Unfolding of phonetic information over time: A database of Dutch diphone perception

Roel Smits[a)]
Max Planck Institute for Psycholinguistics, Postbus 310, 6500 AH Nijmegen, The Netherlands

Natasha Warner
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands and Department of Linguistics, University of Arizona, Tucson, Arizona

James M. McQueen and Anne Cutler
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

(Received 19 December 2001; accepted for publication 3 October 2002)

We present the results of a large-scale study on speech perception, assessing the number and type of perceptual hypotheses which listeners entertain about possible phoneme sequences in their language. Dutch listeners were asked to identify gated fragments of all 1179 diphones of Dutch, providing a total of 488 520 phoneme categorizations. The results manifest orderly uptake of acoustic information in the signal. Differences across phonemes in the rate at which fully correct recognition was achieved arose as a result of whether or not potential confusions could occur with other phonemes of the language (long with short vowels, affricates with their initial components, etc.). These data can be used to improve models of how acoustic-phonetic information is mapped onto the mental lexicon during speech comprehension. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1525287]

PACS numbers: 43.71.Es [KRK]

I. INTRODUCTION

We describe a database of phonetic perception in Dutch, in which 18 listeners judged the first and the second phoneme in gated fragments of all possible Dutch diphones, providing 27 140 identification responses per listener. This database constitutes the largest source of data that is currently available on phonetic perception in Dutch or any other language.

We undertook the project with the aim of motivating a more realistic and fine-grained representation of speech input in computational models of human spoken-language processing such as TRACE (McClelland and Elman, 1986) and Shortlist (Norris, 1994). To this end we wished to determine the accuracy with which human listeners can evaluate acoustic information as speech input unfolds over time, and to compile this information for the entire phoneme inventory of a language, in all potential left and right phonetic contexts. Although phoneme confusion matrices have in the past been obtained from speech in noise (e.g., Miller and Nicely, 1955) as well as from gated signals (e.g., Smits, 2000), we chose the latter method for two reasons. First, we were primarily concerned to examine the detailed temporal resolution of speech perception, and gating easily permits any desired temporal resolution. Second, our interest is in speech perception under general listening conditions. Adding noise to a speech signal creates difficult listening conditions, and moreover differentially affects speech sound categories such as consonants versus vowels.

Our choice of gating does not imply any claim that this task directly reflects online activation of phonemes in speech perception. It is clear that to perform the task, listeners engage a decision process which presumably has no part in normal speech perception (Grosjean, 1996). This decision mechanism will use additional processing time and may incorporate additional information (e.g., phoneme transition statistics) not present in the acoustic stimulus. We believe, however, that gating offers the currently best available window into listeners' resolution of ambiguity as speech signals unfold.

Our materials consisted of a total of 2294 sequences (1179 diphone sequences, of which most were recorded in multiple stress conditions to enable us also to assess effects of stress on acoustic information in phoneme realizations). Each listener heard six gates of each sequence, based on six gating points, three in each sound of the diphone. The shortest gate included only the first third of the first sound; each subsequent gate included another sixth of the entire diphone. The entire stimulus set (all gates from all diphone sequences) was presented to each listener in a different pseudo-random order.

II. METHOD

A. Materials

1. Choice of diphones

We first compiled a list of all possible diphones of the Dutch language. For this purpose, we considered the phonemic inventory of Dutch to be as in Tables I and II.

a)Electronic mail: roel.smits@mpi.nl

TABLE I. The 16 Dutch vowels used in the experiment.[a]

	Front unrounded			Front rounded			Central	Back		
	Diphthong	Long	Short	Diphthong	Long	Short		Diphthong	Long	Short
High		i			y				u	
Mid		e	ɪ, ɛ		œ	ʏ	ə		o	ɔ
Low									a	ɑ
	ɛi			œy				ɑu		

[a]Compared to Booij (1995), we have simplified the vowel system slightly by combining upper and lower mid vowels into a single height.

Decisions as to what constitutes a single phoneme versus a sequence of two phonemes were based on CELEX, an electronic database containing lexical data for English, Dutch, and German (Baayen *et al.*, 1993). We did not, however, include all phonemes and diphones in CELEX (see Appendix A for explanation of exceptions). We constructed a list of diphones consisting of all possible combinations of any two of these phonemes. Appendix B lists the selection rules we applied. Appendix C lists the 2294 diphones included in the experiment, and reasons for exclusion of missing diphones.

2. Recording

Each diphone in Appendix C was placed in a nonsense environment which, with the diphone, formed a phonotactically legal sequence in Dutch. CV and VC diphones were recorded with both stressed and unstressed vowels; VV diphones were recorded with all four possible stress combinations. Table III lists the environments in which the various diphones were recorded.

The nonsense environment always included at least one phoneme after the target diphone, so that the diphone would not be final to the item. This prevented excessive lengthening within the diphone, as would for example apply to the vowel in a CV diphone recorded in isolation. Stressed CV diphones were always followed by the unstressed syllable /kə/, whereas unstressed CV diphones were always followed by stressed /ke/. VCs always straddled a syllable boundary, with one of the syllables stressed and the other unstressed. If unstressed, the final syllable was Cə, if stressed it was Ce. If CC was a legal onset, it formed the onset of the syllable CCa. Otherwise it straddled a syllable boundary, with the first syllable aC stressed and the second Cə unstressed. VV diphones always straddled a syllable boundary. Depending on the stress pattern, the contexts /b/ or /ab/ were prefixed, and the contexts /k/, /kə/, or /ke/ were suffixed, to make the sequences easier to produce with correct stress.

All items (diphones in their environments) were transcribed phonemically, with stress and syllable boundaries marked. A phonetically trained female native speaker of Dutch, whose pronunciation exhibits no strong regional accent, read all of the items from this transcription. The recording was made on DAT in a sound-treated recording booth using high-quality equipment. Any items which were initially mispronounced were rerecorded. The recording was low-pass filtered at 7.5 kHz and resampled at 16 kHz.

3. Stimuli for the perception experiment

Past gating studies have employed two methods for dividing the signal. First, gates can be positioned at fixed time intervals [e.g., 20 ms, as in Smits (2000)], leading to a variable number of gates per diphone. Alternatively, gates can be positioned ''proportionally,'' i.e., using a constant number of gates per phoneme (e.g., Cutler and Otake, 1999), leading to a variable gate duration. We chose proportional gating for two reasons. First, the number of stimuli for our experiment would become unrealistically large if we were to use fixed intervals while at the same time making several gates available for even the shortest diphone. Second, as described above, the ultimate aim of the study was to provide data on which to base computational modeling of the arrival of phonetic information over time; proportional gating provides data which is relatively straightforward to use in this way.

Beginnings and ends of all phonemes were identified manually using the criteria in Appendix D. Each item was

TABLE II. The 22 Dutch consonants used in the experiment.

	Labial/ Labiodental		Alveolar		Postalveolar/ Palatal		Velar/ Uvular		Glottal	
	Voiceless	Voiced	Voiceless	Voiced	Voiceless	Voiced	Voiceless	Voiced	Voiceless	Voiced
Stops	p	b	t	d			k	g		
Nasals		m		n				ŋ		
Fricatives	f	v	s	z	ʃ	ʒ	x[a]			h
Affricate						dʒ				
Liquids				l				r[b]		
Glides		w[c]				j				

[a]This fricative is /χ/, but for ease of transcription we will use /x/.
[b]This liquid is /ʀ/, but for ease of transcription we will use /r/.
[c]This glide is /ʋ/, but for ease of transcription, we will use /w/.

