Chapter 41
Does the Syllable Affiliation of Intervocalic Consonants Have an Articulatory Basis? Evidence from Electromagnetic Midsagittal Articulography[1]

NIELS O SCHILLER, PASCAL HHM VAN LIESHOUT, ANTJE S MEYER and WILLEM JM LEVELT

Introduction

In Levelt's (1989, Levelt et al., 1999) model of speech production syllables are conceived as articulatory units. Previous research has shown that the syllable can function as a processing unit in Dutch speech production (Levelt and Wheeldon, 1994; Wheeldon and Levelt, 1995; Schiller et al., 1996, 1997; Schiller, 1997; Levelt et al., 1998, 1999; Schiller, 1998). However, very little is known about the mechanisms that underlie the articulatory control of speech production, and Levelt's model does not offer a theory of articulatory execution. The model of *gestural phonology* is more explicit about the processes related to phonetic encoding and articulation (Browman and Goldstein, 1992). In this framework, the customary distinction between phonology and phonetics is given up. The activity of the articulatory motor system is described in terms of underlying *articulatory gestures*. These gestures are the basic units of phonological contrast, and at the same time they characterize articulatory events, i.e. movements in space and time. These articulatory events consist of formations and releases of the vocal tract.

[1]For original version of this chapter, see Schiller, 1997.

The dimensions of these constrictions, e.g. constriction location and constriction degree, are specified by *tract variables* (e.g. lip protrusion, tongue tip constriction location, etc.). The targets of the vocal tract variables are achieved by *model articulators* representing relatively independent articulatory subsystems of the vocal tract, e.g. lips, tongue tip, tongue body, jaw, velum, etc. The model articulators are located on different *articulatory tiers*. Within gestural phonology, a gestural *score* takes care of the coordination of the articulatory gestures, i.e. it specifies the phasing between individual gestures in time and space.

The present study investigated whether the articulatory timing of intervocalic consonants was affected by their syllable affiliation. The results of Browman and Goldstein (1988) and Byrd (1995) were inconclusive with respect to the effect of syllable affiliation on articulatory timing of segments. We tested the *syllable timing hypothesis*, which predicts that segments within a syllable are more stably timed relative to each other than segments that belong to different syllables. The rationale behind this hypothesis is that if syllables are articulatory motor units, the motor coordination of segments that belong to the same unit should be less variable than the coordination of segments from different units. Thus the articulatory timing of two consonants, e.g. /f/ and /k/, should be more stable for CVC items, e.g. *fak.tor*, since both segments are in the same syllable, than for CV items, e.g. *fa.kir*, because /f/ and /k/ belong to different syllables. For CV[C] items, such as *fa[kk]el*, in which the intervocalic consonant is ambisyllabic, the timing between /f/ and /k/ should be even more variable than for the CV items because it is not clear with which syllable /k/ is affiliated.

Experiment

Method

Speech materials

In previous research on the role of the syllable in Dutch speech production (see Schiller et al., 1997; Schiller, 1999) three categories of items were investigated, i.e., CVC words such as *faktor* /fɑk.tɔr/ ('factor') beginning with a CVC syllable, CV words like *fakir* /fa.kir/ ('fakir'), and CV[C] words such as *fakkel* /fɑ[k]əl/ ('torch'), which have an ambiguous syllable boundary. We created 10 item triplets that overlapped in the first three segments (disregarding vowel length) but differed with respect to syllable structure, as in the case of *faktor – fakir – fakkel*. All words were stressed on the first syllable. Dutch phonological structure does not allow for short vowels to occur in open syllables (Branching Rhyme Constraint (BRC), Lahiri and Koreman, 1988), hence the CV words had to include a long vowel, and the other words a short vowel.

Participants

Three female and one male speakers took part in the experiment. They were native speakers of Dutch. None of them reported any speech or hearing disorders.

Apparatus

Tongue, lip, and jaw movements were monitored using the AG100 EMMA system (Carstens Medizinelektronik, Göttingen, Germany; Schönle, 1988), which consists of three transmitter and five receiver coils (for details see Tuller et al., 1990; Perkell et al., 1992; Alfonso et al., 1993; van Lieshout et al., 1995; Schiller, 1997). Movement data were recorded at a sampling rate of 400 Hz. Simultaneously with the monitoring of the articulatory movements, acoustic recordings were made at a sampling rate of 16 kHz. Speech and movement data were digitized simultaneously and aligned by means of the AG100 system software (see Figure 41.1)

Procedure

Participants were seated in a chair, and the helmet necessary to monitor their articulatory movements was attached to a suspending device to improve the stability of its position on the participant's head and to compensate for a substantial portion of the helmet's weight (Alfonso et al., 1993; van Lieshout et al., 1995).

