Degemination of Dutch Fricatives in Three Different Speech Rates

Luc Martens and Hugo Quené

1. Introduction

In a previous experiment, we have investigated production data of minimal pairs that differed in their underlying phonological representation, viz. with a single or double consonant. An example is the two-word phrase "zee fijn" vs. "zeef fijn" /ze:(f)#fɛIn/ ('sieve/sea fine'). According to phonological theory, this contrast should be completely neutralised in the phonetic surface form, due to a Degemination rule that deletes one of two adjacent and identical consonants (e.g. Booij, to appear). We compared duration patterns of single and degeminated fricative consonants in two-word phrases as mentioned above, spoken in a sentence context. In comparing these two conditions, we found significant differences in absolute fricative durations, as well as in the ratio of the durations of the fricative consonant and its preceding vowel. The conclusion was drawn that Degemination is not an absolute process (as most phonological accounts seem to imply), but instead a gradual phenomenon (Martens, 1993).

The production research reported here is an extension to this previous study, addressing essentially the same question with Dutch stimulus material, but now using data obtained from three speech rates. We predict that the gradual nature of the Degemination process becomes more clear when comparing underlying single and double consonants in various speech rates. In particular, we predict that the contrast between the two underlying structures disappears as the speech rate increases. In slow speech, Degemination will probably not result in total deletion of one of the members of the double consonant - a trace of its articulation will remain, manifesting itself as a longer fricative duration. In fast speech, by contrast, no such trace of the double consonant will be manifest. These expectations are based on the general hypothesis that assimilation processes such as Degemination are in some way related to the speech rate: assimilation is more pervasive (with stronger phonetic effects) as the speech rate is faster, and vice versa.

2. Experimental Method

The minimal pairs of two-word phrases in this study were a superset of those used in the previous experiment (Martens 1993). All stimulus phrases were Noun-

Linguistics in the Netherlands 1994, 119–126. DOI 10.1075/avt.11.13mar
ISSN 0929–7332 / E-ISSN 1569–9919

-Adjective combinations, consisting of two monosyllabic and monomorphematic words. The members of a minimal pair were segmentally identical, but differed in their underlying representation, viz. with a single or double consonant. In the former case, the (single) consonant always belonged to the second word only (/..V#CV../). In the latter case, this post-boundary consonant also occurred in pre-boundary position (/..VC_i#C_iV../). The pivotal intervocalic consonant was always a voiceless fricative, because these allow easy temporal manipulation of their phonetic surface form, i.e. frication noise (Martens 1993). The resulting minimal pairs are listed in Table 1 below.

Table 1. Minimal pairs of two-word stimulus phrases. The two variants of the first word are given pairwise for each stimulus phrase (see text).

nr	pair of phrases	gloss
1	ra / raaf fraai	'yard / raven beautiful'
2	zee / zeef fijn	'sea / sieve fine'
3	brie / brief fraai	'brie[cheese] / letter beautiful'
4	lei / lijf fors	'slate / body large'
5	roe / roef fors	'rod / deckhouse large'
6	wei / wijf fijn	'meadow / wife fine'
7	ei / ijs slecht	'egg / ice bad'
8	reu / reus slim	'male dog / giant clever'
9	prei / prijs scherp	'leek / price sharp'
10	moe / moes slap	'mom / pulp limp'
11	wee / wees slecht	'woe / orphan bad'
12	ree / race snel	'roe / race quick'

The members of these pairs, viz. the two-word phrases, were embedded in an identical carrier sentence:

> "ik dacht dat die N1 A1 en die N2 A2 was"
> 'I thought that that N1 A1 and that N2 A2 was'

where N1, N2 and A1, A2 represent various Nouns and Adjectives, respectively.

In the present experiment, the crucial factors to be varied in the stimulus material are [1] the underlying phonological representation and [2] the speech rate. In order to make the material suitable for future word perception experiments, additional factors were also varied. This was done by manipulating the sentence containing the stimulus phrase. The sentence contained a critical word which [3] was either semantically related (beach/fork) to the first word of the stimulus phrase (sea/sieve), or it was an unrelated filler word (chair). In addition, [4] the stimulus phrase could occur late vs. early in the carrier sentence, thus following vs. preceding its semantic prime or filler.