TABLE III. Environments in which diphones were recorded (in phonemic transcription). Syllable boundaries are marked by hyphens.

Diphone class	Environment	Proportion with each environment
CV (stressed)	ˈCV-kə	2/3
	a-ˈCV-kə[a]	1/3
CV (unstressed)	CV-ˈke	2/3
	ˈa-CV-ke	1/3
VC (vowel stressed)	ˈV-Cə	1/2
	ˈbV-Cə	1/2
VC (vowel unstressed)	V-ˈCe	1/2
	bV-ˈCe	1/2
CC	ˈCCa	if CC is a legal onset
	ˈaC-Cə	otherwise
VV (stressed–unstressed)	ˈbV-Vk	all
VV (unstressed–stressed)	bV-ˈVk	all
VV (stressed–stressed)	ˈbV-ˈV-kə	all
VV (unstressed–unstressed)	ˈa-bV-V-ˈke	all

[a]For all diphones beginning with /ŋ/, /ɑ/ was used as the preceding vowel instead of /a/ because /ŋ/ cannot follow long vowels.

final-gated at six points during the target diphone, three in each of the target phonemes (with exceptions for initial stops and affricates, see below), to create stimuli consisting of the entire item up to the gating point, including any preceding context.

For phonemes which lack abrupt acoustic changes during the segment, such as nasals, fricatives, and vowels in most environments, gate end points were placed automatically at one-third and two-thirds through the duration of the segment as well as at the end of the segment. For segments with abrupt acoustic changes within the segment, such as stops and affricates, gate end points were determined relative to those abrupt changes. Any preceding environment was always included in the stimuli, but following environment was never included.

With gating it is most important to avoid introducing extraneous acoustic cues in the gated segments. Pols and Schouten (1978), among others, showed that careless truncation of speech signals may bias listeners towards labial and or plosive responses. They also showed, however, that such biases can be minimized by applying smoothing windows and replacing the missing speech by another signal such as noise. At gate end points, items were therefore ramped down to zero using a linear 5-ms ramp. In order to further avoid noise-introduced fricative biases, we used as a replacement signal a 500-Hz square wave, which is not misperceived as a speech sound (Warner, 1998). The square wave had a duration of 300 ms, with the same 5-ms ramp applied at onset and offset, and was overlap-added to the end of the item such that the start of the item's falling ramp coincided with the start of the square wave's rising ramp. The amplitude of the square wave was fixed across stimuli. The rms amplitude of a 50-ms portion of the square wave was 22 dB lower than the rms amplitude of the loudest 50-ms portion across all stimuli.

Mean phoneme duration across all utterances was 138 ms, with a standard deviation of 64 ms. Mean duration of a signal portion between two consecutive gate points was 48 ms, with a standard deviation of 23 ms. The total number of stimuli was 13 570.

B. Subjects and procedures

Twenty-two listeners participated in the experiment, and 19 completed it. All were native speakers of Dutch who had grown up in the Netherlands, and had no known hearing impairment; most were students at the University of Nijmegen. Subjects were paid for each hour of participation, with a bonus on finishing the experiment. Data from the three subjects who did not finish the entire experiment were excluded.

The task involved identifying the two phonemes of the target diphone. Subjects were tested individually in a sound-treated booth. Stimuli were presented over closed headphones. As each stimulus was played, a response screen appeared on a computer screen visible through the booth window. The response screen showed two panels, each containing buttons for each phoneme used in the experiment. Subjects used a computer mouse to click on one button of the left-hand panel for the first sound of the diphone, and one of the right-hand panel for the second sound. If the stimulus included preceding context (/a/, /ɑ/, /b/, or /ab/), the letters ''aa,'' ''a,'' ''b,'' or ''aab,'' respectively, appeared on the screen to the left of the left-hand response panel to inform subjects that those sounds were not the ones to which they should respond. The response buttons for these phonemes were also crossed out in the left response panel to remind subjects not to respond to the preceding environment.

Before beginning the experiment, subjects were trained on the set of symbols to use for responses. Since Dutch orthography is straightforward, most phonemes could be represented orthographically (with double vowels used for long vowels and single vowels used for short vowels); special symbols were necessary only for /ə/ (''@'') and /g/ (''G''). Examples of each phoneme were provided, and special attention was called to phonemes which appear only in loan words. Subjects were told that they would hear the beginning of a nonsense word followed by a beep, and that they should identify the two sounds of the nonsense word using the mouse. They were informed about possible additional initial sounds which they were not to respond to, and warned that they would sometimes hear very little of the nonsense word, making it difficult to identify the two sounds. A native Dutch speaker instructed each subject and checked subjects' understanding of the mapping of response symbols to sounds.

Subjects then completed a practice session, comprising 185 stimuli drawn from the actual experiment. Diphones containing potentially problematic phonemes, such as /ə, ŋ/ and phonemes occurring only in loan words, were well represented in the practice session. The experimenter evaluated subjects' performance on stimuli which included these sounds or a vowel in their entirety to ensure that subjects could perform the task. No subjects were excluded at this stage, since none had difficulty with the task.

Subsequently, subjects completed a series of one-hour experimental sessions, with a break during each session. Subjects returned for as many sessions as needed to respond to all 13 570 stimuli, an average of 27.9 sessions. The total

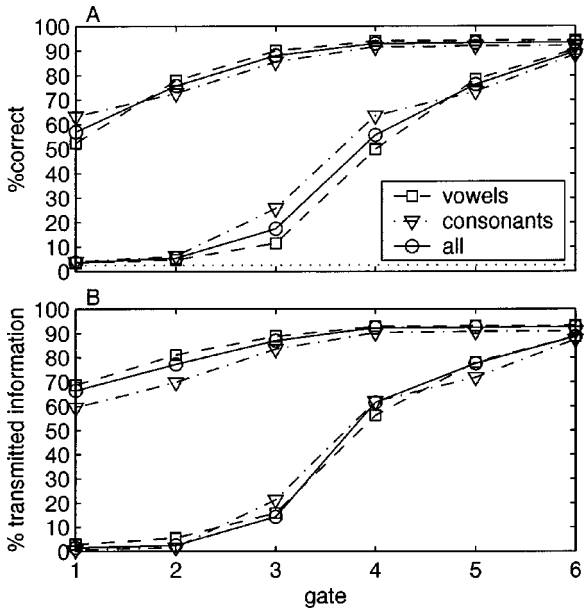

FIG. 1. Correct phoneme recognition rates (a) and percentages transmitted information (b) as a function of gate, averaged across listeners. Results for vowels only, consonants only, and all phonemes are given by separate lines. The upper and lower lines are associated with the first and second phoneme in the diphone, respectively. The dotted line in (a) indicates chance level (2.63%).

set of stimuli was divided into four blocks. For each subject a different pseudo-random order of stimuli within blocks was generated and different subjects received the blocks in a different order. Two gates of the same diphone were separated by at least six stimuli, stimuli from diphones beginning with the same phoneme were separated by at least four stimuli, and no stimuli which appeared in the practice session or other gates of those diphones occurred within the first 1200 experimental stimuli. In total 488 520 phoneme categorizations were collected.

III. RESULTS

A. Summary results

One subject performed much worse than the others in correctly recognizing the first phoneme at gates 1–3. For these gates this subject's recognition rates were more than four standard deviations below the mean recognition rates for all other subjects. The data of this subject were therefore excluded. Figure 1 shows average phoneme recognition rates (panel a) and percentages transmitted information (TI, panel b) as a function of gate, pooled across the remaining 18 subjects, for consonants, vowels, and all phonemes. TI is a measure of the covariance between input and output when both have a categorical nature (e.g., Miller and Nicely, 1955; Smits, 2000).

