ELSEVIER

Speech Communication 18 (1996) 113–130

SPEECH COMMUNICATION

Speaker variability in the coarticulation of /a,i,u/

H. van den Heuvel *, B. Cranen, T. Rietveld

Dept. of Language and Speech, University of Nijmegen, P.O. Box 9103, 6500 HD Nijmegen, The Netherlands

Received 9 November 1994; revised 31 July 1995

Abstract

Speaker variability in the coarticulation of the vowels /a,i,u/ was investigated in /C_1VC_2ə/ pseudo-words, containing the consonants /p,t,k,d,s,m,n,r/. These words were read out in isolation by fifteen male speakers of Dutch. The formants F_{1-3} (in Bark) were extracted from the steady-state of each vowel /a,i,u/. Coarticulation in each of 1200 realisations per vowel was measured in F_{1-3} as a function of consonantal context, using a score-model based measure called COART. The largest amount of coarticulation was found in /u/ where nasals and alveolars in C_1-position had the largest effect on the formant positions, especially on F_2. Coarticulation in /a,u/ proved to be speaker-specific. For these vowels the speaker variability of COART in a context was larger, generally, if COART itself was larger. Studied in a speaker identification task, finally, COART improved identification results only when three conditions were combined: (a) if COART was used as an additional parameter to F_{1-3}; (b) if the COART-values for the vowel were high; (c) if all vowel contexts were pooled in the analysis. The two main conclusions from this study are that coarticulation cannot be investigated speaker-independently and that COART *can* be contributive to speaker identification, but only in very restricted conditions.

Zusammenfassung

Die Sprechervariabilität in der Koartikulation der Vokale /a,i,u/ wurde in /C_1VC_2ə/ Pseudowörtern untersucht, in denen die Konsonanten /p,t,k,d,s,m,n,r/ enthalten waren. Die Wörter wurden von fünfzehn männlichen Sprechern des Niederländischen verlesen. Aus der Mitte jedes Vokals /a,i,u/ wurden die Formanten F_{1-3} (in Barks) extrahiert. Die Koartikulation in F_{1-3} in jeder der 1200 Vokalrealisierungen wurde mithilfe eines modellbasierten Maßes, COART genannt, gemessen. Die stärkste Koartikulation wurde in /u/ aufgefunden. Dabei hatten Nasale und Alveolare im Wortanlaut den größten Effekt auf die Formantlagen, besonders von F_2. Die Koartikulation in /a,u/ erwies sich als sprecherspezifisch. Für diese Vokale war die Sprechervariabilität von COART gemeinhin größer, wenn COART selbst größer war. Schließlich wurde COART in einem Sprecheridentifikationsverfahren überprüft. COART erbrachte dabei nur unter drei kombinierten Voraussetzungen bessere Erkennungsergebnisse: (a) wenn COART den Formanten F_{1-3} hinzugefügt wurde; (b) wenn die COART-Werte des jeweiligen Vokals hoch waren; (c) wenn alle Vokalumgebungen in der Analyse einbezogen wurden. Die zwei wichtigsten Schlußfolgerungen dieser Studie sind, daß Koartikulation nicht sprecherunabhängig untersucht werden kann und daß COART der Sprecheridentifikation behilflich sein *kann*, aber nur in sehr eingeschränkten Bedingungen.

* Corresponding author. E-mail: heuvel@let.kun.nl.

SSDI 0167-6393(95)00039-9

Résumé

La variabilité interlocuteurs dans la coarticulation des voyelles /a,i,u/ a été examinée dans des mots artificiels contenant les consonnes /p,t,k,d,s,m,n,r/. Ces mots ont été lus en isolation par 15 locuteurs néerlandais (m). Les formants F_{1-3} (en Barks) ont été extraits de la partie stable de chaque voyelle /a,i,u/. La coarticulation dans chacune des 1200 réalisations des voyelles a été mesurée en F_{1-3} en fonction du contexte consonantique, par une mesure statistique, nommée COART. L'effect de coarticulation le plus grand a été trouvé sur /u/: les consonnes nasales et alvéolaires en position C_1 y ont l'effect le plus grand sur les formants, plus spécifiquement sur F_2. La coarticulation sur /a,u/ s'avère être spécifique du locuteur. Pour ces voyelles, la variabilité par locuteur de la mesure COART était, généralement, la plus grande pour des valeurs de COART plus grandes. Dans le contexte d'une tâche d'identification de locuteurs, COART n'améliore les résultats d'identification que si trois conditions sont combinées: (a) COART est additionné aux mesures des F_{1-3}; (b) les valeurs de COART pour la voyelle sont grandes; (c) tous les contextes de voyelles sont combinés dans une seule analyse. Les deux conclusions de cette expérience sont que le phénomène de coarticulation ne peut pas être étudié indépendamment de la variabilité interlocuteurs, et que COART *peut* contribuer à l'identification de locuteurs, mais seulement dans des conditions très restreintes.

Keywords: Coarticulation; Speaker variability; Speaker identification; Vowel acoustics

1. Introduction

An important source of variation in the realisation of vowel spectra is given by the vowel's consonantal context. The influence of a segment's context upon its realisation is generally referred to as coarticulation. In this report we deal with the variation in the spectra of the Dutch vowels /a,i,u/ that result from coarticulation. We examine which consonantal contexts exert the largest influence on vowel spectra (Section 3). We will in greater detail focus on speaker variability observed in the coarticulation of these vowels (Section 4) and investigate if this variability can be used to help identify speakers (Section 5). We note in advance that if we speak in the following of "spectral coarticulation", then we in fact refer to the spectral effects of coarticulation phenomena taking place in the articulatory domain.

A large variety of models has been developed to explain coarticulation, but none of them seems to be able to explain all observations found. Critical reviews of models of coarticulation are given by Kent and Minifie (1977), Fowler (1980), Sharf and Ohde (1981) and Tokuma (1993). Each model has to cope with the problem of how vowel realisations come to differ, in varying contexts, from vowels produced in isolation. Here, we adopt the view that some canonical or ideal form is observed when vowels are uttered in isolation, or in a compatible context like /hVd/ (Stevens and House, 1963; Daniloff and Hammarberg, 1973; Pols et al., 1973). The deviation of a vowel realisation from its canonical form may be brought about by e.g. stress, word class, speech rate and consonantal context, cf. (Van Bergem, 1993). In this paper we deal only with the last factor: consonantal context. The effect of this factor is generally called "coarticulation". A more general term for the effect of all factors is "vowel reduction". The terms "target undershoot", "spectral undershoot" or simply "undershoot" are used to describe the vowel formant shift in all types of vowel reduction, including coarticulation (Stevens et al., 1966; Van Son, 1993).

Speaker variability in articulation has not been investigated on a large scale, because invariant aspects of speech production are often considered more interesting. Nonetheless, a few studies have dealt with this topic: Kuehn and Moll (1976) examined the velocity and the displacement of tongue movements as a function of speech tempo. They observed appreciable differences between speakers in the control of the two variables. Similar findings of speaker dependencies in patterns of upper lip and jaw movements were reported by Shaiman et al. (1995). Johnson et al. (1993), who studied x-ray microbeam pellet trajectories during the production of vowels by five speakers, found highly speaker-dependent patterns of inter-articulator coordination (for lip, jaw and tongue movements). They concluded that inter-speaker variation was too large to uphold the hypothesis that

articulatory defined phonetic features are of a universal nature. In the light of these studies it can be hypothesised that acoustic aspects of coarticulation may also exhibit substantial inter-speaker variability.

A confirmation for this hypothesis has been found for some phonemes in the field of (automatic) speaker identification. Su et al. (1974) 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). In his study the coarticulation in /l/ appeared to be more speaker-specific than 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. To study the speaker variability in vowel coarticulation, then, is the objective of the present investigation.

We measured the first three formants (F_{1-3}) of the nucleus vowels in 24 /C_1VC_2ə/ pseudo-words, which were read ten times each in isolation by fifteen male speakers of Dutch. They contained the vowels /a,i,u/ in the context of /p,t,k,d,s,m,n,r/, which appeared in C_1-position and in C_2-position (but not in all combinations). More detailed information on the data used will be given in Section 2.

To illustrate the conceptual problems that arise if one tries to investigate the relation between coarticulation and speaker specificity, let us first consider evidence showing that a relation between the two actually exists. A Linear Discriminant Analysis (LDA) was performed on each of the vowels /a,i,u/ in our data to yield speaker identification percentages on the basis of the formants F_{1-3}. The LDA was instructed to discriminate between the fifteen speakers on the basis of three functions (the maximum possible). Percentages of correct speaker identification were then obtained by classification of the data. Two types of analyses were carried out. The first contained the data of each vowel combined over all contexts, yielding one LDA per vowel (pooled contexts); the second contained the data of each vowel sorted by context (yielding eight LDAs per vowel) whereafter the eight identification percentages obtained were averaged (split contexts). The corresponding results are displayed in Table 1, which shows that the split contexts analysis leads to considerably higher speaker identification percentages than the pooled contexts analysis. This means that the vowel's consonantal context, and therefore coarticulation, may play a significant role in speaker identification by computer, cf. also (Bonastre and Meloni, 1994).

