

Sandhi processes in natural and synthetic speech

Willy Jongenburger—Vincent J. van Heuven

Abstract

Starting from the naive principle that a text-to-speech system should imitate natural speech as closely as possible, the main concern of this project is to make a stock-taking of sandhi phenomena in the natural speech of the model speaker. Next, diphone speech of the model speaker is evaluated in order to decide which sandhi phenomena are and which are not already implicitly present in the diphones. Those that are not implicitly present have to be (explicitly) accounted for by rule in the transcription module MORPHON.

1. Introduction

Let us assume, as a first approximation, that a text-to-speech system should imitate natural speech as closely as possible. Our text-to-speech system should then imitate the human sound adjustments (sandhi processes). We investigated which sandhi rules do apply in the speech of our designated talker PB, and which rules do not. For this purpose we made an inventory of the sandhi processes known to be applicable to Dutch (section 2). Next, we decided, on the basis of informal listening, to what extent each of these processes is actually manifest in the speech of our model speaker (section 3). A pragmatic criterion based on frequency of occurrence was used to decide whether a sandhi process is compulsory or not (section 4). In section 5 the distribution of the glottal stop is discussed. Section 6 deals with the question to what degree the sound adjustments are implicitly present in our diphone inventory. Finally, in section 7 we shall present our recommendations for the transcription module.

2. Sandhi processes

We shall now discuss the different types of sandhi processes. An exhaustive list of sandhi rules plus examples is included in appendix 1; Table 1 presents a summary of the rules investigated. Our primary source is Trommelen—Zonneveld (1979).

We have subdivided the various sandhi rules into the following four classes:

- I. Morpheme modification rules related to “syllable structure” modify sequences of consonants and vowels. They eliminate awkward sequences of only consonants or only vowels by insertion or elision of vowels and consonants, respectively, and generate regularly alternating CV-syllables.
- II. There are several sandhi processes that concern vowels; these are called “vowel-adjustments”. For this category of adjustments, stress is an important factor. Unstressed vowels are weak and often lose a part of their articulation (vowel shortening, vowel reduction); they can even disappear completely.
- III. In the case of “assimilation” spreading of features takes place. When assimilation takes place a phonological feature of a segment changes under the influence of its phonetic context and that segment loses its identity. For instance, an unvoiced [f] can change to a voiced [v]: *afdak* [afɖak] ‘shed’ becomes [avɖak]. Assimilation is not inevitable; the speaker can prevent assimilation by making pauses. Clearly, assimilation processes will apply sooner as speaking rate increases.
- IV. Finally, we also want to investigate the language-specific allophonic rule that turns a light [l] into a dark [ɫ] in certain environments.

3. Sound adjustments in the speech of PB, procedure and results

If we want to imitate the speech of our model speaker PB, we have to know which sandhi rules apply in his speech. For our analysis of the speech of PB we have chosen four texts (approximately 1,500 words). These texts are radio comments and news reports of which we have recordings by PB. Potential cases of sandhi rule application were marked in the texts automatically rather than by hand. For this purpose the above set of sandhi rules was incorporated into an existing letter-to-sound conversion program (Berendsen—Langeweg—van Leeuwen 1986) using the phonological rule compiler TooLip developed by van Leeuwen (1989a).

The first author then listened to the corresponding speech, and indicated for each position where a sandhi rule could apply, whether or not it did.

Before we go on to the results section some remarks are in order:

- Sandhi rules that apply across words, such as assimilation and degemination, are ignored when a prosodic pause is present between the words. Such applications of sandhi rules are left out of the statistics.
- When two sandhi rules can (theoretically) apply at the time to one segment, one of the rules will apply and the other will be blocked. In a sequence as *openkluis* [o:pəŋklɥs] ‘open safe-deposit’ the rule “n-elision after ə” is said to be blocked by the assimilation rule “n ⇒ ŋ/ — k”.
- When two different, independent sandhi rules can lead to the same spoken realization of a word, we cannot tell which sandhi rule has applied. In these cases we say rather arbitrarily that the rule with the less specific context did apply and the rule with the more specific context is blocked. We suppose, for instance, that *miljoenen mensen* [mɪljʊ'nə mɛnsə] ‘millions of people’ is the result of assimilation and degemination and not of n-elision. This means that we treat the rule “n-elision” as blocked.