At gate 1, that is, one-third into the first phoneme of the diphone, the first phoneme (top line) was recognized at almost 60% correct, while TI reaches almost 70%. With increasing gates, levels rose smoothly to about 90% at gate 4 and hardly changed thereafter. The recognition rate for the second phoneme (bottom line) started close to chance level (2.6% correct, or 0% TI) at gate 1 and rose smoothly to almost 90% at gate 6. One-tailed *t*-tests showed that at all gates average recognition rates for both phonemes were significantly above chance level as well as below perfect performance (all p's<0.0005). In these as well as all subsequent tests, subject was the random variable, and the Bonferroni criterion was applied in calculating the significance levels (above, 24 comparisons were made, so the significance level was $\alpha=0.002$).

Recognition rates for gates 4–6 of the second phoneme were quite similar to those for gates 1–3 of the first phoneme. The longer preceding context for the second phoneme therefore did not affect recognition much compared to the first phoneme. The recognition curves for vowels and consonants are very similar. In first position, TI is somewhat lower for consonants than for vowels (about 10% for gates 1 and

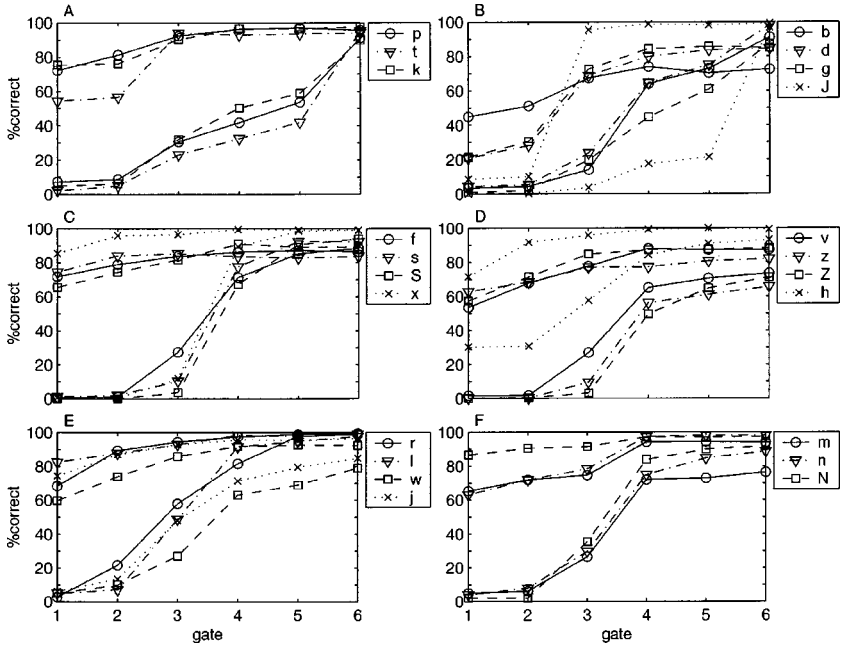

FIG. 2. Correct consonant recognition rates plotted separately for each of the 22 consonants. Phoneme symbols are in accordance with IPA, except for J, S, Z, and N, indicating /dʒ ʃ ʒ ŋ/, respectively. The upper and lower lines are associated with the first and second phoneme in the diphone, respectively.

TABLE IV. Confusion matrix for consonants. Responses were summed across subjects, contexts, and stress conditions. For each stimulus, the first row gives responses to gate 1 for consonants in initial position in the diphone, whereas the second row gives responses to gate 4 for consonants in second position. The last column gives the number of vowel responses to each of the consonants.

	Response																						
Stimulus	p	t	k	b	d	g	dʒ	f	s	ʃ	x	v	z	ʒ	h	r	l	w	j	m	n	ŋ	Vowel
p	325	6	3	62	0	0	0	0	0	0	0	0	0	0	22	0	0	4	0	9	0	0	19
	331	8	34	187	13	16	0	11	0	0	0	11	0	0	58	2	10	43	15	15	13	5	20
t	33	235	4	13	81	1	0	3	0	0	0	5	0	0	25	1	3	15	2	0	0	0	11
	28	258	7	12	340	2	9	2	1	1	1	10	0	0	35	3	5	11	13	6	26	3	19
k	0	0	340	0	0	66	0	0	0	0	3	0	1	0	23	0	1	2	3	0	1	1	9
	26	18	399	11	7	120	1	8	1	0	2	6	0	0	77	6	9	28	13	3	5	5	47
b	77	0	2	275	9	1	0	3	0	0	0	11	1	0	95	2	0	25	1	76	20	0	14
	18	2	0	566	32	18	0	2	0	0	0	10	2	0	11	2	3	98	4	92	10	4	8
d	11	29	0	89	116	2	9	6	1	0	0	11	0	0	99	1	9	67	8	48	37	5	10
	6	5	3	45	571	2	1	1	0	0	3	3	0	0	25	0	40	61	8	6	71	3	28
g	5	0	60	99	19	123	0	1	0	0	6	10	0	0	92	2	4	39	2	58	36	11	9
	6	1	75	82	33	394	4	4	0	1	22	9	1	1	30	4	8	88	22	8	16	11	62
dʒ	8	46	1	95	35	2	49	5	0	0	0	16	0	0	113	2	2	49	21	70	51	6	5
	9	12	1	10	457	2	148	0	0	0	0	2	4	2	17	3	8	8	94	1	47	3	18
f	4	0	0	0	1	0	0	646	1	0	9	172	0	0	43	3	0	6	0	2	0	0	13
	3	0	1	0	1	0	0	565	0	0	12	179	0	0	9	0	1	17	3	0	0	0	1
s	6	2	0	0	0	1	0	1	670	46	0	0	141	3	27	2	0	0	0	1	0	0	0
	3	6	0	1	1	0	0	0	601	26	0	0	107	14	2	2	1	2	1	4	1	0	2
ʃ	4	0	0	0	0	3	2	1	91	590	1	0	17	139	18	0	0	0	26	1	0	0	7
	2	5	1	1	0	1	19	2	112	522	2	2	15	66	5	0	0	2	10	2	1	0	4
x	2	1	2	1	0	0	0	6	0	0	784	1	1	0	52	54	0	1	1	2	0	0	10
	0	0	1	0	0	1	0	3	0	1	709	2	0	0	47	19	0	1	1	0	1	0	6
v	3	2	0	1	1	0	0	126	0	0	0	385	0	0	66	3	0	114	2	4	0	0	13
	1	0	0	0	1	0	0	116	1	0	3	445	2	0	8	4	1	84	9	0	2	0	7
z	5	2	0	2	2	0	5	0	67	23	0	0	452	96	32	3	16	0	7	1	0	0	7
	2	2	0	1	5	0	7	1	106	20	0	1	394	90	12	0	7	5	16	5	1	0	27
ʒ	3	2	0	0	5	4	9	0	17	86	0	0	44	330	31	2	3	0	28	1	2	0	9
	0	3	0	0	3	2	32	0	27	115	1	1	136	428	6	1	1	1	57	2	2	0	46
h	10	1	1	9	1	0	0	25	1	1	16	20	0	0	386	2	6	29	16	5	6	0	5
	2	0	1	2	1	0	0	10	0	1	6	2	0	0	683	3	7	21	12	0	2	0	57
r	1	0	9	1	1	3	0	1	2	0	5	4	0	0	174	628	12	18	5	10	10	3	31
	0	1	0	0	0	0	0	0	1	1	0	4	0	0	16	691	5	6	1	0	1	0	119
l	5	1	4	2	0	0	0	1	0	0	1	1	0	0	67	0	758	10	21	5	3	0	39
	0	0	0	0	1	0	0	0	0	0	1	3	0	0	5	6	759	11	5	1	5	2	29
w	17	0	1	32	3	1	0	0	1	0	0	17	0	0	107	5	20	549	3	101	31	1	29
	2	5	1	3	6	1	0	10	3	0	1	65	1	0	46	22	19	534	15	6	6	2	98
j	0	1	1	3	1	0	0	1	4	0	1	3	5	0	84	0	12	4	683	5	4	3	103
	4	2	2	4	0	0	0	0	3	0	0	0	0	0	20	3	8	15	591	1	3	0	172
m	5	0	0	3	0	0	0	0	0	0	0	2	0	0	120	1	11	30	11	599	113	2	21
	0	0	0	1	0	0	0	1	0	0	0	4	0	0	13	2	11	67	1	609	103	20	14
n	4	0	2	3	2	0	0	0	0	0	0	5	0	0	108	2	9	17	15	140	579	11	21
	1	0	0	2	3	0	0	2	1	0	1	3	4	0	10	1	25	41	4	88	648	7	23
ŋ	0	1	2	0	0	1	0	0	0	0	1	1	0	0	18	6	3	2	5	0	28	810	58
	0	0	1	0	0	0	0	0	0	0	0	0	0	0	8	1	1	2	1	1	13	166	4