Before the recording session participants' occlusal planes were recorded. Participants were instructed to bite on a plate on to which two receiver coils were attached in the midline. The positions of the two coils were recorded and served as an individual anatomical reference plane to which the experimental data could be rotated in order to compare data across subjects. Immediately before and after data collection the static receiver coil positions were recorded for an informal check of the system's stability during the experiment.

On each test trial participants received one test item. Test items were presented visually on sheets of paper. Participants waited until the experimenter gave a *go*-signal, and then produced multiple repetitions of the test item for a period of 10 seconds. The speech rate was self-selected. Inter-trial intervals were approximately 20 seconds. The recording session lasted approximately 40 minutes. The entire experiment took about one and a half hours. Each participant produced all items. The order of items was randomized individually for each participant with the restriction that items belonging to the same triplet or to the same item category were separated by at least one other trial.

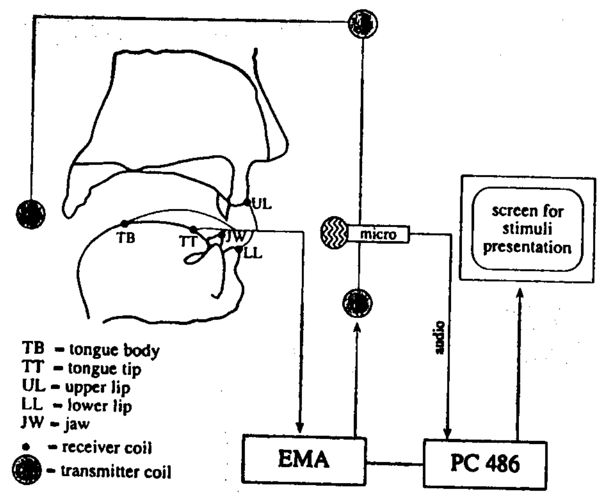

Figure 41.1. Overview of the experimental set-up indicating the location of the transmitter and receiver coils.

Data analysis

The computer routines for the analysis of the articulatory data were similar to the XHADES (Haskins Analysis/Display/Experiment System) software developed at Haskins Laboratories (New Haven, CT, USA) (see Rubin et al., 1991). The analysis routines were integrated into the waves/ESPS speech analysis package (Entropics Inc.), which allows the simultaneous display of the time-aligned acoustic and articulatory signals.

Analysis

After preprocessing the data, the articulatory analyses were based on the displacement data recorded by the coils, which were assumed to reflect most directly reflect the articulatory movement for the constriction gestures under investigation. Research by Gracco and Abbs (1986, 1988; Gracco, 1988) and Hoole et al. (1994) has shown that velocity profiles play an important role in articulatory control during vowel production. Here, we used the velocity characteristics of vertical tongue and lip movements to investigate the articulatory timing of intervocalic consonants.

For the analysis, the second to ninth tokens of each test item were considered. To determine the articulatory timing of the intervocalic consonants, two landmarks were kinematically determined in each test word. The first (anchor point C_1) corresponded to the moment when the articulator(s) forming the release for the onset consonant reached the maximum in the movement signal. The velocity was derived from the displacement signal using a standard differentiation algorithm. The second landmark (C_2) was defined as the articulatory target position of the intervocalic consonant. An articulatory target is generally conceived as a point of minimum velocity (e.g. Perkell et al., 1992). In this study, C_2 corresponded to the temporal midpoints of the consonantal centres (C-centres) of the corresponding consonantal gestures.

At low to moderate speech rates consonantal gestures often display a plateau-like shape, i.e. a quasi-steady state phase, rather than a peak. Hence, the point of zero crossing is not appropriate to determine the articulatory target position. To determine the C-centre of consonantal gestures we adapted a procedure used by Hoole et al. (1994) to define the displacement plateau of the intervocalic consonants. Intervals in the first derivative of the corresponding position signal were demarcated by points in time where for a given cycle both positive and negative velocities became lower than 20% of the peak velocity. Thus, the onset of the interval corresponded to the point in time when the velocity fell below 20% of the peak velocity of a given cycle, whereas the offset corresponded to the point in time when the velocity rose above 20% of the peak velocity. The C-centre was defined as the temporal midpoint of the interval between the two markers, in most cases corresponding to the peak displacement in the position signal.

To quantify stability of the temporal relation between C_1 and C_2 we measured the time interval between C_1 and C_2 for each token (see Figure 41.2), and computed the standard deviation (SD) of the length of this interval C_2-C_1 across the eight repetitions of each test item.