As an example, Table 2 below contains the complete design for one of the twelve minimal pairs.

Table 2. Overview of factors varied, demonstrated for one minimal pair of two-word phrases, viz. "ra / raaf fraai". The carrier sentence has been omitted, relevant slots are indicated as N1, A1, N2, A2. Factors illustrated are [1] the underlying phonological representation, [4] the position of the stimulus phrase, and [3] the presence of a related vs. unrelated word.

[1]REPR	[4]POS	[3]SEM.REL.	N1	A1	N2	A2
single	early	related	ra	fraai	mast	grijs
single	early	filler	ra	fraai	muur	grijs
single	late	related	mast	grijs	ra	fraai
single	late	filler	muur	grijs	ra	fraai
double	early	related	raaf	fraai	mees	grijs
double	early	filler	raaf	fraai	muur	grijs
double	late	related	mees	grijs	raaf	fraai
double	late	filler	muur	grijs	raaf	fraai

All stimulus sentences from all conditions were mixed into pseudo-random order, and printed on a list. All critical Nouns and Adjectives were printed in capitals, as an indication to the speaker to accentuate these words.

The list was read aloud by a male (native Dutch) speaker, seated in a sound-treated booth. The speaker was the same as the one performing in the previous study (Martens, 1993). He was instructed to read the list at three different speech rates, viz. slow, normal and fast. The normal rate was similar to the speech rate as observed in the previous study (Martens, 1993). The list was read twice for each speech rate. If the speaker hesitated or made an error, the stimulus sentence was repeated until both the speaker and the first author were satiesfied (this was necessary a few times). All realisations were recorded on digital audio tape (DAT) at 48 kHz sample frequency, using high-quality recording equipment and a Sennheiser ME 60 microphone. The above procedure resulted in a total number of 576 stimulus phrases [2 representations x 2 positions x 2 sentence contexts x 12 minimal pairs x 3 speech rates x 2 replications].

Each sentence was digitised (at 10 kHz sample frequency, 4.5 kHz low-pass filtering, 12 bits) and the durations were measured of the pivotal consonant, and of the vowel immediately preceding it. In addition, the duration of the whole sentence was measured. Measurements were done by means of software providing both visual (waveform) and auditory feedback, using standard criteria based on van Zanten, Damen and van Houten (1991). The estimated error was less than 5 ms. From the consonant and vowel durations, a ratio was calculated. The advantage of this ratio is that between-utterance variance in absolute consonant and vowel durations is reduced. Such variance may result from different coarticulatory effects between minimal pairs. In addition, differences in speech rate between utterances are normalised in this ratio.

3. Results and Discussion

In order to allow comparison between speech rate conditions, the differences between speech rates should be verified first. Analysis of sentence durations showed an average speaking rate in the Slow condition of 3.6 syl/s (standard deviation 0.4), in the Normal condition of 4.6 syl/s (s.d. 0.2) and in the Fast condition of 5.3 syl/s (s.d. 0.2). Based on these averages, the speech rates were judged to be sufficiently different to justify comparison between speech rate conditions.

Table 3 below presents the average durations of the pivotal consonant, of the preceding vowel, and of the ratio of these two durations.

Table 3. Average durations of vowel and consonant (in ms), and of their ratio (in arbitrary units), broken down by speech rate and by the underlying phonological representation. Standard deviations are given in parentheses. Each average is based on 96 observations.