2), but in second position this difference disappears.

Figure 2 shows correct recognition rates by gate separately for the 22 consonants, grouped by manner and voicing, while Fig. 3 presents those for the 16 vowels, grouped partly according to vowel features and partly according to similarities of the individual curves. Tables IV and V present confusion matrices for consonants and vowels, respectively, summed across listeners, contexts, and stress conditions, in responses to gate 1 for the first phoneme and to gate 4 for the second phoneme.

B. Consonants

(1) *Voiceless stops* /p t k/ [Fig. 2(a)]: As shown in Table III, some diphones were recorded with preceding context and some without. For those without preceding context, gates 1 and 2 were not presented because they contained only silence. Gates 1 and 2 in Fig. 2(a) therefore represent only responses to gated diphones with preceding context—that is, the vowel /a/ with formant transitions plus respectively half or all of the following stop closure. Subjects could recognize

TABLE V. Confusion matrix for vowels. Responses were summed across subjects, contexts, and stress conditions. For each stimulus, the first row gives responses to gate 1 for vowels in initial position in the diphone, whereas the second row gives responses to gate 4 for vowels in second position. The last column gives the number of consonant responses to each of the vowels.

	Response																
Stimulus	ɑ	ɛ	ɪ	ɔ	ʏ	ə	i	u	y	e	o	œ	a	ɛi	œy	ɑu	Consonant
ɑ	640	0	0	4	1	9	0	0	0	0	0	0	26	0	1	11	28
	1275	5	1	27	7	13	0	1	0	0	1	0	131	3	29	45	10
ɛ	0	642	3	0	0	22	0	0	0	1	0	0	0	29	0	0	23
	42	1165	64	2	5	21	2	0	3	15	0	4	37	159	4	0	25
ɪ	2	1	611	0	1	6	32	0	0	30	0	0	0	0	0	0	37
	5	81	1125	2	25	18	90	1	17	127	0	6	4	4	0	0	43
ɔ	3	0	0	634	2	5	0	2	0	0	28	0	0	0	0	0	46
	92	1	2	1291	6	14	1	5	3	0	81	4	1	0	0	1	46
ʏ	0	1	6	1	450	144	0	0	20	1	0	59	0	0	0	0	38
	18	9	5	59	793	404	1	3	36	3	10	119	3	0	9	1	75
ə	10	5	20	8	439	259	4	5	51	4	0	46	12	1	4	0	86
	7	4	21	23	367	205	0	3	21	1	2	53	3	0	1	2	43
i	0	0	34	0	0	13	1671	0	5	2	0	0	2	0	0	0	145
	0	0	163	2	7	13	1260	1	12	3	0	5	1	1	0	0	80
u	0	0	1	18	4	29	0	1732	2	0	2	1	0	0	1	0	82
	0	1	3	47	21	21	1	1307	32	0	4	2	0	0	1	2	106
y	0	0	0	1	59	56	4	4	1588	0	0	4	0	0	0	0	156
	0	0	11	8	104	60	29	115	1048	0	1	8	0	0	4	2	158
e	0	30	1301	0	6	32	6	0	0	411	0	0	1	2	0	0	83
	1	179	989	2	7	11	30	2	0	289	1	2	0	3	1	0	31
o	4	2	0	1189	8	47	0	30	1	0	474	0	1	0	0	0	116
	23	1	0	1136	14	19	0	7	1	0	289	4	1	0	0	7	46
œ	0	9	9	2	1052	400	0	0	5	20	1	290	0	1	2	2	79
	13	4	10	18	814	373	8	5	28	7	4	191	0	0	0	1	72
a	426	23	2	0	0	66	1	1	1	0	0	0	1211	45	10	1	85
	431	90	2	0	8	7	3	0	1	0	1	1	841	76	51	1	35
ɛi	55	828	0	1	0	84	2	1	0	2	0	0	43	815	2	0	39
	149	602	4	0	5	19	3	0	1	3	0	2	248	457	18	3	34
œy	412	78	1	0	12	120	0	0	0	0	0	3	417	135	614	4	76
	602	48	3	2	24	33	1	0	1	1	0	3	452	34	306	12	26
ɑu	1484	1	2	9	3	52	1	0	0	1	0	0	54	0	6	168	91
	1307	3	2	33	2	6	0	0	0	0	1	1	59	0	12	105	17

the stops well from these portions, with recognition rates between 50% and 80%. Note that Dutch voiceless stops are produced without aspiration, while voiced stops are usually produced with negative VOT (voice bar). Recognition of /t/ was somewhat poorer than of /p/ and /k/. This is supported by *t*-tests (all comparisons between /t/ and /p/ or /t/ and /k/ at gates 1 and 2 reached significance, $\alpha = 0.01$). The difference was mainly caused by more place and voicing errors for /t/ than for /p/ and /k/ (see Table IV). Gate 3 included the release burst, which strongly improved recognition.

Recognition of voiceless stops in second position in the diphone at gates 4 and 5 was considerably worse than recognition of the first phoneme at gates 1 and 2 ($p < 0.005$ for all six comparisons, $\alpha = 0.008$). The raw data show that, on average, /a/ as preceding context led to better recognition of the following stop than other preceding contexts. This agrees with reports of Dorman *et al.* (1977) and Smits *et al.* (1996) that formant transitions in /a/ are more informative about place of articulation of an adjacent consonant than transitions in other vowels. At gate 6, when the stop burst is audible, recognition levels exceeded 90%.

(2) *Voiced stops* /b d ɡ/ *and the voiced affricate* /dʒ/ [Fig. 2(b)]: Gates 1 and 2 included half or all of the voice bar, while the third gate included the release burst. In first position, recognition of voiced stops was poorer than for voiceless stops (only 1 out of 18 comparisons did not reach significance, $\alpha = 0.0025$). /b/ fared better than /d/ and /ɡ/ for gates 1 and 2 ($p < 0.001$ for all four comparisons, $\alpha = 0.01$), reconfirming the findings of, among others, Pols and Schouten (1978) and Smits (2000) that an isolated voice bar sounds more like a /b/ than a /d/ or /ɡ/. For later gates, place and voicing confusions were the main source of errors (see Table IV). Voiced stops were more often confused with their voiceless counterparts than vice versa. Especially /b/ was classified relatively frequently as /p/ up to gate 6. The voiced affricate /dʒ/ was not recognized reliably until its final gate, when burst and frication become audible. At earlier gates /dʒ/ was mainly confused with /j/ and /d/.