Table 1
Percentages of correct speaker identification for the three vowels /a,i,u/ in two conditions (see text for further details)

Condition	Vowel		
	/a/	/i/	/u/
Pooled contexts	48.50	43.83	32.33
Split contexts	59.33	62.67	59.84

Does this finding imply that coarticulation as such is speaker-specific? The answer is no. Consider Fig. 1, which shows a hypothetical (and highly simplified) picture of a possible distribution of F_1 and F_2 for /u/ produced by two speakers, S1 and S2. The formant values of speaker S1 are in the left-hand top circle; those of speaker S2 are in the right-hand bottom circle. For both speakers the realisations of context /du/ are found in the left upper part of the circles, and the realisations of context /ku/ are located in the right lower part of the circles. If we perform speaker identification on the pooled contexts, the /ku/-realisations of speaker S1 are confused with the /du/-realisations of speaker S2. However, if we conduct speaker identification on the /du/- and the /ku/-realisations separately, then

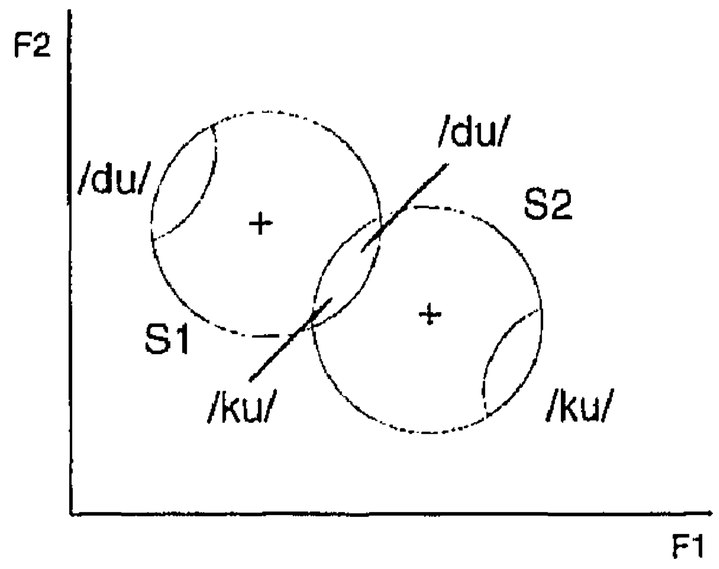

Fig. 1. Illustration of a hypothetical relation between coarticulation and speaker variability. The x-axis and the y-axis denote the first and the second formant respectively. The cross in the centre of each circle indicates the average values for /u/ tokens produced by two speakers, S1 and S2. See the text for a further explanation.

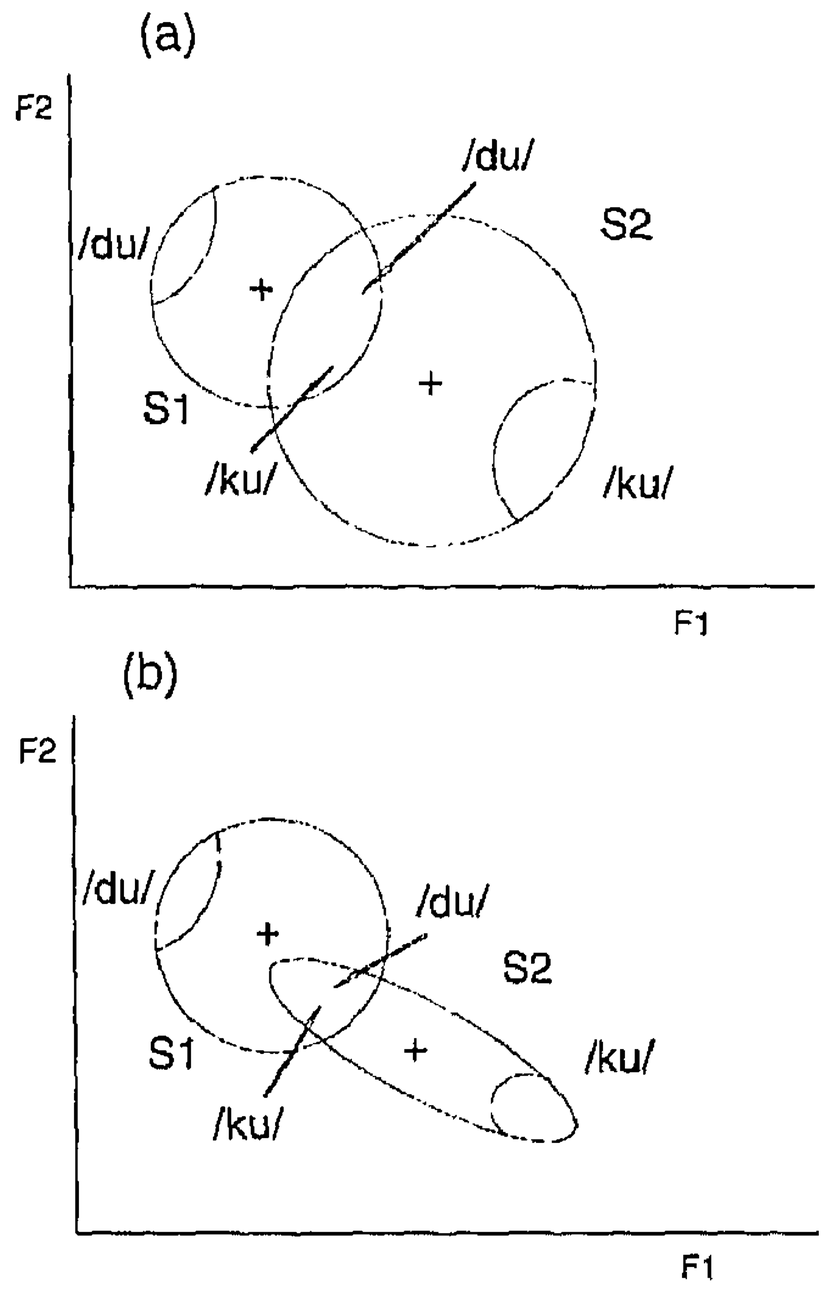

Fig. 2. As Fig. 1, but coarticulation is speaker-specific since the circles differ in their diameters (a), or in their shapes (b).

there is no overlap, and better identification results are obtained. Thus, Fig. 1 adequately explains the results of Table 1. Note, however, that in this figure coarticulation is *not* speaker-specific! If it were, then the circles' diameters would be different (as in Fig. 2(a)) or they would have different shapes (for instance, one of them would be an ellipse, as in Fig. 2(b)). The only difference between S1 and S2 in Fig. 1 concerns the centres of their formant distributions. These, however, do not bear a relation to coarticulation, but are determined by general characteristics of speakers, in particular by their vocal tract lengths.

Thus, we have identified two possible reasons why the consonantal context (and therefore coarticulation) may positively affect speaker specificity.

1. Coarticulation is speaker-specific, or
2. Speakers do not differ in coarticulation, but in some other characteristic (such as vocal tract length). Speaker differences in this other characteristic emerge when speaker specificity is investigated for separate contexts. In this case coarticulation is only a catalyst, but not itself a source of speaker specificity.

The main aims of this paper are to ascertain whether coarticulation in vowels is speaker-specific, and to investigate if it can be beneficially used for speaker identification. (To measure the degree of coarticulation, we will use a self-defined index called COART, which will be clarified in Section 3.1.) The following specific questions will be addressed:

1. Do the consonantal contexts of the vowels examined, viz. (/a,i,u/), cause different degrees of coarticulation, as measured with the COART-index? (→ Section 3)
2. Is the amount of coarticulation observed in the vowels speaker-specific? And if so, is there a relation between the amount of coarticulation in a vowel (as quantified by COART) and the speaker variability in this coarticulation? (→ Section 4)
3. Is the COART-index useful as a parameter in speaker identification? (→ Section 5)

2. Speakers and speech data

We opted for a rather restricted data set, to keep the experiment within practical limits and to have control over the number of factors that may effect the vowel formants. In larger, more complex data bases it is difficult to identify the exact source of the variation because the variation may be the result of many (interacting) factors. As mentioned earlier, speaker variability in the coarticulation of vowels has seldom been investigated. Given this situation our restricted data set can be considered as a useful starting point. The data set used consisted of 24 /C_1VC_2ə/ (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/. The consonants selected are representatives of different places of articulation (cf. /p,t,k/) and manners of articulation (cf. /t,s,n,r/). /d/ represented the voiced plosives. It was also included because of its noted coarticulatory effect on /u/ (Stevens and House, 1963; Stevens et al., 1966; Ohde and Sharf, 1975; Schouten and Pols, 1979; Tokuma, 1993). Since we also wanted to analyse the effect of all consonants in C_2-position, and since in Dutch word-final obstruents are devoiced (/d/ be-

comes /t/), disyllabic words were used to prevent the devoicing of /d/. Each consonant occurred once in C_1-position and once in C_2-position. This was done to reduce the data set to manageable proportions (a completely balanced consonant set would have required well over three hours recording time per speaker). As can be seen from Table 2, it was safeguarded that not the same consonant combinations were used for more vowels: if we find that the /a/ in /dakə/ is as highly coarticulated, as the /i/ in /disə/ and the /u/ in /dupə/, we can infer with some certainty that the coarticulation was brought about by C_1 (/d/). Such a conclusion could not be drawn if C_2 did not vary for the three vowels, as in /dakə/, /dikə/ and /dukə/; in such a case both C_1 and C_2 might have caused the effect. The test words had an open syllable boundary between the nucleus vowel and the word-medial consonant. It can, hence, be expected that the coarticulatory effect of C_1 on the vowel will be larger than that of C_2, cf. (Rietveld and Frauenfelder, 1987). Some of the disyllables existed as Dutch words: 'mieke' is a Dutch given name; 'koene' is a declined adjective in Dutch, and 'kade' and 'mare' are Dutch nouns. All but five (/pinə/, /rusə/, /sapə/, /surə/ and /tumə) can be changed into real Dutch words by adding a final /n/, /r/ or /l/. The 24 words were printed in a random order on ten 30-word word lists, which were read out by each speaker in one recording session. The initial three words served as fillers, as did the final three, yielding 240 words (24 words × 10 repetitions) per speaker, and 3600 words in all.

The subjects were all native Dutch males, between 20 and 30 years of age. They were instructed to read the words in a relaxed speech tempo. The speakers whose data were used were selected from a larger group by five speech therapists who screened and approved them with respect to (1) pronunciation of Standard Dutch, (2) naturalness of production and (3) absence of voice and articulation disorders. For the selected speakers there are no indications that the real words were pronounced more naturally than the pseudo-words.