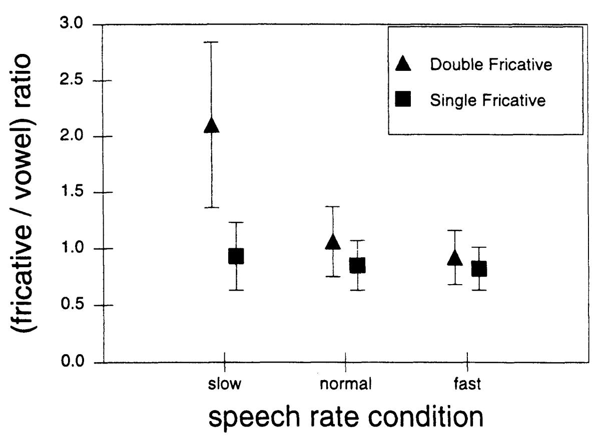
speech rate	phon. repr.	vowel duration	fricative duration	fricative/vowel ratio
SLOW	single	205 (48)	181 (44)	0.93 (0.30)
	double	169 (38)	334 (72)	2.10 (0.74)
NORM	single	150 (27)	123 (15)	0.85 (0.22)
	double	141 (28)	141 (17)	1.06 (0.31)
FAST	single	103 (14)	103 (14)	0.82 (0.19)
	double	112 (16)	112 (16)	0.92 (0.24)

Five-way analyses of variances were performed on these three dependent variables. We will concentrate here on results regarding the consonant-to-vowel ratio (for reasons given above); results regarding the two absolute durations are summarised in the Appendix. Independent variables were [1] the underlying phonological representation, [2] the speech rate, [3] the presence of a related vs. unrelated word, [4] the position of the stimulus phrase in the carrier sentence, and [5] the minimal pair. Only the latter was treated as a random factor in the analysis of variance.

The results showed a significant main effect of [1] underlying representation on consonant-to-vowel ratio [$F(1,11)=56.7$, $p < .001$]. Inspection of the results in Table 3 shows that consonant duration varies more between representation conditions than vowel duration does. Hence, this main effect is largely due to the longer duration of double consonants as compared to single ones. Apparently, Degemination does not always result in total deletion of one of the members of the double consonant. In addition, a main effect of [2] speech rate was observed [$F(2,22) = 90.9$, $p<.001$]. This effect is due to the fact that - in this particular stimulus material - the pivotal fricative consonant has greater temporal 'elasticity' than its preceding vowel. The consonant-to-vowel ratio decreases as speech rate increases: hence the consonant shortens more than the vowel at faster speech rates. This finding contradicts results from other studies (e.g. Nooteboom 1972, Klatt 1976). This unexpected effect is probably due to the specific phonemic make-up of the stimulus sentences. Thirdly, the factor [5] minimal pair yielded a significant main effect [$F(11,288)=59.9$, $p<.001$], indicating temporal differences between the minimal pairs of phrases. This effect is due to phonemic differences between minimal pairs, which may yield differences in segmental duration (Klatt, 1976). Other main effects were not significant.

Interestingly, a significant two-way interaction was observed between underlying representation and speech rate [$F(2,22)=65.9$, $p<.001$]. This interaction is illustrated in Figure 1 below. For all speech rates, the consonant-to-vowel ratio differs between single and double consonants; however, the temporal contrast decreases as speech rate increases. This supports the primary hypothesis in this study. Impressionistic auditory analysis of the stimulus phrases provides additional support: at slow speech rates, a drop in the amplitude contour of the frication noise was audible quite regularly, namely in 61 out of 96 relevant cases. This decrease and subsequent increase of the amplitude contour clearly indicates that there are two separate articulatory gestures involved. The resulting fricative speech segment has an extremely long duration (mean 334 ms), corresponding to two separate phonemes rather than one.

Figure 1. Average ratio of pivotal consonant with respect to preceding vowel, for phrases with underlying single and double consonants, for three speech rates.



A second two-way interaction that reached significance was phonological representation x minimal pair [$F(11,288)=10.1$, $p<.001$]. This effect indicates different single vs. double consonant contrasts for individual minimal pairs (as the phonemic structures differ). A third two-way interaction that reached significance was speech rate x minimal pair [F(22,288)=4.2, p<.001]. This effect points out that speech rate affects individual minimal pairs differently (as the phonemic structures vary). The last two-way interaction that reached significance was speech rate x position [F(2,22)=23.4, p<.001]. This effect indicates that speech rate has a different effect as the positions of the fragments differ (due to other prosodic environments of early vs. late fragments). Other two-way interactions were not significant.