(3) *Voiceless fricatives* /f s ʃ x/ [Fig. 2(c)]: For all fricatives, the three gates comprise one-third, two-thirds and all of the frication noise, respectively. At gate 1 of the first phoneme recognition was already good, with levels between

60% and 90%. Recognition gradually improved with increasing amounts of frication and subsequent context. Remaining confusions of /f s ʃ/ were with their voiced counterparts. In addition, there was some confusion between /s/ and /ʃ/ (see Table IV). The voiceless velar fricative /x/ was recognized very well at all gates. Note that /x/ has no voiced counterpart in most regional variants of Dutch, including that of our speaker. Recognition levels for gates 4–6 of the second phoneme resembled those for gates 1–3 for the first. Note the marked jump in recognition between gates 3 and 4, that is, when some frication noise became audible.

(4) *Voiced fricatives* /v z ʒ h/ [Fig. 2(d)]: In initial position, voiced fricatives were generally recognized as well as their voiceless counterparts (only 1 out of 18 comparisons reaches significance, $\alpha = 0.0025$). In second position, however, voiced fricatives were recognized less well than their voiceless counterparts at gates 4–6 ($p < 0.0005$, $\alpha = 0.0025$). Although the pattern is thus less clear than for the stop consonants, it has the same cause, namely asymmetric confusions of the voicing feature. Voiced fricatives were categorized as their voiceless counterparts more often than the reverse (see Table IV). This pattern may be related to the fact that for many regional variants of Dutch, including the one spoken in Nijmegen (but not the native variant spoken by the talker), the voicing distinction in fricatives is weak, with voiced fricatives being pronounced as their voiceless counterparts.

The glottal fricative /h/ was recognized better than the other fricatives (in initial position 11 out of 18 comparisons reach significance, $\alpha = 0.0025$; in second position 17 out of 18 comparisons reach significance, $\alpha = 0.0025$). In first position recognition already exceeded 90% at gate 2. Note that /h/ has no voiceless counterpart, so if manner and place of articulation are recognized, there is no room for voicing errors. In second position /h/ was recognized well even at gate 1. This is an artifact of the gating method: some subjects used a default /h/ response for the second phoneme when they had no information about that phoneme. As the second phoneme sometimes actually was /h/, this response bias increased recognition rates for the early gates of /h/ in this position.

(5) *Liquids* /r l/ and *glides* /w j/ [Fig. 2(e)]: Positioning of begin and end points for these phonemes varied greatly depending on context, but the three gate points always divided the phoneme into equal thirds (see Appendix D). Recognition in first position was already good at gate 1, with recognition rates between 60% and 85%. At later gates recognition further increased to very high levels. In second position, recognition of the labiodental glide /w/ was significantly poorer than of the liquids for gates 3–6 ($\alpha = 0.001$); confusions occurred with the voiced labiodental fricative /v/ and the vowels /ʏ/ and /ə/ (N.B. /w/ was hardly ever confused with the vowel /u/). Recognition of liquids and glides in second position gradually increased across all six gates. From gate 4 onwards, however, recognition of the glides was substantially lower than that of the liquids ($\alpha = 0.002$ is reached for all 12 comparisons), and asymptoted at levels close to 80%. /w/ was again mainly confused with /v/ and /j/ was mainly confused with /i/, while the main confusions for /r/ were with /h/ and /ɑ/. The confusions for /l/ were rather scattered and include consonants /d j h r w/ and vowels /i ɪ ə/.

(6) *Nasals* /m n ŋ/ [Fig. 2(f)]: The three gate points divided the nasal murmur into equal thirds. For nasals in first position it is striking that /ŋ/ was recognized much better than /m n/ at gates 1 to 3 (all six comparisons reach significance, $\alpha = 0.008$). This is again an artifact: Because /ŋ/ cannot occur in syllable-initial position, recognition levels of /ŋ/ in initial position were based on tokens with preceding context /ɑ/, which therefore includes formant transitions into the nasal. In contrast, /m/ and /n/ occurred in initial position in two-thirds of the tokens, without informative preceding transitions. For nasals in second position a marked increase in correct recognition can be seen at gates 3 and 4, which include the speech signal up to oral closure and one-third into the murmur, respectively. Table IV shows that at gate 1 in first position and at gate 4 in second, confusions were mainly across place, while at later gates the remaining confusions were across manner and place was recognized reasonably well. At gates 5 and 6, recognition of /m/ was some 15% lower than that of /n/ and / ŋ/. The raw data show that /m/ was often confused with /n/ at these gates.

C. Vowels

(1) *Short vowels* /ɑ ɛ ɪ ɔ/ [Fig. 3(a)]: At gate 1, recognition of these vowels in first position was already very good, with levels close to 90% correct. In second position, recognition jumped to levels between 70% and 85% at gate 4 and rose further at subsequent gates. When listeners heard one-third or more of the target vowel, the remaining confusions were as follows. /ɑ/ was mainly confused with /a/, /ɛ/ with /ɛi/ and /ɪ/, /ɪ/ with /e/ and /i/, /ɔ/ with /ɑ/ and /o/ (see Table V). That is, short vowels were confused with any nearby long counterpart.

(2) *Long vowels* /i u y/ [Fig. 3(b)]: These, like the short vowels, were recognized well in first position at gate 1. Note that these vowels do not have short counterparts (Booij, 1995). When a third or more of the vowels was audible, the remaining confusions tended to be with similar short vowels: /i/ was confused with /ɪ/, /u/ with /ɔ ə w/, and /y/ with /ə ʏ u/ (see Table V).

(3) *Short vowels* /ʏ/ and /ə/ [Fig. 3(c)]: Recognition of /ə/ was poor, showing little improvement over the six gates and never exceeding 40% correct. /ʏ/ was recognized better, but still much worse than the short vowels in Fig. 3(a). As shown in Table V, /ʏ/ and /ə/ more or less form a single category: responses to both stimuli were very similar, and listeners seem to have selected at random between the two responses, with a bias against /ə/ (such a bias has also been encountered by others, Van Son, personal communication). We therefore grouped stimuli and responses for these two vowels together and calculated recognition rates for the compound vowel class. The resulting recognition curves are displayed in Fig. 3(c) with the label ''Y/@.'' In first and second position, recognition for the new class was significantly better than that of /ə/ at all gates ($\alpha = 0.002$). Compared to /ʏ/,