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 segmented into a steady-state portion flanked by transitions. Segmentation boundaries were derived from the sampled waveform, the RMS intensity curve and the four lowest formants as fitted by an LPC-analysis. This is illustrated in Fig. 3. The segment boundaries were proposed by a DTW-based segmentation algorithm and, if needed, corrected manually. A strict segmentation protocol was used to carry out the labellings. Performance of the automatic segmentation algorithm was good, but not good enough to accept the labellings of words with a nasal or /r/ without subsequent visual inspection (and manual correction, if needed). The formants F_{1-3} were extracted from each nucleus vowel from the middle frame of the steady-state by means of an LPC-analysis (pitch-asynchronous autocorrelation method). The analysis was conducted using a Hamming window of 25 ms and a frame shift of 5 ms. The prediction order M was 20. From the filter coefficients that were thus obtained for each analysis frame, formant resonance frequencies (in Hz) were computed by root-solving. For 24 vowel tokens of each of two speakers the LPC-spectra taken from the vowel mid-

Table 2
The /C_1VC_2ə/ pseudo-words used in the experiment, both in phonemical and orthographical representation

V =	/a/		/i/		/u/	
C_1 = /p/	/pasə/	"pase"	/pinə/	"piene"	/pudə/	"poede"
C_1 = /t/	/tanə/	"tane"	/tirə/	"tiere"	/tumə/	"toeme"
C_1 = /k/	/kadə/	"kade"	/kimə/	"kieme"	/kunə/	"koene"
C_1 = /d/	/dakə/	"dake"	/disə/	"diesse"	/dupə/	"doepe"
C_1 = /s/	/sapə/	"sape"	/sidə/	"cide"	/surə/	"soere"
C_1 = /m/	/marə/	"mare"	/mikə/	"mieke"	/mutə/	"moete"
C_1 = /n/	/natə/	"nate"	/nipə/	"niepe"	/nukə/	"noeke"
C_1 = /r/	/ramə/	"rame"	/ritə/	"riete"	/rusə/	"roesse"

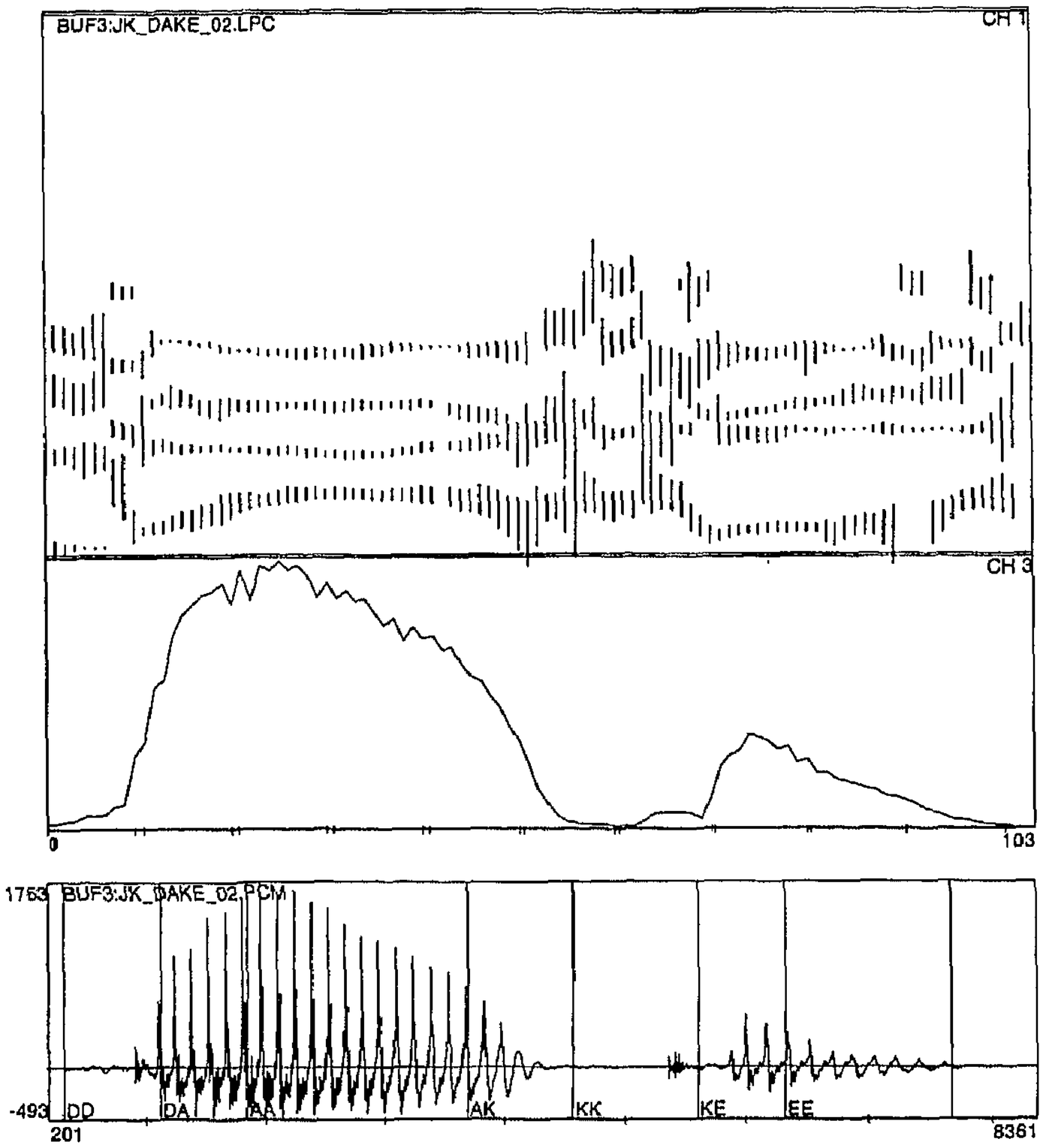

Fig. 3. Segmentation of a token of the word /dakə/. The top panel displays the first four formants F_{1-4} (range 0–8 kHz); the middle panel gives the intensity curve (frame numbers along the horizontal axis), and the bottom panel displays the corresponding waveform (sample numbers along the horizontal axis). See text for further details.

dle frame were visually compared to corresponding (ninth order) FFT-spectra to evaluate the quality of the LPC-spectra. Close matches in the locations of F_{1-3} were observed between the two types of vowel spectra.

The formant positions in all selected vowel middle frames were checked by hand. If a formant from the middle frame of a vowel segment was an outlier (compared to the other nine realisations available), then the vowel formant track was inspected visually. The middle frame of the vowel deviated from the neighbouring frames within the vowel segment in 6% of the instances. For these tokens a neighbouring frame was selected. This correction reduced the proportion of outliers to less than 1.5%. To prevent these remaining outliers from influencing the results, the following precaution was taken. The particular formant that caused a frame to be an outlier was replaced by the mean value of that formant for the vowel (leaving all other formants unchanged).

The resulting mean formant values of F_{1-3} (in Hz) and their standard deviations are depicted for the vowels in all contexts in Fig. 4. The formant frequencies were converted into Barks using the equation given by Hermansky (1990, p. 1739):

$$b = 6 \ln\left(f/600 + \left[(f/600)^2 + 1 \right]^{0.5} \right),$$

where f is the formant frequency in Hertz and b is the converted frequency in Barks. The reason for this conversion is that on a *linear* frequency scale (Hz) variations in the higher formants F_2 and F_3 would

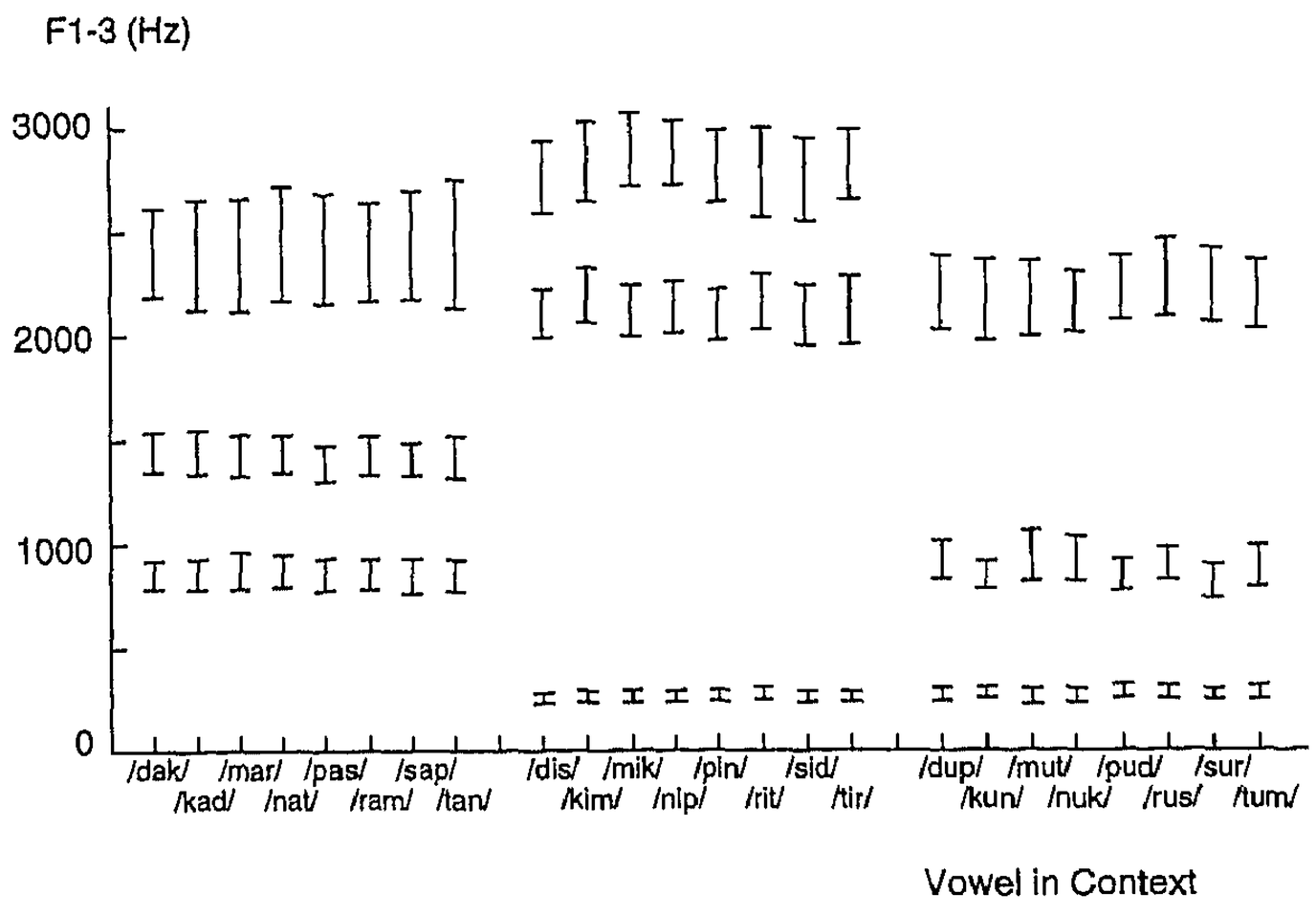

Fig. 4. Averages and standard deviations of F_{1-3} in the middle frames of /a,i,u/. On the x-axis each vowel is shown in its consonantal context. The averages are based on 15 speakers and 10 replications, which makes 150 observations.

obtain a dominating weight. We selected F_{1-3} only, because higher formants could not be reliably established.