In Table 1 the results are presented. By means of the acquired overview we can determine how often a sandhi rule has applied in the speech of PB, how often the rule has not applied, and how often the rule is blocked by another rule.

Table 1. Frequency of application of 32 sandhi rules in text corpus (further see text)

	applied	not applied	blocked	N	p	c/o
degemination	95.9	4.1	-	73	<.01	c+
t-elision	22.6	48.4	29.0	31	.13	c
χ-elision	-	-	-	0	-	o
n-elision <i>volgas</i>	11.1	88.9	-	9	.04	o-
schwa-insertion	11.1	88.9	-	18	<.01	-
j-insertion	25.9	74.1	-	27	.02	o
w-insertion	10.0	90.0	-	10	.02	o-
n-elision after ə	26.5	64.2	9.3	162	<.01	
glottal stop	56.2	43.8	-	397	.02	
e:-shortening	70.6	29.4	-	17	.14	c
a:-shortening	31.0	69.0	-	29	.06	o
o:-shortening	14.3	85.7	-	21	<.01	o-
i:-shortening	9.4	90.6	-	32	<.01	o-

Table 1—Continued

	applied	not applied	blocked	N	p	c/o
vowel reduction	8.0	84.0	8.0	50	<.01	o-
vowel raising	-	100.0	-	1	-	o
i-gliding <i>pja: no:</i>	-	-	-	0	-	o
i-gliding <i>so: sja:l</i>	100.0	-	-	6	.03	c+
car.ov. ass. <i>op sak</i>	79.7	20.3	-	59	<.01	c+
antic.ass. <i>azbak</i>	90.8	9.2	-	77	<.01	c+
antic.ass. <i>batpak</i>	100.0	-	-	2	.50	c
final devoicing	100.0	-	-	77	<.01	c+
n ⇒ ŋ / _ ʒ, ʒ, k	89.5	10.5	-	19	<.01	c+
n ⇒ m / _ p, b, m	91.4	2.9	5.7	35	<.01	c+
n ⇒ ɲ / _ j	100.0	-	-	1	-	c
n ⇒ ɲ / _ f, v, w	84.2	10.5	5.3	57	<.01	c+
s, j ⇒ ʃ	100.0	-	-	5	.06	c
t, j ⇒ tʃ	-	-	-	0	-	c
t, j ⇒ tʃ	100.0	-	-	6	.03	c+
p, j ⇒ pʃ	-	-	-	0	-	c
d, j ⇒ dʃ	-	-	-	0	-	c
d-weakening	50.0	50.0	-	8	1.00	o
dark ʈ	45.3	54.7	-	95	.41	o

Note: In the first column the relative frequency of the cases in which the sandhi rules do apply is indicated, in the second column the relative frequency of the cases in which the sandhi rules do not apply is and in the third column the relative frequency of the cases in which the sandhi rules are blocked. The fourth column lists the number of cases for each sandhi rule, the fifth column for each sandhi rule the value of p for the proportion applied/not applied by means of a binomial test (50 percent-50 percent), and finally the sixth column specifies whether a sandhi rule is compulsory (c) or optional (o); c+ means incontestably compulsory, o- means incontestably optional.

4. Interpretation of the results

In order to decide whether or not a sandhi rule is compulsory in PB speech, we apply two criteria:

- $p < .05$, the proportion applied/not applied should significantly deviate from the 50 percent-50 percent proportion
- one of the categories “applied” or “not applied” should contain at least 75 percent of all cases (75/25-test).

The results for 15 out of the 33 sandhi rules satisfy both conditions. This means that we have a strong cue in the decision whether or not these rules are compulsory. Rules that have been applied in at least 75 percent of all cases we call compulsory (these cases are marked in Table 1 with c+), rules that have not been applied in at least 75 percent of all cases we call optional (marked by o-).

For the remaining rules that satisfy neither condition, we will base our decision whether or not the sound adjustments are compulsory, on the consideration that PB has a rather formal speaking style. Therefore, informal sandhi rules are optional.¹ Further, we assume that, on the analogy of compulsory assimilation processes of voice and place, the sandhi rules of palatalization are compulsory as well in the speech of PB.