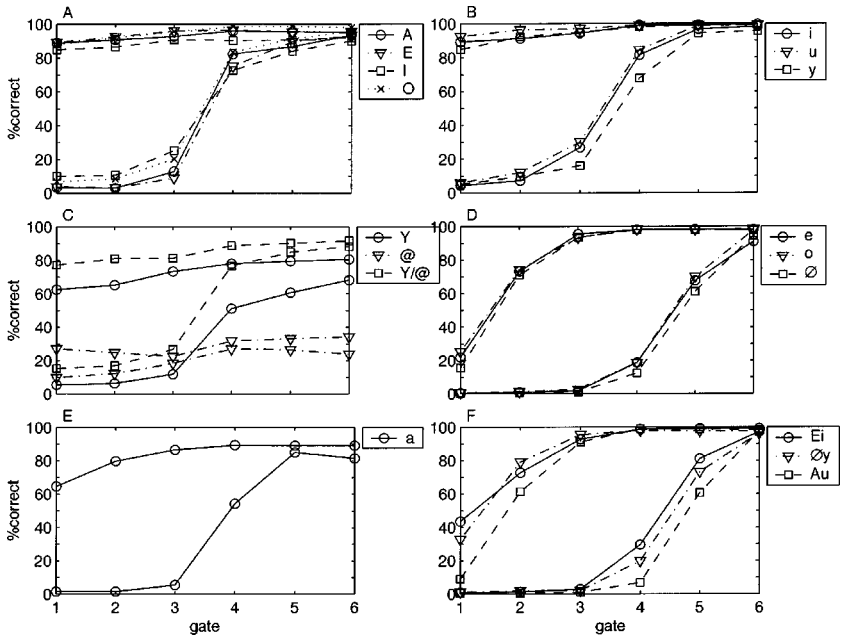

FIG. 3. Correct vowel recognition rates plotted separately for each of the 16 vowels. Phoneme symbols are in accordance with IPA, except for A, E, I, O, @, and ø, indicating /ɑ ɛ ɪ ɔ ə œ/, respectively. The upper and lower lines are associated with the first and second phoneme in the diphone, respectively.

the new class was better for gates 1, 4, 5, and 6 in second position only ($\alpha = 0.002$). This shows that at gates where at least a third of the vowel is audible, the majority of confusions were indeed between /ʏ/ and /ə/. The remaining confusions were mainly with /œ/ (see Table V).

(4) *Long vowels* /e o œ/ [Fig. 3(d)]: In most regional variants of Dutch, including that of our speaker, these vowels are slightly diphthongized, ending in articulatory positions corresponding to /i u y/, respectively (Booij, 1995). In first position, these phonemes were initially not well recognized. At gate 1, recognition levels were between 15% and 25%, which is much lower than for other vowels discussed so far. At gate 1, /e/ and /o/ were mainly confused with /ɪ/ and /ɔ/, respectively, while /œ/ was mainly confused with /ʏ/ and /ə/ (see Table V). This is partly supported by Booij's (1995) position that the short counterparts of /e/ and /œ/ are indeed /ɪ/ and /ʏ/ (with /ʏ/ and /ə/ being highly confusable, as discussed earlier), while /o/ and /ɔ/ do not form a long-short pair because /o/ is higher than /ɔ/. Our data suggest, however, that, perceptually, the relation between /ɔ/ and /o/ is very similar to that between /ɪ/ and /e/. At gate 2, recognition levels were just above 70%, and the full three gates were necessary for recognition to exceed 90%. The recognition results for /e o œ/ in second position are very similar to those for the first position, shifted by three gates.

(5) *Vowel* /a/ [Fig. 3(e)]: This vowel is depicted separately because it shows a pattern between that of /i u y/, which have no short counterpart, and that of /e o œ/, which do. This finding tallies with the description of /a ɑ/ as "almost" a long-short pair, with the qualification that both vowels are back, but /a/ is somewhat fronted compared to /ɑ/ (Booij, 1995). Another aspect which sets /a/ apart from the other long vowels is that its recognition asymptoted just below 90%, whereas the others were eventually recognized at levels close to 100%.

The raw data show that at all gates /a/ was recognized better when stressed than unstressed. When it was unstressed /a/ was mainly confused with /ɑ/ and to a lesser extent with /ə/ and /ɛi/. The pattern is, however, more subtle. When /a/ was part of a VV diphone (which always has a syllable boundary in the middle), and the stress pattern of this diphone was either weak–strong or strong–weak, the confusion with /ɑ/ was much less than when it was part of an unstressed CV or VC diphone, or a VV diphone with a weak–weak stress pattern. We hypothesize that when /a/ is stressed or it is possible to hear that /a/ is unstressed (by contrast to the adjacent syllable), listeners are more likely to choose the (correct) /a/ response. The data show that the same general pattern applies to /e/ and /œ/, but the effect is much weaker, possibly due to their slight diphthongization, which makes confusions with their short counterparts less likely.

(6) *Diphthongs* /ɛi œy ɑu/ [Fig. 3(f)]: The general picture is similar to that for the diphthongized long vowels [Fig. 3(d)], but there is more variability. When only part of the diphthong was audible, /ɑu/ was recognized worse than the other two diphthongs (in first position both comparisons reached significance at gate 1 while only /œy/ versus /ɑu/ did so at gate 2; in second position all comparisons involving /ɑu/ reached significance for gates 4 and 5, $\alpha = 0.001$). Not surprisingly, /ɑu/ was predominantly classified as /ɑ/ for these gates (see Table V). /œy/ was mainly confused with /ɑ/, /a/, and /ə/ at early gates, while /ɛi/ was mainly confused with /ɛ/ (see Table V). When the diphthongs were fully audible, recognition levels were close to 100%.

IV. SUMMARY AND CONCLUSIONS

We have presented the method and results of a large-scale study of the perception of gated versions of all possible Dutch diphones. For the consonants we found the following six confusion patterns. First, inclusion of bursts considerably improved recognition of both voiced and voiceless stops. This finding agrees with past studies on stop recognition

(e.g., Schouten and Pols, 1983; Smits *et al.*, 1996). Second, voiceless stops were recognized better than voiced stops. This difference was caused by asymmetrical voicing confusions: voiced stops were classified as voiceless more often than the reverse. This pattern has not been reported earlier. Third, fricatives were recognized well from only a third of their frication noise. This had already been established for English (Jongman, 1989; Smits, 2000), but not for Dutch. Fourth, the same asymmetrical pattern of voicing confusions that we found for stops applied to the fricatives. This pattern has been documented for American English by Jongman (1989). Fifth, perceptually relevant information was temporally more spread out for liquids and glides than for other consonants. A similar pattern was reported by Klaassen-Don (1983). Sixth and finally, in accordance with Kurowski and Blumstein (1984) and Smits (2000), our results show that transitions into the nasal murmur, together with the first few pulses of the murmur, contain important information for nasal recognition.

The confusion patterns for vowels were dominated by the long–short distinction. This corresponds well with previous studies employing gated vowels (e.g., Strange *et al.*, 1976; van Bergem, 1993). Short vowels were recognized well as soon as a third of their duration became available. However, /ʏ/ and /ə/ formed an exception to this rule, mainly because they were mutually confused. Long vowels that do not have short counterparts were also recognized well from a third of their duration. For long vowels with short counterparts, on the other hand, as well as for diphthongs, the entire vowel was needed for correct recognition. The pattern for the long vowel /a/, which forms an approximate long–short pair with /ɑ/, fell between the two extreme patterns.

The database of Dutch diphone perception described here is available at http://www.mpi.nl/world/dcspdiphones. It was collected with the aim of improving existing models of spoken word recognition. In particular, we plan to replace the input representation of the Shortlist model (Norris, 1994), which currently consists of a string of phoneme labels, by phoneme activation patterns that are graded and temporally more fine-grained. These activation patterns will be derived from the present database. The planned improvements will enable a start to be made on modeling the match between the speech signal and competing word candidates in a more realistic manner.