3. The score model approach to spectral coarticulation

3.1. Introduction and method

Several methods have been suggested to measure spectral coarticulation. A simple one is to calculate the difference between the observed values of a formant (generally F_2) to a reference value for the same formant (Ohde and Sharf, 1975; Whalen, 1990). A more sophisticated method was described by Van Bergem (1993). He computed euclidean distances between two vowel realisations on the basis of mel-scaled formants (F_1 and F_2) to determine acoustic vowel reduction. The computation of our COART-index is very similar to the latter approach, but is designed specifically for the computation of coarticulation. We use a score model in which the coarticulation in a formant i in a specific context c in replication r as realised by speaker s is given by

$$\mathrm{COART}(s,c,r,i) = \left(f_{scr}(i) - f_{s(\mathrm{ref}).}(i)\right)^2, \qquad (1)$$

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

In a strict score model based approach the reference should be the vowel centroid (i.e. the average formant position for a speaker). However, the use of a reference *other* than the vowel centroid is rather compelling for spectral coarticulation. Stevens and House (1963) reported that formant values averaged over a set of contexts deviate greatly from formant values that are obtained from vowels spoken in isolation or in a /hVd/-context. Since in both latter contexts, the formant target values of the vowel are thought to be realised, we should use one of these contexts as reference context and, obviously, not the phoneme centroids. However, our speakers did not produce /a,i,u/ in isolation nor in a /hVd/-context. In order to obtain good estimates of these formant values for our experiment, we opted for the following solution. The vowel formant frequencies as published by Pols et al. (1973, p. 1094) for 50 male speakers of Dutch for /a,i,u/ in a /hVt/-context (which is compatible to isolated vowels) were taken as initial references. We will refer to these values as the PTP-references. Using the PTP-references as $f_{s(\mathrm{ref}).}(i)$ we computed the coarticulation for

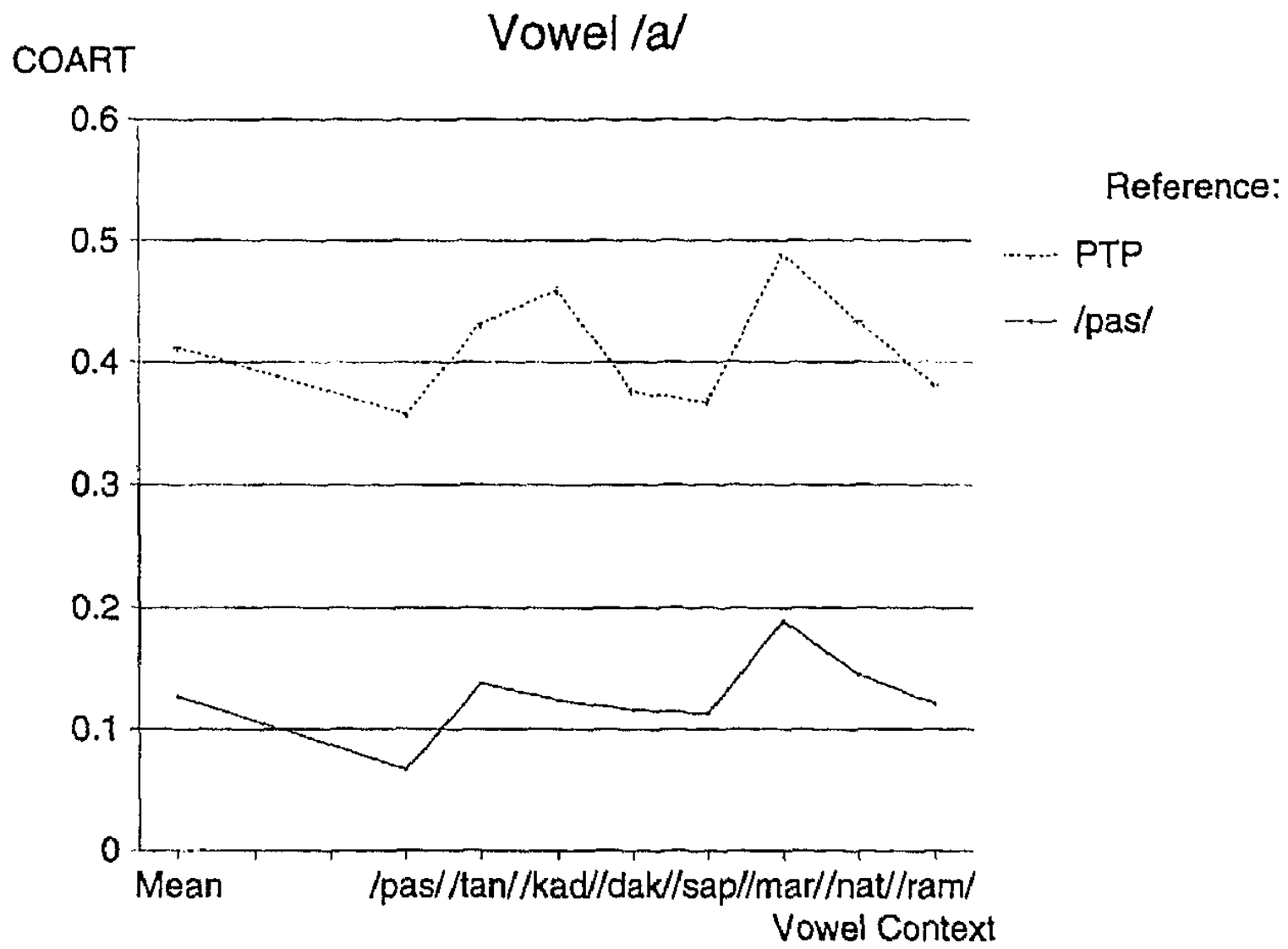

Fig. 5. COART-values for the vowel /a/ both for the PTP-reference and the best matching context being the /a/ from /pasə/. The contexts shown in the figure equal the corresponding words, with the final schwa being omitted. See text for further explanation.

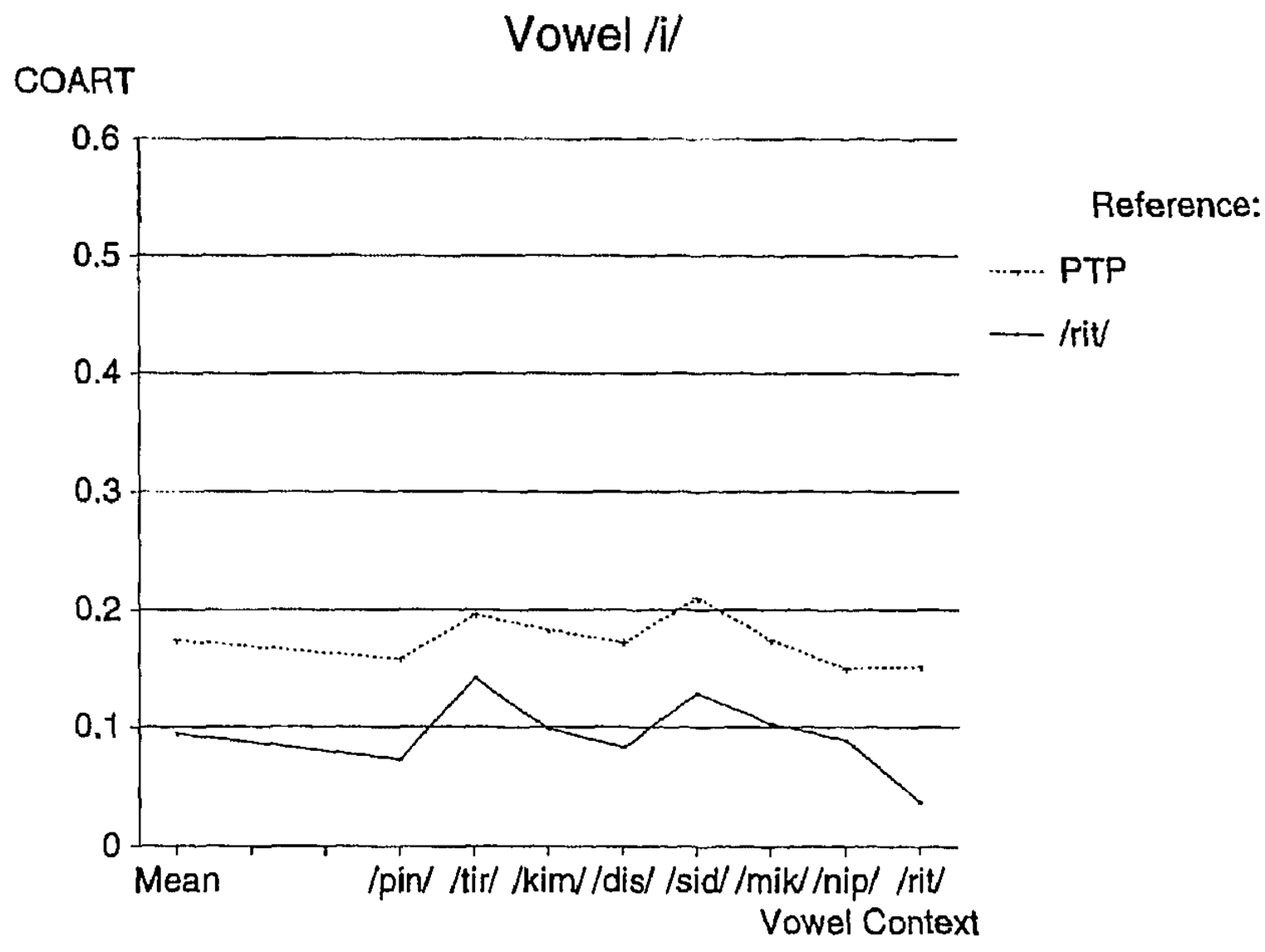

Fig. 6. COART-values for the vowel /i/ both for the PTP-reference and the best matching context being the /i/ from /ritə/. See text for further explanation.

the vowels in each consonantal context from Eq. (1). To obtain one COART-value for each consonantal context we averaged over speakers, replications and formants, according to

$$\mathrm{COART}(c) = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{R}\sum_{r=1}^{R}\frac{1}{I}\sum_{i=1}^{I}\mathrm{COART}(s,c,r,i). \qquad (2)$$

The results are shown by the dashed lines in Figs. 5–7 for /a,i,u/, respectively. Since we required previously that our reference be speaker-specific, the PTP-references could not establish the ultimate vowel references. Therefore, in a next step, we used $f_{sc.}(i)$ (i.e. the speaker averages of F_{1-3} in each context) as the reference. Each single $f_{sc.}(i)$ served once as the reference for computing the average COART-value for all consonantal contexts. We selected as the ultimate reference for each vowel /a,i,u/ the reference context which yielded the best match to the PTP-reference, i.e. the context that resulted in a between-context pattern of COART-values that most closely corresponded to the pattern found for the PTP-references.