The rule of “t-elision in consonant clusters” does not apply in 48.4 percent, applies in 22.6 percent and is blocked in 29.0 percent of all potential cases. Research in order to find sub-regularities in the application of this rule failed. Therefore, the assumption that the articulation of [t] in consonant clusters demands rather difficult, acrobatic movements of the organs of speech must settle the matter. We expect that t-elision will apply in order to increase speech comfort.

Research in order to detect subregularities in the application of the rule “n-elision after schwa” has shown that this sandhi rule respects the ideal structure of a word, i.e., a regular alternation of consonants and vowels. When a syllable or word final /n/ is followed by another consonant, elision of /n/ will change the undesirable sequence of consonants to a more favorable sequence, for instance *kijken door* [kɛlkən do:r] ‘look through’ becomes [kɛlkə do:r]. This is exactly what happens in PB speech: the /n/ is elided within a phonological phrase before a consonant; /n/ is not elided when followed by a vowel. This means that n-elision respects the ideal structure -CVCVC-.

In order to find subregularities in the application of the rule “l ⇒ ɫ” we have taken into consideration the transcriptions of the Speech Database project (van Zanten—Damen—van Houten, this volume). These data show that an “in-between” form of [l] appears in word final position and that in non-word-final position the light [l] appears. This means that a truly dark [ɫ] hardly ever appears in PB speech. This fact is probably due to the cultivated speaking style of PB.

Glottal stop insertion is a rather special case. Although the proportion 56.2 percent applied / 43.8 percent not applied (with N=397 and p=.02) significantly deviates from a 50/50 proportion, this ratio is hardly satisfactory.

In order to discover subregularities in the insertion of the glottal stop we shall examine our database more closely in section 5.

5. Distribution of the glottal stop

Several factors prove to be of importance for glottal stop insertion. We have investigated the influence of the following factors on glottal stop insertion:

- *prosodic pause*: it appears that when a prosodic pause precedes the (morpheme-)initial vowel, a glottal stop is always inserted; within a phonological phrase no regularity is found;
- *feature voice of the preceding phoneme*: it appears that when a (by definition voiced) vowel precedes the (morpheme-)initial vowel, a glottal stop is generally inserted; when the preceding phoneme is an unvoiced consonant in general no glottal stop is inserted; when the preceding phoneme is a voiced consonant no regularity is found;
- *length of the target word and its left neighbor*: it appears that, independently of the length of the left neighbor, a glottal stop is inserted when the target word is polysyllabic; when the target word is monosyllabic no glottal stop is inserted;
- *prominence* of the syllable where the glottal stop is to be inserted: a prominent syllable is more likely to undergo glottal stop insertion than a non-prominent syllable.

The effects of the different factors on the glottal stop insertion seem to be interwoven with each other. Intricate interactions between the factors determine whether a glottal stop is or is not inserted. When we take into consideration all possible combinations of the three most important conditioning factors (prosodic pause, prominence and length of the target word), we end up with Table 2.

With the help of this table we can predict glottal stop insertion for each combination of the conditioning factors. When we predict, for instance, that a glottal stop is inserted when the target word is polysyllabic with lexical stress on the initial syllable in a phonological phrase, this prediction will be true in 61 cases and false in 15 cases. This means that in this specific condition we make a wrong prediction in 20 percent of the cases. In the table, “*” marks the correctly predicted cases and the underlined cases are the incorrectly predicted. Out of the total of 397 cases of potential glottal stop insertion we make 95 wrong predictions, i.e., 24 percent. This means that we can predict correctly whether a glottal stop will or will not be inserted in 76 percent of all potential glottal stop positions in the corpus. For a more elaborate treatment of glottal stop insertion based on the PB-corpus, the reader is referred to Jongenburger—van Heuven (1991). For a discussion of acoustic correlates of glottal stops on a more general level, see Krech (1968) for German and Malécot (1975) for French.