ACKNOWLEDGMENTS

We are grateful to Dennis Norris for discussion of this material. We further thank Mattijn Morren, Keren Shatzman, Petra van Alphen, Niels Janssen, Tau van Dijck, Anne Pier Salverda, and Aoju Chen for their great efforts in preparing and running the experiment, and Saskia Bayerl for assistance with preparing the database for the Internet. Finally, we thank James Hillenbrand, Rob van Son, and an anonymous reviewer for very helpful comments on an earlier version of this paper.

APPENDIX A: PHONEME SELECTION CRITERIA

Reasons for selection or exclusion of certain phonemes are as follows:

(1) Besides the voiceless velar fricative /x/, CELEX and Booij (1995) recognize the voiced velar fricative /ɣ/. We excluded /ɣ/ because many Dutch speakers—including the speaker for the experiment—neutralize the distinction, maintaining only /x/ (Gussenhoven, 1992).
(2) The vowels /iː, yː, uː, ɔː, œː, ɛː/ occur only in a few unassimilated loan words (e.g., *analyse, centrifuge, cruise, zone, oeuvre, serre*, respectively), and contrast with native phonemes only in length. We excluded these non-native vowels as Gussenhoven (1992) and Booij (1995) both hold them to be marginal.
(3) We did include some consonants which occur in Dutch only in unassimilated loan words: the voiced velar stop /g/, the fricative /ʒ/, and the affricate /dʒ/. These appear in a relatively large number of loans, many quite frequent (e.g., *goal; jam*, /ʒɛm/; and *jazz*).
(4) There are inconsistencies in the CELEX inventory, e.g., the fact that [tʃ] is treated as a sequence of a stop and a fricative, /tʃ/, while [dʒ] is treated as a single affricate segment /d͡ʒ/. In these cases we observed the CELEX standard.

APPENDIX B: DIPHONE SELECTION CRITERIA

The following criteria were applied in selection of the diphones:

(1) For each sequence of two phonemes containing a vowel other than /ə/ (which is never stressed), one diphone was included with the vowel stressed, and another with it unstressed. For vowel–vowel diphones, all four stress combinations (stressed–stressed, unstressed–unstressed, stressed–unstressed, unstressed–stressed) were included.
(2) We included diphones which can only occur across word or morpheme boundaries in Dutch (e.g., /ŋp/), but we excluded sequences which, because of phonotactic constraints, could never occur even across word boundaries.
(3) In cases where phonotactic constraints were violated by large numbers of loan words, we included the diphones. Thus Booij's (1995) claim that short vowels cannot be followed by a glide within the syllable might be considered to be violated by *timing*, *tranquilizer*, and *boiler*.
(4) We excluded certain diphones which are possible (at least across morpheme boundaries) according to a phonemic transcription, but unlikely ever to be produced as a sequence of the two sounds, e.g., /sʃ, ʃs, td͡ʒ/.
(5) We excluded all sequences of identical consonants ($C_1 = C_2$), since Dutch phonology requires that these be degeminated within the prosodic word (Booij, 1995), and they are likely to be reduced to a single consonant even across word boundaries unless produced with a pause.
(6) A few diphones which probably never occur in Dutch, e.g., /ɑ, ɛ, ʏ/ followed by /ʒ/, were included simply because no known phonotactic constraint excludes them.

APPENDIX C: DIPHONE TEST SET

TABLE VI. Diphones included in the experiment, and reasons for exclusions. Each row represents diphones X_1X_2, where X_1 is each of the phonemes in the X_1 column and X_2 is each of the phonemes in the X_2 column.

Class	X_1	X_2
CV diphones		
C=stop, affricate, nasal, liquid, or glide	p, t, k, b, d, g, dʒ, m, n, ŋ, r, l, j, w	all full vowels stressed, all vowels unstressed
C=fricative	f, v, s, z, ʃ, ʒ, x, h	all full vowels stressed
	f, v, s, z, ʃ, ʒ, x	all vowels unstressed
	h	all full vowels unstressed
	Exclusion: */hə/ within the syllable, and /h/ cannot be syllable-final[a]	
VC diphones		
C=stop, affricate, liquid, or glide	all full vowels stressed, all vowels unstressed	p, t, k, b, d, g, dʒ, r, l, j, w
C=fricative	all full vowels stressed, all vowels unstressed	f, s, ʃ, ʒ, x, h
	all long vowels and diphthongs stressed; all long vowels, diphthongs, and /ə/ unstressed	v, z
	Exclusion: short vowels before /v z/ not possible within the syllable, and short vowels cannot be syllable-final[b]	
C=nasal	all full vowels stressed, all vowels unstressed	m, n
	all full short vowels stressed; all short vowels unstressed	ŋ
	Exclusion: /ŋ/ cannot follow long vowels within the syllable[c] and cannot be syllable-initial	
VV diphones		
stressed–unstressed	all long vowels and diphthongs	all vowels
unstressed–stressed	all long vowels, diphthongs, and /ə/	all full vowels
unstressed–unstressed	all long vowels, diphthongs, and /ə/	all vowels
stressed–stressed	all long vowels and diphthongs	all full vowels
	Exclusion for all VV categories: short vowels cannot be V_1 because they cannot be syllable-final	
CC diphones		
C_1=voiceless stop, nasal, liquid, or glide	p, t, k, m, n, ŋ, l, r, j, w	all consonants except $C_1=C_2$ and /ŋ/
	Exclusion: /ŋ/ cannot follow a stop or another sonorant within the syllable or be an onset	
C_1=voiced stop	b	d, g, dʒ, v, z, ʒ, n, l, r
	d	b, g, v, z, ʒ, m, n, r, j, w
	Exclusions for /b d/: */bw bj bm dl/ in syllable onset, and voiced stops must devoice if not in onset unless followed by a voiced obstruent;[d] cannot be followed by /ŋ/ because /ŋ/ cannot be an onset	
	g	b, d, v, z
	Exclusions: syllable-final /g/ without devoicing is only followed by these consonants, and /g/ is never word-final[e]	
C_1=fricative	f	all consonants except f, v, ŋ
	Exclusion: /fv/ too difficult for speaker to produce without assimilation	
	s, ʃ	all consonants except s, ʃ, ŋ
	Exclusions; /sʃ/ and /ʃs/ are unlikely, unless assimilated	
	x	all consonants except x, ŋ
	v	b, d, g, z, ʒ, dʒ, n, l, r
	Exclusions: */vj vw vm/ as onsets and /v/ must devoice if not in onset	
	z	b, d, g, v, dʒ, m, n, j, w
	Exclusions: */zl zr/ as onsets and /z/ must devoice if not in onset; /zʒ/ is likely to assimilate	
	Exclusion for /v z/: cannot be followed by a voiceless fricative within the syllable, and will devoice in coda position unless followed by a voiced obstruent	

Table VI. (*Continued.*)

Class	X_1	X_2
	ʒ	w
	Exclusions: /ʒ/ never occurs syllable-finally and in onset occurs only before vowels or /w/ (e.g., in *bourgeois*)	
	Exclusion for all fricatives: /ŋ/ cannot follow a fricative within the syllable and cannot be an onset	
C_1=affricate	dʒ	m
	Exclusions: /dʒ/ never occurs word-finally, occurs syllable-finally only in the word *management*, and cannot be followed by any other consonant within an onset	
	Exclusion for all CC diphones: no geminates	