3.2. Results

The vowels in the following words were found to establish the closest match to the PTP-references: /pasə/ for /a/; /ritə/ for /i/; /surə/ for /u/. The COART-values obtained by using these contexts as references are displayed in Figs. 5–7 for /a,i,u/, respectively. For comparison, the COART-values obtained by using the PTP-reference are also shown in the figures. The connections between the data points in the figures have been added to facilitate visualisation; they are not intended to suggest other relationships.

A number of observations can be made from the figures. The first is that the PTP-references yield considerably higher COART-values than the contextual references from our own data. This must be attributed to the fact that the PTP-references involved other speakers, whereas the best matching reference stemmed from our own speakers.

We may expect that the context that has the largest vowel duration is bound to establish the closest match to the PTP-reference. Since the inertia of the articulators plays a smaller role in the middle

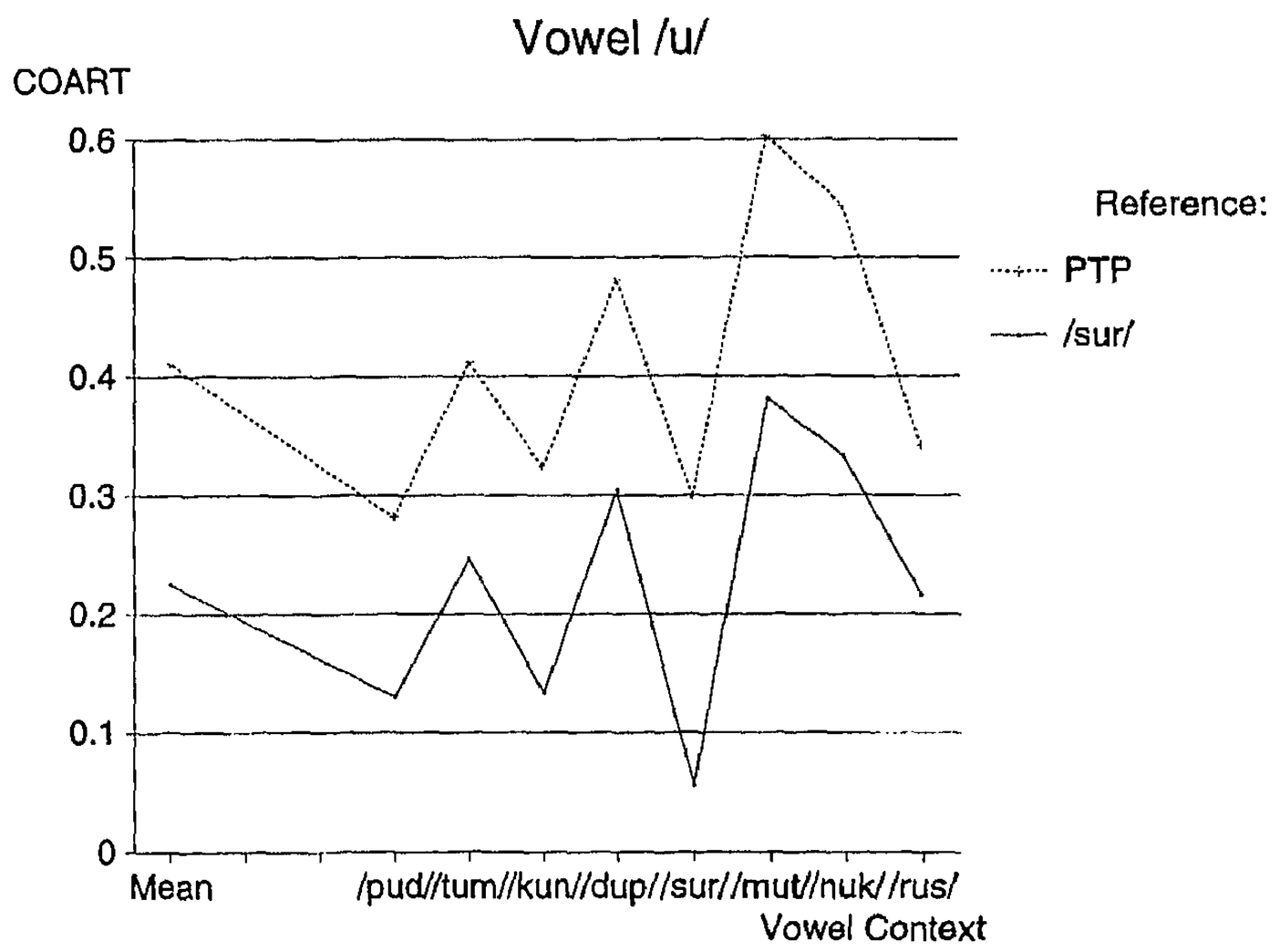

Fig. 7. COART-values for the vowel /u/ both for the PTP-reference and the best matching context being the /u/ from /surɔ/. See text for further explantion.

frame of a long than of a short vowel, it is reasonable to assume that the target values can be optimally attained in long vowels. This expectation was only partially corroborated by our data. For /a/ and /i/ other reference contexts emerged than those having the longest vowel duration (being /marə/ and /tirə/, respectively). Since for /a/ and /i/ the differences in coarticulation between contexts appeared to be minor, it is arbitrary which context comes out as the best match. Therefore, we should not be surprised that these references did not satisfy our expectation. For /u/, however, where the COART-values and the between-context differences were much more substantial, our expectation was entirely confirmed. /surə/, indeed, contains the longest realisation of /u/ [1]. A further observation is that, not surprisingly, the COART-values obtained for the vowels in the selected reference contexts (i.e. /a/ from /pasə/, /i/ from /ritə/ and /u/ from /surə/) were always the lowest. Evidently, the realisations in these contexts have the shortest distance to the reference, because the reference is the mean of the same context. As noted, coarticulation was largest in /u/-contexts. This was confirmed by an ANOVA on the combined (3 vowels × 8 contexts =) 24 COART-values of the three vowels. The factor Vowel was significant in this analysis ($F_{2,21} = 7.47$, $p < 0.005$). A subsequent Tukey HSD post-hoc comparison ($\alpha = 0.05$) showed that the COART-values found for /u/ were significantly higher than those for /a/ and /i/. Furthermore, for /u/ the largest between-context differences in coarticulation were observed, as can be seen from the figures. The smallest COART-values and between-context differences were found for /i/. The next step was to determine in which contexts coarticulation was largest. An ANOVA was carried out on the COART-values of each of the vowels /a,i,u/. COART was computed for individual speakers, contexts and replications as

[1] It could be argued against this account that, in Dutch, vowels are coloured due to postvocalic /r/. However, for /a,i,u/ this colouring takes place in the last part of the vowel, manifesting itself as a shift towards schwa. It has been verified that the vowel middle frames were not taken from this schwa-like tail, but from the long stable vowel part preceding it.

Table 3
Subsets of /u/ allophones not differing significantly in amount of coarticulation according to a Tukey HSD post-hoc comparison ($\alpha = 0.05$, HSD = 0.205). In the left column the COART-values of the vowel allophones. In the subset columns /u/ is shown with both its adjacent consonants

	Subset		
COART	1	2	3
0.056	/sur/		
0.130	/pud/	/pud/	
0.134	/kun/	/kun/	
0.216	/rus/	/rus/	/rus/
0.247	/tum/	/tum/	/tum/
0.305		/dup/	/dup/
0.333		/nuk/	/nuk/
0.381			/mut/

an average over the three formants, in accordance with the formula

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

Factors Speaker (fifteen levels) and Context (eight levels) were crossed in the ANOVAs, with ten values per cell. Speaker was considered random and Context fixed. Detailed results of the analyses are presented in Section 4.2 (Table 5). Context was significant only for /a/ and /u/ ($p < 0.05$). Tukey HSD tests ($\alpha = 0.05$) revealed that, for /a/, a significant difference (HSD = 0.086) existed only between /pasə/ and /marə/ (cf. Fig. 5). For /u/ the significant differences are listed in Table 3. From the table we infer that coarticulation in /mutə/, /nukə/ and /dupə/ was significantly stronger than in /surə/, and that the vowel coarticulation in /mutə/ was also significantly stronger than in /pudə/ and /kunə/. It appears that the largest coarticulation effects upon /u/ were due to nasal and alveolar C_1-consonants (see subset 3), with nasality being the more prominent of the two.

For /u/ we investigated which of the three formants showed most coarticulation. We calculated COART for individual formants by averaging COART over replications and speakers, using

$$\mathrm{COART}(c,i) = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{R}\sum_{r=1}^{R}\mathrm{COART}(s,c,r,i). \qquad (4)$$

Table 4
COART-values for individual formants of /u/-contexts. The second column lists the COART-values for the combined formants. SD: standard deviation (over contexts)

Context	COART			
	F_{1-3}	F_1	F_2	F_3
/sur/	0.056	0.019	0.105	0.043
/pud/	0.130	0.056	0.236	0.099
/kun/	0.134	0.040	0.255	0.106
/rus/	0.216	0.074	0.481	0.092
/tum/	0.247	0.059	0.603	0.078
/dup/	0.305	0.042	0.745	0.128
/nuk/	0.333	0.089	0.792	0.118
/mut/	0.381	0.119	0.936	0.087
Mean	0.225	0.062	0.519	0.094
SD	0.113	0.031	0.300	0.026

Table 4 displays the results of this computations.