Table 2. Glottal stop distribution as a function of three dichotomous linguistic variables

pause	prominence	polysyllabic	+ [ʔ]	- [ʔ]	%error
+	+	+	7	0	0
+	+	-	6	0	0
+	-	+	4	0	0
+	-	-	64	0	0
-	+	+	61	15*	20
-	+	-	17*	19	47
-	-	+	15	14*	48
-	-	-	49*	126	28

Note: The first column indicates whether a pause does (+) or does not (-) precede the initial vowel; the second column whether the initial syllable is prominent (+) or not (-); the third column whether the word is polysyllabic (+) or monosyllabic (-); the fourth column indicates how often a glottal stop is inserted in the specific condition; the fifth column how often no glottal stop is inserted in the specific condition; the sixth column lists the percentage of errors that will be made by the prediction. Incorrect decisions are marked by an asterisk

6. Sound adjustments in diphones

In this stage of the project we know which sandhi rules apply in the speech of PB. Our text-to-speech converter which is based on diphone concatenation, should imitate these sounds adjustments. Since diphones contain the transitions between two sounds, we expect that sound adjustments (such as assimilation) will already be present to a certain degree in some, if not all, diphones of the PB inventory. We will have to decide for each sandhi rule whether the adjustments that are implicitly present in the diphones satisfy the listener, that is are strong enough, or whether the adjustments in the diphones are not strong enough and will have to be accounted for by explicit rules in the transcription module. We will illustrate this issue as follows. For a word such as *tentzeil* [tentsɛi] 'canvas' we do not know whether in the diphone concatenation the diphone /tz/ with implicit assimilation of voice should be used or the diphone /ts/ with explicit assimilation as brought about by the rule system MORPHON (see Nunn—van Heuven, this volume). A listening test is needed to decide this matter. For this purpose we prepared a set of stimulus words, which we

shall describe in section 6.1. In section 6.2 we will present and discuss the results of the listening test and relate the results to our earlier findings on the sandhi phenomena in PB's speech.

6.1. Stimuli and set-up of the listening test

The set of stimuli we use will have to satisfy several demands that we will discuss now.

In the first place we want to compare two versions of one and the same word: one version with implicit adjustment in the diphones, for instance *tentzeil* by means of concatenation of the diphones: /tɛ ɛn nt tz zɛɪ ɛɪ t̪ / and the other version with explicit adjustment by rule (/z/ ⇒ [s] / — t) which in our example is obtained by concatenation of the diphones: /tɛ ɛn nt ts ɛɪ ɛɪ t̪/. Other examples: the word *zeeën* /ze:ən/ 'seas' is presented in the versions /ze: ɛ:ə ən/ and /ze: ɛ:j jə ən/, the word *aanpak* /a:npɑk/ 'approach' is presented in the versions /a:n np pɑ ak/ and /a:m mp pɑ ak/.

Secondly, the stimulus words have to be presented to the listener in a carrier sentence in order to give the listener an indication of speech rate. We assume that speech rate will influence sandhi processes such that in fast, sloppy speech more sandhi phenomena will occur than in slow, carefully articulated speech. Therefore, the listener needs cues to speech rate in order to properly evaluate the stimuli in terms of pleasantness. We have chosen the carrier sentence *En toen zei ze ... tegen me* /ɛn tu:n zɛɪ zə ... tɛ:ʒən mɛ/ 'and then she said ... to me'. The target words are embedded in the middle of the utterance.

In the third place the diphones that are used for the stimuli should be taken from the set of diphones with the highest intelligibility. That is why we have chosen the set of PB diphones with 30 reflection coefficients, which are normalized with respect to duration in order to ensure that specific vowels and consonants have the same length in all the diphones in which they occur (van Bezooijen 1990; van Bezooijen—Pols, this volume). We want to note here that no integrally articulated complex consonant clusters are present in our selection of PB diphones. This means that sound adjustments such as insertions and elisions in consonant clusters such as in *melk* [mɛlək] 'milk' and *postbode* [pɔsbɔ:də] 'post man' cannot be implicitly present in the diphones and thus will always have to be accounted for by rules. For these sandhi processes we cannot make a distinction between implicit and explicit adjustments. Still we can make two versions of each stimulus word, for example /mɛ ɛɪ lk/ versus /mɛ ɛɪ lə ək/ and /pɔ əs st t̪ bɔ: ɔ:d də/ versus /pɔ əs sb bɔ: ɔ:d də/. When the listeners give a negative judgment for both versions, that is if both versions sound unpleasant, this might mean that we need integrally articulated complex consonant clusters after all.

