}{3}\sum_{i=1}^{3} COART(s,c,r,i).$$

Factors Speaker (fifteen levels) and Context (eight levels) were crossed in the ANOVAs. The interaction CxS proved to be significant ($p < .001$) for /a/ ($F_{98,1080} = 3.21$) and /u/ ($F_{98,1080} = 8.28$), but not for /i/ ($F_{98,1080} = 1.64$). This demonstrates that, indeed, the pattern of coarticulation over the contexts is speaker-dependent for /a,u/, but most for /u/. The speaker variability in the coarticulation of /u/ is shown in figure 1.

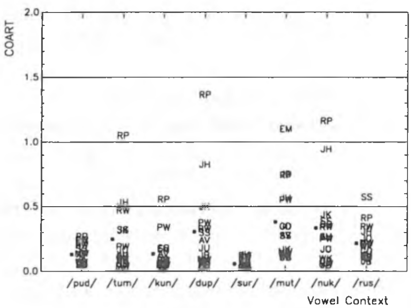

Figure 1: The speaker distribution of COART for each context of /u/. The speaker means are denoted by the speakers' initials; the context's mean is denoted by a black circle, slightly shifted to the left.

The figure illustrates that, indeed, speakers do not coarticulate uniformly. The most salient observation is that the mean COART-values in contexts with alveolars in C_1-position (/nuk/, /dup/ and /tum/) are pushed up due to the behaviour of two speakers: JH and, in particular, RP. Nonetheless, the interaction CxS remains significant if the data of these two speakers are removed from the ANOVA for /u/ ($F_{84,936} = 8.34$, $p < .001$).

SPEAKER IDENTIFICATION USING COART

Our results have indicated that coarticulation in the vowels /a/ and /u/ of our data set (as expressed by COART) is speaker-specific. This suggests that the amount of coarticulation may be used for speaker identification. The question that we tested was: do speaker identification scores improve if COART is used as an additional parameter to F_{1-3}? This is, of course, only a sensible question if COART is not highly correlated to (one of) the formants. The highest correlation observed between COART and a formant is the one between COART and F_2 of /u/; it was $r = 0.55$ ($n = 1200$), which is rather low. This makes it interesting to evaluate the question.

Speaker identification percentages were acquired by utilising the classification option of Linear Discriminant Analysis (LDA) and by introducing the 15 speakers as the groups to discriminate. For the present purpose, LDAs were carried out (a) for separate contexts of each vowel, and (b) for the pooled contexts of each vowel. In condition (a) there were (15 speakers $\cdot$ 10 replications =) 150 cases for each LDA; in condition (b) this number was multiplied by 8 (contexts), yielding 1200 cases. The identification percentages were based on three functions both for the LDAs on F_{1-3} and for the LDAs on F_{1-3} combined with COART. In this manner the analysis results were kept compatible. The results are presented in table 2.

Table 2: Percentages for correct speaker identification of the three vowels /a,i,u/. See text.

Condition	Vowel	F_{1-3}	COART + F_{1-3}
Pooled contexts	/a/	48.50	49.42
	/i/	43.83	42.33
	/u/	32.33	40.17
Split contexts	/a/	59.33	59.92
	/i/	62.67	66.58
	/u/	59.84	60.33

The differences between the two analysis settings (F_{1-3} vs COART $+F_{1-3}$) are clearly not substantial. The only exception is /u/ in the pooled contexts condition, where the improvement is about 8%. Thus, we find that the COART-index was not found a useful cue in speaker identification for most analysis settings.

GENERAL DISCUSSION

In this study we examined the speaker variability in the coarticulation of /a,i,u/. The amount of coarticulation in a vowel context was quantified using a score-model based measure COART.

We concentrated first on the coarticulation effects in the vowel contexts as such. It was observed that the effect of coarticulation upon /u/ was much stronger than upon /a/ and /i/.

In the next section we looked at the speaker variability in the observed coarticulation phenomena. It was found that the coarticulation in precisely the vowels that showed significant differences between contexts, proved to be speaker-specific as well (i.e. the vowels /a/ and especially /u/).

Guided by the finding that COART in /a,u/ had turned out to be speaker-specific, we tested if COART is a useful additional parameter for automatic speaker identification. Our results quite convincingly indicate that the COART-index is not a valuable coefficient for this task. Similar findings have been reported for /l/ and /r/ by Nolan (1983) [4]. As yet high speaker identification scores for a coarticulation measure have been presented only by Su, Li & Fu (1974) [3] for /m/. But also in their paper there is no proof that the coarticulation measure performs better than or just as good as simple spectral coefficients of /m/.

It is evident that the experimental setting described in this paper deviates considerably from the conditions normally encountered in (automatic) speaker recognition. There, the setting will be less formal and the recording background and transmission channels more noisy. Moreover, automatic speaker recognition nowadays operates increasingly more on sentence material and less and less on isolated words. Hence, stronger coarticulation and probably more between-speaker variability in coarticulation can be expected in such more complex speech data.

REFERENCES

[1] D.P. Kuehn and K.L. Moll (1976), "A cineradiographic study of VC and CV articulatory velocities", *J. of Phonetics*, Vol. 4, pp. 303-320.

[2] K. Johnson, P. Ladefoged and M. Lindau (1993), "Individual differences in vowel production", *JASA*, Vol. 94, pp. 701-714.

[3] L.-S. Su, K.P. Li and K.S. Fu (1974), "Identification of speakers by use of nasal coarticulation", *JASA*, Vol. 56, pp. 1867-1882.

[4] F.J. Nolan (1983), *The phonetic bases of speaker recognition* Cambridge: Cambridge University Press.

[5] L.C.W. Pols, H.R.C. Tromp and R. Plomp (1973), "Frequency analysis of Dutch vowels from 50 male speakers", *JASA*, Vol. 53, pp. 1093-1101.