It shows that coarticulation in /u/ was largely restricted to F_2: F_2 had the largest mean COART (0.519) of the three formants and also the largest spread over contexts (0.300). Furthermore, the rank order of contexts for F_2 is identical to the rank order obtained for F_{1-3} combined.

To summarise the answer to our first research question (see the end of Section 1), we may say that the effect of coarticulation in /u/ was considerably larger than in /a,i/, and that the coarticulation effect of nasal and alveolar word-initial consonants upon /u/ was significantly larger than of other consonants.

We will next examine the speaker specificity of the coarticulation effects observed.

4. Speaker variability in COART

4.1. Introduction and method

In the previous section we encountered significant differences in consonantal coarticulation across vowels. We may now ask whether the same pattern of between-context differences is observed for all fifteen speakers. If not, then we may conclude that coarticulation is speaker-specific.

In Section 4.2 we present the results of a set of ANOVAs, referred to earlier (Section 3.2), which were carried out to answer this question. The ANOVAs were performed on the COART-values of each vowel /a,i,u/. COART was computed for individual speakers, contexts and replications, in accordance with Eq. 3. Factors Speaker (fifteen levels) and Context (eight levels) were crossed in the ANOVAs. The factors Speaker and Context were crossed; each cell contained ten replications. Speaker was considered random and Context fixed. If Speaker and Context show a significant interaction, then the between-context pattern of COART is speaker-dependent and, consequently, coarticulation is speaker-specific.

We also examined the relation between COART and its speaker variability, to determine if speaker-related variation in COART was larger in those contexts where COART itself was larger. If a positive relationship is found, then the mean of COART in a context may be a good predictor of its speaker specificity. To explore this question we performed ANOVAs on the COART(s,c,r)-values for each context of /a,i,u/. These ANOVAs involved the (random) factor Speaker only and were conducted on data sets of (15 speakers × 10 replications =) 150 COART-values each. The strength of association of the factor Speaker was computed using ω^2:

$$\omega_S^2 = \sigma_{\text{speaker}}^2 / \Sigma\sigma^2 = \sigma_{\text{speaker}}^2 / \left(\sigma_{\text{speaker}}^2 + \sigma_{\text{error}}^2\right), \quad (5)$$

and was regarded as a measure of speaker variability of COART in the specific context. Next, the Pearson product-moment correlation between ω_S^2 and COART for the eight contexts per vowel was calculated.

Table 5
Results of the ANOVAs on COART for each vowel /a,i,u/. Factors were Context (C) and Speaker (S). ns: $p > 0.05$

Vowel	Factor	df_1, df_2	F	sign.
/a/	C	7,98	3.03	$p < 0.007$
	S	14,1080	11.06	$p < 0.001$
	C×S	98,1080	3.21	$p < 0.001$
/i/	C	7,98	1.64	ns
	S	14,1080	3.60	$p < 0.001$
	C×S	98,1080	1.18	ns
/u/	C	7,98	5.79	$p < 0.001$
	S	14,1080	52.61	$p < 0.001$
	C×S	98,1080	8.28	$p < 0.001$

4.2. Results

We will first turn to the ANOVAs on COART(s,c,r) that were carried out for each vowel /a,i,u/ to determine if coarticulation was speaker-specific. The results of these ANOVAs are summarised in Table 5.

The table shows a significant interaction between Context and Speaker for the vowels /a/ and /u/. Thus, it appears that the amount of coarticulation, as expressed by COART, is speaker-specific for /a/ and /u/. Interestingly, these are exactly the vowels that also showed significant differences between contexts. In Section 3.2 it was shown that the contextual differences were small for /a/, but larger for /u/. Therefore, we will focus on /u/ to assess the effect of speaker variability on the COART-values in more detail. This speaker variability is shown in Fig. 8 where the speaker distribution of COART for each context of /u/ is plotted.

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. A large part of the apicalisation of /u/ in the alveolar contexts is, therefore, attributable to these two speakers. (Nonetheless, the interaction C × S remains significant if the data of these two speakers are removed from the ANOVA for /u/: $F_{84,936} = 8.34$, $p < 0.001$).

Table 6
Correlation (r) between COART and its speaker variability expressed as ω_S^2. p: significance of r; n: number of observations (eight contexts per vowel)

Vowel	n	r	p
/a/	8	0.226	0.590
/i/	8	−0.479	0.229
/u/	8	0.690	0.058
/a,u/	16	0.730	0.001

So far we have said nothing about the relation between the magnitude of COART and its speaker variability. As explained earlier (Section 4.1), ω_S^2, being a measure of speaker variability in COART, was computed for each of these ANOVAs and correlated with the corresponding COART-value. The results are summarised in Table 6, which shows that for /a/ and /u/ the speaker specificity of COART

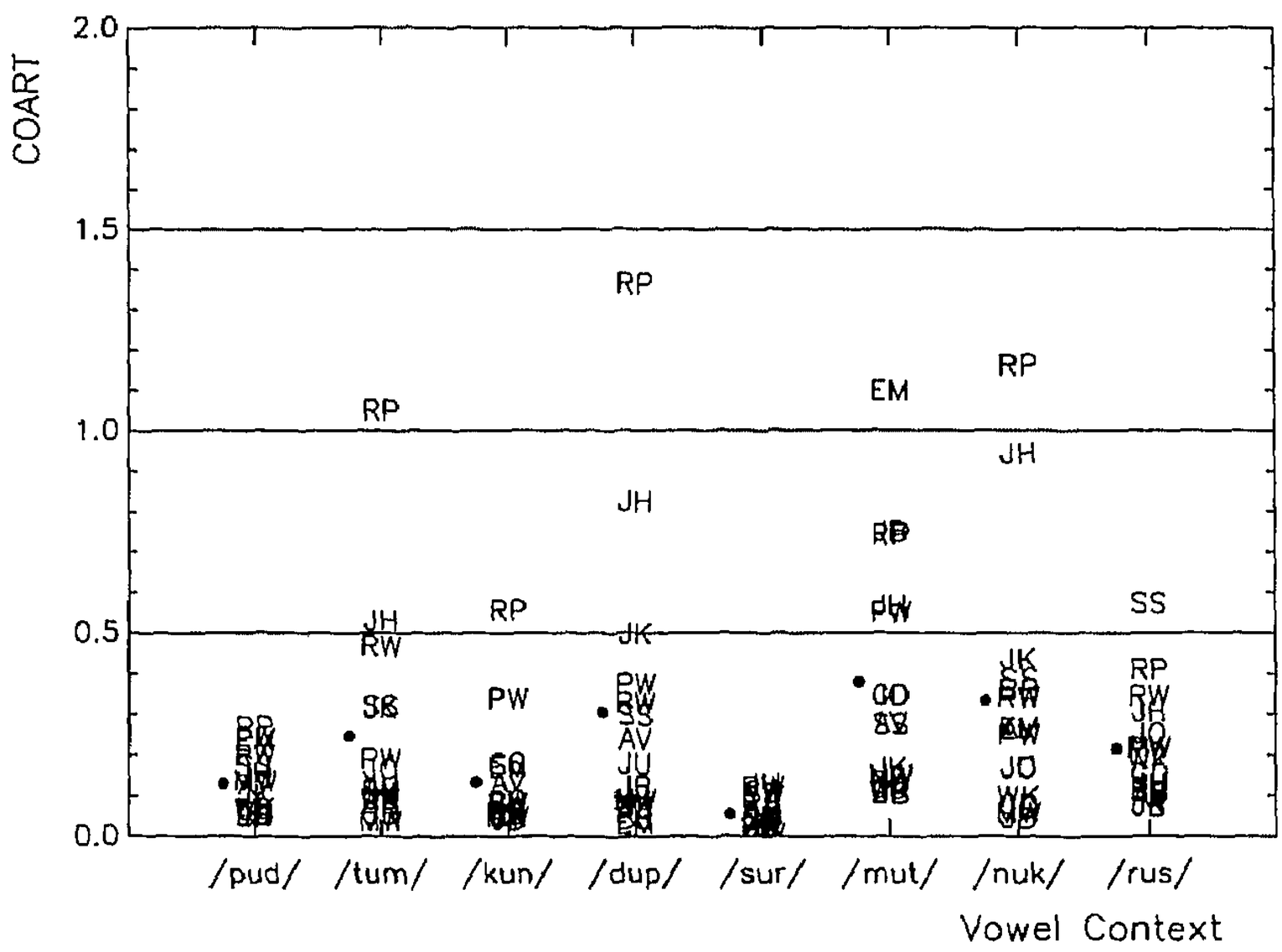

Fig. 8. The speaker distribution of COART for each context of /u/. The speaker means are denoted by the speakers' initials; the context's mean is denoted by a black circle, slightly shifted to the left. The contexts of (4) on the x-axis are displayed with both adjacent consonants.

is positively correlated to the mean value of COART in a context. This correlation is significant if the data for /a/ and /u/ are combined. (Significance here is also attained by the doubling of n to 16.) The correlation suggests that idiosyncratic variation in COART is relatively large if COART is high. An illustration of this can be seen in Fig. 8. For contexts with a low mean COART-value, like /surə/ and /pudə/, speaker variability is small, whereas for contexts with a high mean COART-value, like /mutə/, /nukə/ and /dupə/, speaker variability is large.

To conclude, our answer to research question 2 is that (1) the amount of coarticulation in contexts, as expressed by COART, is speaker-specific and (2) is more speaker-specific if the mean COART-value for a context is higher.

To answer our third research question we will examine if COART can be used to identify speakers.

5. Spectral coarticulation and speaker identification

5.1. Introduction

Our results indicate that coarticulation (as expressed by COART) is speaker-specific, which suggests that amount of coarticulation may be an effective and useful parameter for speaker identification. To collect additional evidence as to the speaker specificity of coarticulation, and to put our COART-index to the test, the COART-index was used as a speaker-discriminating tool in an actual speaker identification task. Our third research question will be divided into three subquestions.

1. Is it possible to identify speakers above chance level solely on the basis of their COART-index?
2. Are the identification scores obtained by using the COART-index comparable to the identification scores obtained by using F_{1-3}?
3. Do speaker identification scores improve if COART is used as an additional parameter to F_{1-3}?