In the case of vowel adjustments we can often make sets of stimulus words that consists of more than two versions. For example the word *minuut* 'minute' can be realized in three versions: /mi' i'n ny' y't/, /mɪ lɪn ny' y't/, /mø ən ny' y't/. Note that these adjustments, too, cannot be implicitly present in the diphones.

In some cases the diphones needed for a version do not exist. For example the word *afdak* 'shelter' can only be made with implicit adjustment in the diphones: /af fd da ak/. Because the diphone /vd/ is not contained in the diphones set, we cannot concatenate the diphones /av vd da ak/ for the version with explicit adjustment by rule.

A further possible imperfection of the diphone set is that there are no geminates. This means that we cannot generate the word *stampot* 'hutch potch' in the version [stampɔt] without degemination, but can only create the version with degemination [stampɔt]. For the complete set of stimulus words we refer to Jongenburger (1991).

Subjects of the rather informal listening test were the two authors and a colleague at Leiden University. For each set of stimulus words, subjects were asked to note their preference for one word of the set. Where a set is larger than two stimulus words, they had to rank the alternatives in terms of acceptability. In the case where there is only one stimulus word, they were asked if the realization of the word was acceptable or not.

6.2. Results of the listening test

We want to compare the results of the present experiment with our earlier findings on the sandhi phenomena in PB speech. For this purpose we present the data of the listening test and information on sandhi in the speech of our designated talker in Table 3.

Table 3. Subjects' preferences for each sandhi rule and application of sandhi rules in natural PB speech

sandhi rule	subjects' preference in diphone speech	natural PB speech
degemination	+/*	+
t-elision	+	+
χ-elision	+!	-
n-elision before s	+!	-

Table 3—Continued

sandhi rule	subjects' preference in diphone speech	natural PB speech
schwa-insertion	-	-
homorganic glide insertion:		
j-insertion after i'	-	-
j-insertion after e:	+!	
w-insertion after y'	-	-
w-insertion after o:	+!	-
n-elision after schwa	+	+
a:-shortening	+!	-
o:-shortening	-	-
e:-shortening	+	+
i'-shortening	-	-
vowel reduction to schwa	-	-
schwa-elision	-	-
vowel raising	+!	-
i-gliding [pja:no:]	-	-
i-gliding [so:fa:lə]	*	+
assimilation of voice:		
v ⇒ f in a [1 stress] syll.	+	+
v ⇒ f in a [≥ 2 stress] syll.	+	+
z ⇒ s in a [1 stress] syll.	+	+
z ⇒ s in a [≥ 2 stress] syll.	+	+
f ⇒ v	-	+
s ⇒ z	-	+
p ⇒ b	-	+
χ ⇒ ʃ	-	+
k ⇒ g	-	+
t ⇒ d	-	+
homorganic nasal adjustment:		
n ⇒ ŋ / _ χ	+	+
n ⇒ ŋ / _ k in a [1 stress] syll.	-	+
n ⇒ ŋ / _ k in a [≥ 2 stress] syll.	+	+
n ⇒ m / _ p	-	+
n ⇒ m / _ b	+	+
n ⇒ m / _ m	-	+
n ⇒ ɱ / _ f	-	+
n ⇒ ɱ / _ v	-	+
n ⇒ ɱ / _ w	-	+

Table 3—Continued

sandhi rule	subjects preference in diphone speech	natural PB speech
homorganic nasal adjustment:		
n ⇒ ŋ / _ j	+	+
palatalization:		
t- j ⇒ c	+	+
p- j ⇒ ɸj	-	+
d- j ⇒ ʤ	*	+
s- j ⇒ ʃ	*	+
d-weakening	-	-

Note: The left column identifies the sandhi processes. The column in the middle gives the preferences of the subjects for each sandhi process; the symbol '+' means that the subjects prefer the adjustment to be explicit (by rule), '-' means that they prefer the adjustment to be implicit, '*' means that required diphones are lacking, '!' means discrepancy between natural and diphone speech. The right column indicates for each sandhi rule whether in PB speech the rule does or does not apply.