Results and Discussion

The SD of the interval C_2-C_1 across the eight repetitions was determined per item and participant to compare the stability of the articulatory timing within each item. The means of the SDs per participant and category are shown in Table 41.1. As can be seen, the SD was markedly smaller for the CVC items than for the remaining item types for only one participant. Analyses of variance were run per participant to compare the three item categories (CV, CVC, or CV[C]). The main effect of item category was not significant for any of the participants (participant 1: $F(2,27) = 2.28$, $F < 1$ for the other participants).

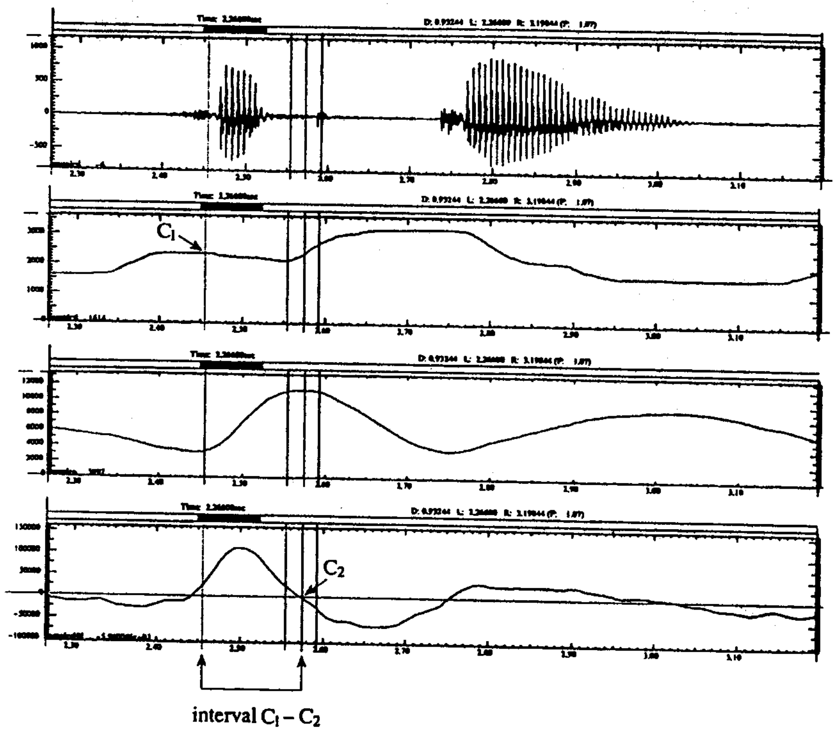

Figure 41.2. Display of the speech signal, the jaw and tongue body movement signal, and the velocity profile of the tongue body movement for one token of the experimental trial *faktor* produced by participant 3. The upmost panel displays the acoustic signal, the upper middle panel the jaw movement (recorded from coil JW), the lower middle panel the tongue body movement (recorded from coil TB), and the lowest panel the velocity profile of the tongue body movement. The anchor point C_1 and the C-Centre (C_2) as well as the interval C_2-C_1 are indicated in the figure.

Table 41.1. Mean SDs of the interval C_2-C_1 per participant and item category

	Mean SD of the interval C_2-C_1 (in ms)		
participant	CV items	CVC items	CV[C] items
1	28	15	29
2	55	47	56
3	31	36	31
4	38	38	35

The length of the first vowel significantly differed between the three item categories as was determined in the acoustic signal by sonagraphic analyses (means: CVC: 78 ms, CV[C]: 80 ms, and CV: 167; $F\ (2957) = 803.14$, $MS_e = 1022.33$, $p < 0.001$). The mean length and the SD of the interval C_2-C_1 correlated significantly for three of the four participants (participant 1: $r = 0.50$, $p = 0.01$; participant 2: $r = 0.42$, $p = 0.05$; participant 3: $r = 0.64$, $p < 0.001$). Therefore, we ran analyses of covariance entering the mean length of the interval C_2-C_1 as a covariate in order to take into account the vowel length differences between the item categories. However, the differences of the SDs between the categories only reached significance for one participant (participant 3: $F\ (2,26) = 5.88$, $p = 0.01$).[2]

The results did not provide any evidence for the hypothesis that the syllable affiliation of an intervocalic consonant plays a role in articulatory timing. The variability of the articulatory timing in different item categories did not show a stable pattern across participants. The fact that the timing of consonants within a syllable was *not* more stable than the timing of onset consonants of successive syllables came as a surprise. This result clearly contradicts the syllable timing hypothesis according to which the variability of the articulatory timing should be highest for the CV items since C_1 and C_2 belong to different syllables. The fact that items including ambisyllabic consonants did not significantly differ from the other two item categories with respect to the articulatory timing of the intervocalic consonant was also unexpected.