Question 2 pertains to the issue if the (one-dimensional) COART-index is a better dimension for speaker identification than the original three-dimensional formant space. It might be that for speaker identification less favourable (because noisy) properties of the original speech signal are eliminated in the coarticulation index, which will then be a more effective predictor of speaker identity.

Question 3 examines whether it is useful to employ COART as an extra speaker-identifying tool if the formants F_{1-3} are already available. This is a sensible question only 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.

5.2. Method

Speaker identification percentages were obtained by utilising the classification option of Linear Discriminant Analysis (LDA). For the present purpose, LDAs were carried out (a) for split contexts of each vowel, and (b) for the pooled contexts of each vowel. In condition (a) there were (15 speakers × 10 replications =) 150 cases for each LDA; in condition (b) this number was multiplied by 8 (contexts), yielding 1200 cases.

To answer question 1, COART was used as the only predictor variable. Consequently, identification percentages were determined on the basis of one discriminant function only. The LDAs on the predictors F_{1-3} yield a maximum of three discriminant functions. To answer question 2, only the first function of these analyses was used (to keep the resulting identification percentages compatible to those obtained for COART). However, to answer the third question, 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 also these analysis results were kept compatible.

5.3. Results

Table 7 presents the speaker identification scores based on COART on the one hand and F_{1-3} on the other hand. Also the identification percentages for the individual formants F_1, F_2 and F_3 are shown. In the pooled contexts condition all eight contexts of a vowel were combined in one LDA. In the split

Table 7
Percentages for correct speaker identification of the three vowels /a,i,u/ using different predictors for speaker identity: COART, F_{1-3} combined, and each individual formant

Condition	Vowel	COART	F_{1-3}	F_1	F_2	F_3
Pooled contexts	/a/	10.50	18.17	18.08	16.50	19.83
	/i/	10.25	21.83	19.50	22.92	18.00
	/u/	11.25	20.00	20.25	8.33	19.75
Split contexts	/a/	13.50	27.25	20.83	23.25	24.67
	/i/	16.17	28.08	23.92	27.25	23.17
	/u/	18.33	29.08	25.42	19.17	25.67

contexts condition, the data belonging to each vowel context were analysed separately and the average identification percentage was calculated. The identification percentages for F_{1-3} are not identical to those presented in Table 1, where identification scores were computed using all three discriminant functions, whereas only one discriminant function was used for the LDAs in the present table. This, of course, leads to much lower recognition scores.

Table 7 contains information useful for answering questions 1 and 2. Fifteen speakers were entered into the analyses, yielding an identification-by-chance level of $100/15 = 6.67\%$. The presented identification percentages for COART exceed this chance level, but, at least for the pooled contexts condition, only marginally (be it still significantly in χ^2-terms: $\chi^2(2) = 7.26$, $p < 0.05$). As for question 2, we note that the identification percentages obtained for F_{1-3} were much higher than those found for the COART-index, which holds for both analysis conditions. Also the scores for individual formants were higher (except for F_2 of /u/ in the pooled context condition). Obviously, the COART-index is not a very effective transformation for capturing the speaker-discriminating information in the three formants.

Apparently, COART cannot replace the three formants in a speaker identification task, but it may still contain some useful complementary speaker information. Thus we arrive at the third subquestion and ask if the identification scores improve if COART is used as an additional parameter to F_{1-3}. In Table 8 recognition percentages were based on three discriminant functions. (The percentages for F_{1-3} are from Table 1.) Pairwise comparisons of the identification percentages demonstrated that the differences between the two analysis settings (F_{1-3} versus F_{1-3} + COART) are significant for all three vowels in both conditions, but the differences are small, except for /u/ in the pooled contexts condition, where the improvement was about 8%. In general terms, then, also the third subquestion has to be answered negatively.

This leads us to conclude that the COART-index was not found a useful cue in speaker identification for most analysis settings.

Table 8
As Table 7. However, the LDAs were based on other combinations of the predictor variables: F_{1-3} and F_{1-3} combined with COART

Condition	Vowel	F_{1-3}	F_{1-3} + COART
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

6. Discussion

In this study we examined the speaker variability in the coarticulation of /a,i,u/. F_{1-3} were extracted from the middle frame of the vowel's steady-state and coded on a Bark scale. The vowels were taken from /C_1VC_2ə/-words, spoken in isolation by fifteen male speakers of Dutch. The consonants surrounding the vowels were /p,t,k,d,s,m,n,r/. The amount of coarticulation in a vowel context was quantified using a score-model based measure

COART. Two safeguards should prevent that the length of the vocal tract of a speaker was reflected in COART: (1) the reference of coarticulation was made speaker-dependent; (2) formant values were transformed to the Bark-scale, which is more or less logarithmic.

A still better removal of the average vocal tract characteristics of speakers can be achieved by using an entirely logarithmic frequency scale. The Bark-scale tends to be linear up to say 400 Hz, which may conflate the COART-scores for F_1 of /i/ and /u/ with a speaker's average vocal tract characteristics. For /a/ and /u/ logarithmically scaled formants yielded results very similar to those for Bark-scaled formants. For /i/ the COART-scores were somewhat higher than for the Bark-scaled formants (but still relatively low), and the interaction $C \times S$ in Table 5 became significant, thus endorsing the speaker dependency of COART. Considering that the results for logarithmic and Bark-scaled frequencies were very similar and that the Bark-scale is the more commonly used, we adhered to the Bark-scale in this paper.

We concentrated first on the coarticulation effects in the vowel contexts as such (Section 3). It was observed that the effect of coarticulation upon /u/ was much stronger than upon /a/ and /i/ and that especially nasal and alveolar word-initial consonants introduced extensive formant shifts in the F_2 of /u/. Carry-over effects of nasal consonants onto /u/ have been attested by e.g. Flege (1988), as well as a strong effect of initial /d/ on /u/ (Schouten and Pols, 1979; Tokuma, 1993), especially with respect to F_2 (Stevens and House, 1963; Stevens et al., 1966; Ohde and Sharf, 1975). The alveolar character of C_1 in /nukə/, /dupə/, /tumə/ and perhaps /rusə/, which are all contexts yielding a high COART-value (cf. Table 4), suggests that tongue-tip movement may have played a vital role. It can be put forward that due to its sluggishness the tongue-tip dwells at the alveolar ridge for some time during the realisation of /u/. This leads to an apicalisation of the /u/. As a result, the frontal mouth cavity is kept relatively small, which leads to a relatively high F_2-value. This explains, too, why /u/ coarticulates far more with the alveolars than /a/ and /i/ do: the locus of alveolars (about 2000 Hz) is much nearer to the F_2 of /i/ (about 2200 Hz) and, to a less extent, of /a/ (about 1400 Hz) than it is to the F_2 of /u/ (about 850 Hz). This, probably, is the major reason why we found hardly any coarticulation in /i/, only some in /a/ and quite a lot in /u/. Nonetheless, we have to concede that this interpretation does not make clear why most coarticulation was observed in the /u/ of /mutə/. Presumably, the effect of nasalisation upon /u/ is more profound than the effect of apicalisation. For nasal and alveolar consonants we may conclude from our data that the $/C_1V/$-part (with V = /u/, but tendentially the same was found for /a,i/) constitutes a stronger production unit than the $/VC_2/$-part. This is in line with findings published by Ohde and Sharf (1975) and Suomi (1987). It also confirms our expectation that coarticulation is attenuated by the syllable boundary between V and C_2.

The word durations produced by our speakers exhibited a relatively large range, even though the speakers were instructed to use a relaxed speech tempo. Average word durations ranged from 368 ms to 563 ms. We took a closer look at the relationship between the vowel duration and COART for the vowel /u/. For every speaker we compared the average duration of /u/ with the average COART of /u/. A (non-significant) Spearman rank correlation coefficient of $r_S = -0.326$ ($n = 15$) was observed, which indicates not more than a weak relation between a speaker's vowel duration and his COART-value. We also compared the average duration of /u/ with the average COART of /u/ for the eight contexts. We found a (non-significant) Spearman rank correlation of $r_S = -0.381$ ($n = 8$), which is a rather weak relation, too. These observations indicate that vowel duration as such was not a major determinant of COART in our data.

On the whole, between-context differences in coarticulation were substantial only for /u/. Perhaps our speech material (disyllabic words, spoken in isolation without a carrier-phrase, in a relaxed speech tempo) allowed rather near-target realisations of /a/ and /i/ in all contexts. Seen from this perspective, it is noteworthy that the coarticulation phenomena observed for /u/ apparently persist even in such unfavourable conditions. In more extensive data sets containing spontaneous speech or read out sentences stronger effects of vowel reduction and coarticulation may be expected due to the interfer-

ence of additional factors, such as speech tempo, syllable structure, sentence position and stress, cf. (Van Bergem, 1993). Our findings with the presented data set can be considered as a base-line effect of coarticulation as such.

Next, we looked at the speaker variability in the observed coarticulation phenomena (Section 4). 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/). It was further concluded for /a/ and /u/ that the speaker variability in COART correlated positively with the mean value of COART in a context. This makes the mean value of COART in a context a reasonable predictor for the speaker specificity of COART.

Guided by the finding that COART in /a,u/ had turned out to be speaker-specific, a test was performed to assess if COART is a useful parameter for automatic speaker identification (Section 5). Our results indicate that it is not a valuable parameter for this task. Three findings support this view.

1. Used as a single predictor COART is only marginally able to identify speakers above chance level.
2. Identification scores obtained by using F_{1-3} as predictor variables are considerably higher than those obtained for COART.
3. Adding COART to F_{1-3} does not improve identification results.

Similar findings have been reported for /l/ and /r/ by Nolan (1983, p. 112–114). Nolan found that his coarticulation measure (which differed somewhat from ours) yielded identification percentages that were above chance level, but the spectral (filterband) coefficients as such constituted better predictors for speaker identity than did the coarticulation measure. So far, high speaker identification scores for a coarticulation measure have been presented only by Su et al. (1974) for /m/. But this study did not prove that the coarticulation measure performs better than, or as well as, simple spectral coefficients of /m/. Not for a single phoneme, then, has it been reported to date that a coarticulation measure has as much speaker-discriminating power as the spectral coefficients (formant or filterbank values) from which it was derived.