It appears from Table 3 that sound adjustments by assimilation of voice implicitly present in the diphones are strong enough when these processes are anticipatory. In case of carryover assimilation, however, the adjustments need to be explicitly accounted for by rule; here the adjustments implicit in the diphones are not strong enough.

The homorganic nasal adjustments implicitly present in the diphones are satisfactory with the exception of the processes:

- n ⇒ ŋ / _ ɣ
- n ⇒ ŋ / _ k before a [1 stress] syllable
- n ⇒ m / _ b
- n ⇒ ŋ / _ j

Subjects prefer these adjustments to be explicitly accounted for by rule. Observe that such anticipatory homorganic nasal adaptation is sufficiently reflected in the diphones when the nasal occurs in a syllable with primary stress with the conditioning segment following in a secondarily stressed syllable (e.g. *'in'koop* ⇒ [ɪn'ko:p] 'purchase'). However when the nasal occurs in a syllable with secondary stress followed by the conditioning phoneme in a syllable bearing primary stress, the strength of the assimilation is considered insufficient, (e.g., [ɔn'kʌys] *onkuis* 'without chastity'). Clearly, a precise, stressed, articulation of a speech sound makes it resistant

to the assimilation process. When we assume that listeners are aware of this fact, they expect less or weaker adjustment in primary stressed than in secondary stressed syllables. This explains why the relatively weak implicit adjustment in primary stressed syllables does satisfy the subjects whereas the implicit adjustment in secondary stressed syllables is insufficient.² While selecting the stimulus material for this experiment little attention was paid to this matter but it is interesting to investigate this matter in greater detail in the future.

For the processes of palatalization we note that two types of adjustments cannot be accounted for by the present diphone set. Stimulus words which contain palatalization phenomena of the type $dj \Rightarrow j$ and $sj \Rightarrow f$ are judged unsatisfactory by the subjects. In both cases only the versions with adjustment by rule are presented to the subjects; the versions with implicit adjustment in the diphones cannot be made because the diphones /dj/ and /sj/ are absent from the set of diphones. This also means that i-gliding of the type [so:fa:lə] cannot be realized with the current diphone set. It seems that the diphone set should be expanded with the diphones /sj/ and /dj/.

It appears from Table 3 that the subjects prefer the elision processes to apply with the exception of schwa elision. The process of weakening should not be applied. Further, we see that shortening of /a:/ and /e:/ are evaluated as acceptable, whereas shortening of /o:/ and /i:/ are evaluated as unacceptable. For the other vowel adjustments we see that schwa insertion, vowel reduction to schwa, and i-gliding should not apply in order to get acceptable speech but that vowel raising should apply.

Looking at the results for the homorganic glide insertion we find that explicit insertions are required only after [-high] vowels (e:, o:) and not after a [+high] vowel (i:, y:).

We mentioned in section 6.1 the absence of integrally articulated complex consonant clusters in our diphone set. From our data in Table 3 we conclude that processes such as insertions and elisions in consonant clusters (t-elision, χ -elision, n-elision before s, schwa insertion and schwa elision) can be accurately accounted for by the diphone set. Target words where degemination has taken place, seem to be less acceptable with respect to temporal organization. A duration rule that doubles the duration of the plosive might solve the problem. So we draw the conclusion that there is no great need for integrally articulated complex consonant clusters.

Now let us compare these results with what actually happens in PB speech (articulation rate of the diphone speech is comparable to the articulation rate of PB "live": (5.7 syll/s versus 5.4 syll/s, respectively). We see that some sandhi rules do not apply in the speech of PB while application of these very rules is preferred by the subjects (marked in Table 3 by '!').

Apparently, we cannot just copy speech characteristics from our model speaker to less than optimal quality speech such as diphone speech. The recordings of the speech of PB were made under favorable conditions so that the speaker knew that the intelligibility of his speech is optimal. Diphone speech, however, is not optimally intelligible and therefore requires different speech features. This has implications for the acceptability of synthetic speech based on the sandhi phenomena such as they appear in PB's speech. This observation leads us to the decision to make two different sets of sandhi rules so that the user of the text-to-speech system can choose which set to use. One set includes all the sandhi rules in the way that PB applies them, leading to a close imitation of PB, the other set of rules leads to more acceptable diphone speech, at least as far as sandhi is concerned, and is not necessarily based on what PB does. This set of rules includes the former set extended with the rules preferred by the subjects.