[a]CELEX does list three forms with /hə/, all based on the word *coherent*.
[b]Short vowel-/h/ diphones should be impossible, and thus should have been excluded, since short vowels cannot be syllable-final and /h/ cannot be in a coda. Also, although Booij (1995) states the prohibition of short vowels followed by /v z/ within the syllable as a phonotactic constraint, another rule in the phonology voices underlying /f s/ before a voiced stop (Booij, 1995). Thus, a short vowel can be followed by [v z] if a voiced stop follows, as in *zesde* [zɛzdə], sixth; *afdeling* [ɑvdelɪŋ], department; etc. These diphones should have been included.
[c]Although Booij (1995) states this phonotactic constraint, CELEX includes many words with long vowels followed by [ŋ]. However, the [ŋ] is always derived from underlying /n/ by assimilation to a following velar, e.g., *aangelegenheid*, affair; *woonkamer*, living room. Place assimilation in these cases tends to be optional.
[d]Booij (1995) states that coda voiced stops only remain voiced if followed by another voiced stop, not a voiced fricative or a sonorant. Since /bv dz/, etc. are unlikely onsets, these diphones, as well as /bn dm/ etc., may also be impossible. We included them since Booij mentions that some stop-fricative and stop-nasal onsets do occur in a few words. CELEX lists words with the excluded diphones /bw/ (*clubwedstrijd*, club contest), /bj/ (*objectief*, objective), /bm/ (*schrabmes*, scraping knife), and /dl/ (*woordloos*, wordless), but in all these cases the voiced stop is in coda position and should be devoiced.
[e]/gr, gl/ do occur as onsets in some loan words (e.g., *groupie*, *glamour*) and should have been included.

APPENDIX D: GATE POSITIONING CRITERIA

The following criteria were applied in establishing phoneme beginnings (B) and ends (E).

(1) Nasal: Sudden change in spectral distribution of energy (B, E).
(2) Fricative, after or before consonant: onset (B) or cessation (E) of frication.
(3) Voiceless fricative, after or before vowel: cessation (B) or onset (E) of voicing.
(4) Voiced fricative, after or before vowel: cessation (B) or onset (E) of vowel's first formant.
(5) Voiceless stop, after or before consonant: beginning of stop closure (B) or end of release burst (E).
(6) Voiceless stop, after or before vowel: cessation (B) or onset (E) of voicing.
(7) Voiced stop: beginning of prevoicing (B) or end of burst (E).
(8) Affricate /dʒ/: beginning of prevoicing (B) or end of frication (E).
(9) Trilled /r/: amplitude minimum just before first tap of trill (B), or after last tap, sometimes realized as slight burst (E).
(10) Approximant or fricative /r/: changes in formant frequencies or frication (B, E).
(11) Onset (light) /l/: sudden change in the spectral distribution of energy (B, E).
(12) Coda (dark)/l/: moment of maximum decline of energy in the first and second formants of the preceding vowel (B).
(13) Glide or /l/, after or before consonant: use criteria for the other consonant (B, E).
(14) Glide, after or before other glide or vowel: point halfway through the duration of the $F2$ transition (B, E).
(15) Vowel, after or before consonant: use criteria for the consonant (B, E).
(16) Vowel to vowel: vowels were always separated by creaky voicing, the silence of a glottal stop, or both. Boundary was set at onset of creaky voicing or silence (B, E).

As a default, gate end points were positioned at one-third, two-thirds, and the end of a phoneme. For certain phonemes in certain environments, however, the following special gate end points were used:

(1) Vowel to vowel: First gate end point for second vowel at the end of creaky voicing or silence. Third gate end point at the end of second vowel. Second gate end point halfway between the other two.
(2) Stops: First gate end point halfway through the silence or prevoicing. Second gate end point just before the beginning of the stop burst.
(3) Initial voiceless stops: only the final gate end point was used, because earlier gate end points, during the stop closure, would produce stimuli containing only silence. Therefore, diphones with a voiceless stop as the first phoneme, if recorded without preceding environment, had only four gates instead of the usual six.
(4) Voiced stops without prevoicing: In Dutch, /b d g/ are often produced without prevoicing (van Alphen, 2000). If no prevoicing was visible in the waveform at all in initial position, gate end points were placed as for a voiceless stop, producing four gates for the diphone.

Baayen, H., Piepenbrock, R., and Rijn, H. van (**1993**). *The CELEX Lexical Database (CD-ROM)* (Linguistic Data Consortium, Univ. of Pennsylvania, Philadelphia).
Booij, G. (**1995**). *The Phonology of Dutch* (Clarendon, Oxford).
Cutler, A., and Otake, T. (**1999**). ''Pitch accent in spoken-word recognition in Japanese,'' J. Acoust. Soc. Am. **105**, 1877–1888.
Dorman, M. F., Studdert-Kennedy, M., and Raphael, L. J. (**1977**). ''Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues,'' Percept. Psychophys. **22**, 109–122.
Grosjean, F. (**1996**). ''Gating,'' Language and Cognitive Processes **11**, 597–604.
Gussenhoven, C. (**1992**). ''Illustrations of the IPA: Dutch,'' J. Intern. Phonet. Assoc. **22**, 45–47.
Jongman, A. (**1989**). ''Duration of frication noise required for identification of English fricatives,'' J. Acoust. Soc. Am. **85**, 1718–1725.
Klaassen-Don, L. E. O. (**1983**). ''The influence of vowels on the perception of consonants,'' unpublished doctoral dissertation, Leiden University.
Kurowski, K., and Blumstein, S. E. (**1984**). ''Perceptual integration of the murmur and formant transitions for place of articulation in nasal consonants,'' J. Acoust. Soc. Am. **76**, 383–390.
McClelland, J. L., and Elman, J. L. (**1986**). ''The TRACE model of speech perception,'' Cognit. Psychol. **18**, 1–86.
Miller, G. A., and Nicely, P. E. (**1955**). ''An analysis of perceptual confusions among some English consonants,'' J. Acoust. Soc. Am. **27**, 338–352.
Norris, D. (**1994**). ''Shortlist: a connectionist model of continuous speech recognition,'' Cognition **52**, 189–234.
Pols, L. C. W., and Schouten, M. E. H. (**1978**). ''Identification of deleted consonants,'' J. Acoust. Soc. Am. **64**, 1333–1337.
Schouten, M. E. H., and Pols, L. C. W. (**1983**). ''Perception of plosive consonants—The relative contributions of bursts and vocalic transitions,'' in *Sound Structures: Studies for Antonie Cohen*, edited by M. P. R. van den Broecke, V. J. J. P. van Heuven, and W. Zonneveld (Foris, Dordrecht), pp. 227–243.
Smits, R. (**2000**). ''Temporal distribution of information for human consonant recognition in VCV utterances,'' J. Phonetics **27**, 111–135.
Smits, R., Ten Bosch, L., and Collier, R. (**1996**). ''Evaluation of various sets of acoustical cues for the perception of prevocalic stop consonants: I. Perception experiment,'' J. Acoust. Soc. Am. **100**, 3852–3864.
Strange, W., Verbrugge, R. R., Shankweiler, D. P., and Edman, T. R. (**1976**). ''Consonant environment specifies vowel identity,'' J. Acoust. Soc. Am. **60**, 213–224.
van Alphen, P. (**2000**). ''Does subcategorical variation influence lexical access?'' in *Proceedings of the Workshop on Spoken Word Access Processes*, edited by A. Cutler, J. M. McQueen, and R. Zondervan (MPI for Psycholinguistics, Nijmegen), pp. 55–58.
van Bergem, D. R. (**1993**). ''Acoustic vowel reduction as a function of sentence accent, word stress, and word class,'' Speech Commun. **12**, 1–23.
Warner, N. (**1998**). ''The Role of Dynamic Cues in Speech Perception, Spoken Word Recognition, and Phonological Universals,'' unpublished doctoral dissertation, Univ. of California, Berkeley.