Summary and conclusion

Syllables are seen as articulatory-motor units in Levelt's model of speech production. We investigated whether the phonological syllable affiliation of intervocalic consonants is reflected on the articulatory output level, i.e. at the stage of motor execution. The results revealed no significant differences between the timing of segments within a syllable and the timing of the same segments when a syllable boundary occurred between them. But since the difference in vowel length is problematic for comparisons between items, future research may focus on the articulatory timing of onset and coda consonants in monosyllables with short versus long vowels.

Nevertheless, this study has proved the usefulness of the EMMA method for the observation of articulatory movements during speaking in a non-clinical setting with real speech material.

---

[2]Analyses of variance using the quotient of the SD and the mean of the C_2-C_1 interval as a coefficient of variation revealed the same pattern of results (participant 1: $F\ (2,27) = 1.47$, $MS_e < 0.01$, ns; participant 2: $F\ (2,27) < 1$; participant 3: $F\ (2,27) = 5.29$, $MS_e < 0.01$, $p = 0.012$; participant 4: $F\ (2,27) < 1$).

References

Alfonso PJ, Neely JR, van Lieshout PHHM, Hulstijn W, Peters HFM (1993) Calibration, validation, and hardware-software modifications to the Carstens EMMA System. Haskins Laboratories Status Report on Speech Research, SR-114, 101–12.

Browman CP, Goldstein L (1988) Some notes on syllable structure in articulatory phonology. Phonetica, 45, 140–55.

Browman CP, Goldstein L (1992) Articulatory phonology: An overview. Phonetica, 49, 155–80.

Byrd D (1995) C-centers revisited. Phonetica, 52, 285–306.

Gracco VL (1988) Timing factors in the coordination of speech movements. The Journal of Neuroscience, 8, 4628–39.

Gracco VL, Abbs JH (1986) Variant and invariant characteristics of speech movement planning. Experimental Brain Research, 65, 156–66.

Gracco VL, Abbs JH (1988) Central patterning of speech movement timing. Experimental Brain Research, 71, 515–26.

Hoole P, Mooshammer C, Tillmann HG (1994) Kinematic analysis of vowel production in German. Proceedings of the Third International Conference on Spoken Language Processing, Yokohama, pp. 53–6.

Levelt WJM (1989) Speaking. From Intention to Articulation. Cambridge, MA: MIT Press.

Levelt WJM, Roelofs A, Meyer AS (1999) A theory of lexical access in speech production. Behavioral and Brain Sciences, in press.

Levelt WJM, Schiller NO (1998) Is the syllable frame stored? Commentary on the BBS target article 'The frame/content theory of evolution of speech production' by MacNeilage PF. Behavioral and Brain Sciences, 21, 520.

Levelt WJM, Wheeldon L (1994) Do speakers have a mental syllabary? Cognition, 50, 239–69.

Perkell JS, Cohen MH, Svirsky MA, Matthies ML, Garabieta I, Jackson MTT (1992) Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. Journal of the Acoustical Society of America, 92, 3078–96.

Rubin P, MacEachron M, Tiede M, Maverick V (1991) HADES - Haskins Analysis/Display/Experiment/System: Documentation. New Haven: Haskins Laboratories.

Schiller NO (1997) The role of the syllable in speech production. Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. PhD dissertation, Nijmegen University.

Schiller NO (1998) The effect of visually masked syllable primes on the naming latencies of words and pictures. Journal of Memory and Language, 39, 484–507.

Schiller NO, Meyer AS, Baayen RH, Levelt WJM (1996) A comparison of lexeme and speech syllables in Dutch. Journal of Quantitative Linguistics, 3, 8–28.

Schiller NO, Meyer AS, Levelt WJM (1997) The syllabic structure of spoken words: Evidence from the syllabification of intervocalic consonants. Language and Speech, 40, 101–39.

Schönle PW (1988) Elektromagnetische Artikulographie. Ein neues Verfahren zur klinischen Untersuchung der Sprechmotorik. Berlin: Springer.

Tuller B, Shao S, Kelso JAS (1990) An Evaluation of an Alternating Magnetic Field Device for Monitoring Tongue Movements. Journal of the Acoustical Society of America, 88, 674–9.

van Lieshout PHHM, Alfonso PJ, Hulstijn W, Peters HFM (1995) Electromagnetic Midsagittal Articulography (EMMA). In Maarse FJ, Akkerman AE, Brand AN, Mulder LJM, van der Stelt MJ (eds) Computers in Psychology: Applications, Methods, and Instrumentation. Lisse: Swets & Zeitlinger, pp. 62–76.

Wheeldon L, Levelt WJM (1995) Monitoring the time course of phonological encoding. Journal of Memory and Language, 34, 311–34.