7. Recommendations for the transcription module

In order to obtain synthetic speech that imitates natural PB speech as closely as possible, the following set of sandhi rules should be used:

- t-elision in consonant clusters
- n-elision after schwa before a consonant (within a phonological phrase)
- e:-shortening
- carryover assimilation of voice
- homorganic nasal adjustment:
 - n \Rightarrow ŋ / _ χ
 - n \Rightarrow ŋ / _ k before a [1 stress] syllable
 - n \Rightarrow ŋ / _ b
 - n \Rightarrow ɲ / _ j
- palatalization:
 - t-j \Rightarrow c
- final devoicing
- glottal stop under specific conditions (see section 4).

This set can be extended by the following rules to obtain more acceptable diphone speech:

- χ-elision
- n-elision before s
- j-insertion after e:
- w-insertion after o:
- a:-shortening
- vowel raising

Appendix: Survey of sandhi rules

Modification rules related to syllable structure

- Degemination
reduction of two identical consonants to one occurs in Dutch whenever two identical consonants are adjacent.
stampot /stamppɔt/ ⇒ [stɑmpɔt] ‘hotchpot’, *misstap* /misstap/ ⇒ [mɪstap] ‘misstep’
- t-Elision in consonant clusters
postbode /pɔstbɔ:də/ ⇒ [pɔsbɔ:də] ‘postman’, *kastje* /kastjə/ ⇒ [kɑsjə] ‘cupboard’
- χ-Elision
schrijven /sχrɛivən/ ⇒ [srɛivən] ‘write’
schroeven /sχru:vən/ ⇒ [sru:vən] ‘screws’
- n-Elision before s
every n after schwa and before s can be elided.
volgens /vɔlχəns/ ⇒ [vɔlχəs] ‘according to’, *wegens* /we:χəns/ ⇒ [we:χəs] ‘because of’
- Schwa-insertion after /l/ and /r/ and before another tautosyllabic consonant
arm /ɑrm/ [ɑrəm] ‘arm’, *lantaarn* /lɑnta:rn/ ⇒ [lɑnta:rən] ‘lantern’, but *armen* /ɑrmən/ *[ɑrəmən] ‘arms’
- Homorganic glide insertion
between a stem-final vowel (not schwa or /a:/) and a following suffix that begins with a vowel, a glide (/w/ after a back vowel and /j/ after a front vowel) is inserted in order to prevent a succession of vowels.
drieën /dri:ən/ ⇒ [dri:jən] ‘threes’, *kanoën* /ka:nɔ:ən/ ⇒ [ka:nɔ:wən] ‘to canoe’
- n-Elision after schwa
particularly in western dialects of Dutch syllable-final /n/ after schwa can disappear, except when the context of /n/ is /ə _ də/.
ramen /ra:mən/ ⇒ [ra:mə] ‘windows’, *eigendom* /ɛiχəðɔm/ ⇒ [ɛiχəðɔm] ‘property’, but *volgende* /vɔlχəndə/ ⇒ *[vɔlχəðə] ‘next’
- Glottal stop
a glottal stop is inserted when a vowel is in hiatus position, for instance at the beginning of a word, in the sequence [schwa - unreduced vowel]

or in the sequence of two identical vowels. Rules are unclear in the literature.

aap /#a:p/ ⇒ [ʔa:p] ‘monkeys’, *beamen* /bəa:mən/ ⇒ [bəʔa:mən] ‘to assent’, *moe oefent* /muː uːfənt/ ⇒ [muːʔuːfənt] ‘mum is practising’

Vowel adjustments

— a:-, o:-, e:-, i-shortening

these are the optional processes that lead to a shortening of vowels in non primary stressed initial syllables. Shortening only applies to long vowels in open syllables, followed by a consonant in the next syllable.

banaan /ba:na:n/ ⇒ [bana:n] ‘banana’, *politie* /pɔ:liːtsi/ ⇒ [pɔliːtsi] ‘police’, *melaats* /me:lɑ:ts/ ⇒ [mlɑ:ts] ‘leprous’, *minuut* /miːnyʊt/ ⇒ [mɪnyʊt] ‘minute’

— Vowel reduction to schwa

this rather informal rule applies to unstressed vowels.