A three-way interaction that reached significance was [1] phonological representation x [2] speech rate x [5] minimal pair [F(22,288)=4.2, p<.001]. Again this should not be surprising as the phonemic structures differ among minimal pairs. Quite remarkable, an other three-way interaction that reached significance was [1] phonological representation x [3] position x [4] filler/related Noun [F(1,11)=9.5, p<.01]. We do not have an explanantion for this effect. Other three-way and all four-way interactions were not significant.

4. General Discussion

The results presented above clearly demonstrate that Degemination is a gradual phenomenon, there is never complete deletion of one of the members of the double consonant. Even at fast speech rates, a trace of the double consonant can still be observed in the phonetic surface form. We predicted that the temporal contrast would be strongest in slow speech, and that no such contrast would be observed in fast speech. The observed interaction pattern was as predicted, although the relevant contrast can still be observed in fast speech. The results of this study suggest that Degemination is a gradual phenomenon, and that it applies stronger as the speech rate increases.

From a perceptual point of view, this implies that fast speech is phonetically more ambiguous than slow speech, at least in this respect. Listeners have to recognise words in the connected speech signal. To this end, both acoustic-phonetic and non-sensory information is used (e.g. Marslen-Wilson 1987). In faster speech, however, it seems that acoustic-phonetic information is less reliable, because assimilation phenomena (such as Degemination) have a more drastic effect on the phonetic surface form. Consequently, word recognition may be impeded in faster speech (or, recognition tends to rely more upon non-sensory information).

In future research, we will investigate the perception of ambiguous two-word stimulus phrases, while manipulating the amount of non-sensory information that can contribute to recognition of the target words. The stimulus phrases will have to be maximally ambiguous, so that the effect of non-sensory cues on solving the ambiguity can be assessed reliably. Hence, the stimulus phrases will have a fricative duration which corresponds to the cross-over point between "zee fijn" and "zeef fijn" recognition responses. To this end, a calibration study has been conducted, in which the duration of the pivotal fricative consonant was varied systematically. In the future, we hope to report in more detail about these word recognition experiments.

References.

Booy, G. (to appear) *The phonology of Dutch*, Oxford University Press.

Klatt, D.H. (1976) 'Linguistic uses of segmental duration in English: acoustic and perceptual evidence', *Journal of Acoustical Society of America*, 59 (5), 1208-1221.

Marslen-Wilson, W.D. (1987) 'Functional parallelism in spoken word recognition', *Cognition*, 25, 71-102.

Martens, L.M.H. (1993) 'Production of boundary-ambiguous two-word phrases: vowel - and fricative durations as disambiguating cues', ms., Utrecht University.

Nooteboom, S.G. (1972) *Production and perception of vowel duration, A study of durational properties of vowels in Dutch,* Doctoral dissertation, Utrecht University.

van Zanten, E., L. Damen and E. van Houten (1991) 'The ASSP speech data base', SPIN/ASSP-report 41, Foundation for Speech Technology, Utrecht.

Appendix

Table A1. Summary of analysis of variance results with the duration of the pivotal consonant as dependent variable. (Only main effects and significant interaction effects are given).

effect	F	df	p
[1] Phonological Representation	270.2	1,11	<.001
[2] Speech Rate	935.4	2,22	<.001
[3] Position	8.0	1,11	<.05
[4] Filler/Related	2.5	1,11	n.s.
[5] Minimal Pair	3.1	11,288	<.001
1 x 2	177.2	2,22	<.001
1 x 3	6.4	1,11	<.05

Table A.2. Summary of analysis of variance results with the duration of the pre-boundary vowel as dependent variable. (Only main effects and significant results are given).

effect	F	df	p
[1] Phonological Representation	61.6	1,11	<.001
[2] Speech Rate	148.0	2,22	<.001
[3] Position	11.1	1,11	<.01
[4] Related/Filler	1.6	1,11	n.s.
[5] Minimal Pair	90.4	11,288	<.001
[1] x [2]	28.6	2,22	<.001
[2] x [3]	10.3	2,22	<.001
[2] x [5]	3.3	22,288	<.001