Although we conclude that COART is not generally useful for speaker identification, one exception came to light. For the vowel /u/ in the pooled

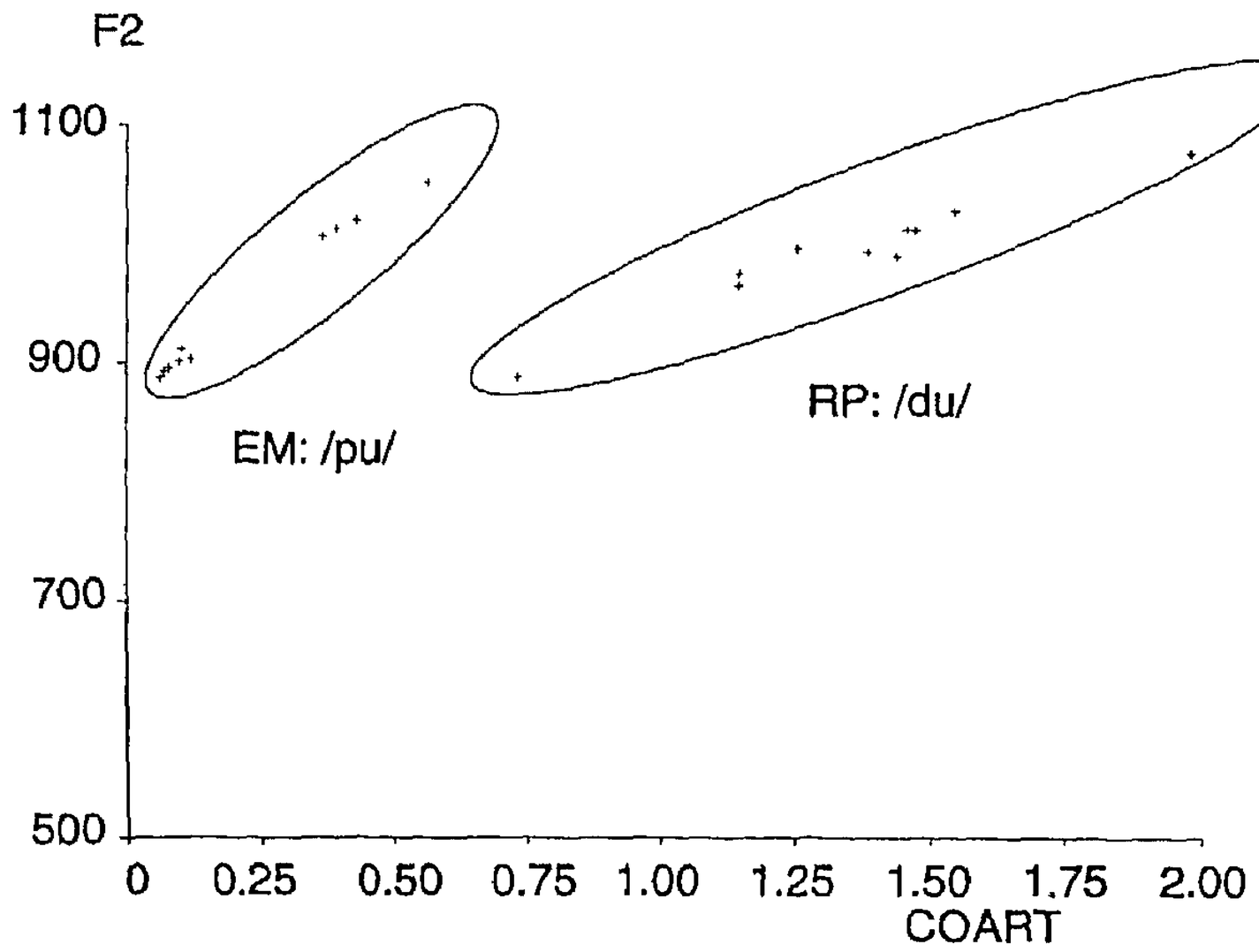

Fig. 9. Illustration of the relation between F_2 and COART in /u/ for speaker EM, and /u/ in /dupə/ for speaker RP. Ellipses around the data points were drawn by hand for clarity. The figure shows that speakers can be separated on the COART dimension, but not on the F_2-dimension.

contexts condition, an 8% identification improvement was achieved by adding COART as a predictor to the formants F_{1-3} (see Table 8). Notably, this effect was not found for the split contexts condition. Fig. 1 may serve as an illustration to explain why. Let us start with only F_{1-3} as predictors in our LDAs. If every context is analysed separately, optimal speaker discrimination is possible, as is indicated by the figure (in a two-dimensional formant space). In the pooled contexts condition, the overlap between speakers precludes this optimal speaker discrimination, since a vowel in one context as realised by a speaker is confused with the realisations of the vowel in another context by another speaker. Now, by adding the COART dimension to the predictor set of the pooled contexts the overlapping formant space can be better decomposed in speaker-specific subspaces and, obviously, better speaker identification scores will be attained. An illustrative example for this, taken from the actual data, is shown in Fig. 9. It can be seen that speakers EM and RP have similar F_2-values for /u/ in two different contexts, but that the speakers can be kept apart by introducing COART as an additional parameter. Thus, COART may turn out to be useful as an extra speaker-identifying cue only if three combined conditions are fulfilled: (a) if COART is used as an additional parameter to the formant values *and* (b) if a phoneme exhibits large between-context differences in coarticulation *and* (c) if the identification procedure is performed on pooled contexts.

It should be kept in mind that our first objective in this study was to examine speaker variability in speech segments from a more fundamental, phonetic point of view. For this reason a simple data set-up as chosen here was preferred. In a next step more complex data sets need to be explored. 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. Probably, the coarticulatory effects observed in the present study will be stronger in sentence material, since it can be expected that articulation rate will be mostly faster. But it is difficult to predict if also the speaker specificity of these coarticulatory effects will be stronger in sentence material. Our finding that the speaker specificity of COART is larger if COART is larger points to the affirmative, but future research is needed here.

Acknowledgements

The authors thank Jan-Willem Hoogakker for his assistance in analysing and discussing the data of this experiment. The detailed and valuable comments of two anonymous reviewers on an earlier version of this paper are acknowledged with gratitude.

References

J.F. Bonastre and H. Meloni (1994), "Inter- and intra-speaker variability of French phonemes. Advantages of an explicit knowledge-based approach", *Proc. ESCA Workshop on Speaker Recognition, Identification and Verifation, Martigny, Switzerland*, pp. 157–160.

R.G. Daniloff and R.E. Hammarberg (1973), "On defining coarticulation", *J. Phonetics*, Vol. 1, pp. 239–248.

J.E. Flege (1988), "Anticipatory and carry-over nasal coarticulation in the speech of children and adults", *J. Speech and Hearing Research*, Vol. 31, pp. 525–536.

C.A. Fowler (1980), "Coarticulation and theories of intrinsic timing", *J. Phonetics*, Vol. 8, pp. 113–133.

H. Hermansky (1990), "Perceptual linear prediction (PLP) analysis of speech", *J. Acoust. Soc. Amer.*, Vol. 87, pp. 1738–1752.

K. Johnson, P. Ladefoged and M. Lindau (1993), "Individual differences in vowel production", *J. Acoust Soc. Amer.*, Vol. 94, pp. 701–714.

R.D. Kent and F.D. Minifie (1977), "Coarticulation in recent speech production models", *J. Phonetics*, Vol. 5, pp. 115–133.

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

F.J. Nolan (1983), *The Phonetic Bases of Speaker Recognition* (Cambridge Univ. Press, Cambridge).

R.N. Ohde and D.E. Sharf (1975), "Coarticulatory effects of voiced stops on the reduction of acoustic vowel targets", *J. Acoust. Soc. Amer.*, Vol. 58, pp. 923–927.

L.C.W. Pols, H.R.C. Tromp and R. Plomp (1973), "Frequency analysis of Dutch vowels from 50 male speakers", *J. Acoust. Soc. Amer.*, Vol. 53, pp. 1093–1101.

A.C.M. Rietveld and U.H. Frauenfelder (1987), "The effect of syllable structure on vowel duration", *Proc. 11th Internat. Congress of Phonetic Sciences, Tallinn, Estonia*, Vol. 4, pp. 28–31.

M.E.H. Schouten and L.C.W. Pols (1979), "Vowel segments in consonantal contexts: a spectral study of coarticulation – Part I", *J. Phonetics*, Vol. 7, pp. 1–23.

S. Shaiman, S.C. Adams and M.D.Z. Kimelman (1995), "Timing relationships of the upper lip and jaw across changes in speaking rate", *J. Phonetics*, Vol.23, pp. 119–128.

D.J. Sharf and R.N. Ohde (1981), "Physiological, acoustic and perceptual aspects of coarticulation. Implications for the remediation of articulatory disorders", in *Speech and Language. Advances in Basic Research and Practice*, ed. by N.J. Lass, Vol. 5, pp. 154–247.

K.N. Stevens and A.S. House (1963), "Perturbation of vowel articulations by consonantal context: An acoustical study", *J. Speech and Hearing Research*, Vol. 6, pp. 111–128.

K.N. Stevens, A.S. House and A.P. Paul (1966), "Acoustical description of syllabic nuclei: An interpretation of a dynamic model of articulation", *J. Acoust. Soc. Amer.*, Vol. 40, pp. 123–132.

L.-S. Su, K.P. Li and K.S. Fu (1974), "Identification of speakers by use of nasal coarticulation", *J. Acoust. Soc. Amer.*, Vol. 56, pp. 1867–1882.

K. Suomi (1987), "On spectral coarticulation in stop-vowel-stop syllables: Implications for automatic speech recognition", *J. Phonetics*, Vol. 15, pp. 85–100.

S. Tokuma (1993), "Some arguments on vowel formant shift", *Speech, Hearing and Language: Work in Progress, UCL*, Vol. 7, pp. 233–254.

D.R. van Bergem (1993), "Acoustic vowel reduction as a function of sentence accent, word stress, and word class", *Speech Communication*, Vol. 12, No. 1, pp. 1–23.

R.J.J.H. van Son (1993), Spectro-temporal features of vowel segments, Dissertation, University of Amsterdam, IFOTT Studies in Language and Language Use, Vol. 3.

D.H. Whalen (1990), "Coarticulation is largely planned", *J. Phonetics*, Vol. 18, pp. 3–35.