Canada /ka:na:da:/ ⇒ [ka:nəda:] ‘Canada’, *economie* /e:ko:no:mi:/ ⇒ [e:kəno:mi:] ‘economics’, *salaris* /sa:lɑ:rəs/ ⇒ [səla:rəs] ‘salary’

— Vowel raising

if the vowel /e:/ is followed by another vowel or by a glide (as a result of homorganic glide insertion), then /e:/ is raised to /i/. This rule only applies to unstressed non-initial syllables.

ideaal /iːde:a:l/ ⇒ [iːdiːa:l] ‘ideal’, *Koreaan* /ko:re:a:n/ ⇒ [ko:riːa:n] ‘Korean’

— i-Gliding

a non-initial, unstressed [iː] loses its syllabic properties between an obstruent and a vowel.

kopieer /ko:piːe:r/ ⇒ [ko:pje:r] ‘copy’, *Aziaat* /a:ziːa:t/ ⇒ [a:zja:t] ‘Asian’, *sociaal* /so:siːa:l/ ⇒ [so:sja:l] ‘social’

Assimilation processes

— Carryover assimilation of voice

fricatives are unvoiced if preceded by an obstruent.

op zak /ɔp#zak/ ⇒ [ɔpsak] ‘in ones pocket’, *zegt veel* /zɛχt#ve:l/ ⇒ [zɛχt fe:l] ‘says a lot’, *afzien* /af#ziːn/ ⇒ [afsiːn] ‘give up’

— Anticipatory assimilation of voice: [-voice] ⇒ [+voice]

obstruents are voiced if followed by a voiced plosive.

opdiene /ɔpdiːnən/ ⇒ [ɔbdiːnən] ‘serve’, *asbak* /usbak/ ⇒ [azbak] ‘ash tray’

- Anticipatory assimilation and final devoicing
all syllable final obstruents are unvoiced.
broodplank /bro:dplɑŋk/ ⇒ [bro:tplɑŋk] ‘bread-board’, *bed* /bɛd/ ⇒ [bɛt] ‘bed’
- Homorganic nasal adjustment
the place of articulation of the nasal becomes identical to the place of articulation of the following consonant.
bilabial:
/n/ ⇒ /m/ / _ p,m,b *onbewust* /ɔnbəwœst/ ⇒ [ɔmbəwœst] ‘unconscious’
velar:
/n/ ⇒ /ŋ/ / _ k,χ,g *inclusief* /ɪnklyˈziːf/ ⇒ [ɪŋklyˈziːf] ‘inclusive’
labiodental:
/n/ ⇒ /ɱ/ / _ f,v,w *onwel* /ɔnwɛl/ ⇒ [ɔɱnwɛl] ‘ill’
palatal:
/n/ ⇒ /ɲ/ / _ j *onjuist* /ɔɲjʌyst/ ⇒ [ɔɲjʌyst] ‘incorrect’
- Palatalization
the place of articulation of plosives, fricatives and nasals becomes identical to the place of articulation of [j] (palatal).
/t/ dental ⇒ palatal [c] in *weet je* /wɛːt jə/ [wɛːcə] ‘do you know’
/p/ bilabial ⇒ palatal [pʲ] in *heb je* /hɛp jə/ [hɛpʲə] ‘do you have’
/d/ alveolar ⇒ palatal [dʲ] in *djatie* /djati/ [dʲati] ‘teak’
/s/ dental ⇒ palatal [ʃ] in *was je* /was jə/ [wɑʃə] ‘were you’
- d-Weakening
the obstruent [d] becomes sonorant after a long vowel or a diphthong and before a schwa.
rode /ro:də/ ⇒ [roːjə] ‘red’, *houden* /hɑvdə/ ⇒ [hɑʋwə] ‘keep’
- l ⇒ ɫ
a word initial “light” [l] sounds different from a word final “dark” [ɫ]
leren [leːrə] ‘learn’, *dal* [da ɫ] ‘valley’

Notes

1. It appears from Table 1 that seven sandhi processes hardly occur in the text: for these rules no p-value is given. We assume that the informal sound adjustments are optional in PB’s speech.
2. Note that no such discrepancy is found for carryover assimilation of voice.