Schwa-assimilation in Danish Synthetic Speech

Jensen, Christian

Publication date:
2001

Document version
Peer reviewed version

Schwa-assimilation in Danish Synthetic Speech

Christian Jensen

Department of General and Applied Linguistics
University of Copenhagen, Denmark
chrjen@cphling.dk

Abstract

Assimilation of schwa into surrounding sonorant consonants is a vital feature of natural Danish speech. It varies with speaking rate and speaking style and is more likely to occur in some phonological contexts than in others. This presents some problems for the implementation of the process into a Danish text-to-speech system.

1. Introduction

Assimilation of unstressed schwa ([ə]) into surrounding vowels or sonorant consonants is a process which entered into standard Copenhagen Danish around the middle of the 19th century and has been evolving ever since. See [1] for a fairly thorough description of the process and its historical development. It is found in words such as *pige, katten, bade, stemme*, "girl, the cat, bathe, voice": ['pi:i], ['kadn̩], ['bæ:ð̩], ['sdɛmm̩] (transcriptions are modified, broad IPA throughout the paper. See [2] for details).

As these examples suggest ə-assimilation can be described as a process whereby ə is replaced by the most sonorant adjoining sound, which then takes over the syllabic function of the schwa. All the above words will normally be perceived as bisyllabic by Danish speakers − even in their assimilated forms. This is achieved through a lengthening of the sound which assimilates ə, and, when following a stressed syllable, through the special tonal pattern which characterises a stressed plus unstressed syllable(s) in Danish (see [3]). However, ə can also be affected after voiceless consonants in words such as *huse, hoppe*, "houses, (to) hop": ['hu:s], ['hʌb].

In the latter case ə is simply elided, normally without any compensatory lengthening of the preceding consonant (or of the stressed vowel). That is, this phenomenon is strictly speaking syncope of ə, but it is normally treated as a special case of ə-assimilation. This means that assimilation or elision of ə can in principle occur in (more or less) any position where we find ə.

2. The conditions for ə-assimilation

There are, however, some restrictions on this principle in most natural speech. First of all, ə-assimilation is not equally likely to occur in all phonological contexts. Secondly, ə-assimilation is highly dependent on speaking rate, or tempo, and speaking style. That is, the faster the tempo and the more casual the speaking style the more likely ə-assimilation is to occur. It should be noted, though, that ə-assimilation is found even in careful, read speech. And thirdly, there seem to be some lexico-semantic considerations in operation too, so that some words which we would expect to have ə-assimilation in a given speaking rate and style seem to maintain ə. I will go through each of these three factors and how they affect the implementation of ə-assimilation rules into a Danish text-to-speech system, the DST (*Dansk Syntetisk Tale*) Project,[1] henceforth just DST.

2.1. Speaking rate and speaking style

Although we know that ə-assimilation varies with speaking rate and style, the precise nature of this variation is still unexplored. Fortunately, we do not need to mimic this complex variation in DST. We decided to make ə-assimilation and other reductions a feature of speaking style, and to vary tempo independently. In fact, the three different styles that we operate with are defined by having different degrees of reduction and assimilation. While this may not be in total agreement with the variation found in natural speech, it does make the system very flexible. The three style levels are distinguished mainly by the phonological contexts in which we allow ə-assimilation to occur.

2.2. Phonological contexts

ə-assimilation is (almost) never an obligatory process, in the strictest sense of the word. There are phonological contexts where ə-assimilation is very likely to occur in all speaking styles and others which seem to require more casual speech. If we regard ə-assimilation in different contexts as different processes it is possible to list these processes in a more or less hierarchical system, and then to select a specific speaking style simply by choosing which of these processes to implement in the phonetic rule complex. I will outline these different processes or contexts below.

There is one other process that deserves special attention, namely when combinations of /ə/ and /r/ result in [ʌʁ] or [ʁʌ] because of the lowering/opening effect of the [ʁ], or in [ʌ] through a complete merger of the two sounds. Ex:

raseri:	/'raːsəriːˀ/	['ʁɑːsʌʁiːˀ]	rage
kasse:	/'kasə/	['kasə]	box
kasser:	/'kasər/	['kasʌ]	boxes

This is not an instance of ə-assimilation although there are some similarities (mainly that it involves ə), but it does have an effect on ə-assimilation. Since this process is obligatory the relevant changes have been entered into the DST lexicon, making "r-coloured" ə identical with the stressed vowel ʌ in *holde*: ['hʌlə] since the two are phonetically indistinguishable. However, as we shall see later, there are some rules which impose restrictions on the assimilation of the *second* ə in a word, and this includes ə following an r-coloured ɔ: [ʌ].

---

[1] This project was commissioned by the Danish Ministry of Information Technology and Research and is being developed by the Department of General and Applied Linguistics, University of Copenhagen, Center for PersonKommunikation, Aalborg University and Tele Danmark.

2.2.1. ə *in bisyllabic words*

This word structure, with stress on the first syllable, is prototypical of Danish, and is suitable for illustrating the variation in ə-assimilation depending on context.

After long vowels
In this position ə-assimilation is more or less obligatory in all speaking styles:

pige:	['piːə]	['piːi]	girl
due:	['duːə]	['duːu]	pigeon

When there is one intervening, or medial, consonant we find the following.

Between an obstruent and a sonorant (very common)

katten:	['kadən]	['kadn̩]	the cat
nissen:	['nesən]	['nesn̩]	the elf
cykel:	['sygəl]	['sygl̩]	bicycle
hoppet:	['hʌbəð]	['hʌbð̩]	jumped

After semi-vowels (very common)
Note that ð, in narrow IPA [ð̞ʸ], is treated as a semi-vowel in our project.

have:	['hæːwə]	['hæːw̩]	garden
leje:	['lɑjə]	['lɑjj̩]	rent
bade:	['bæːðə]	['bæːð̩]	bathe

After other sonorant consonants (common)

tale:	['tæːlə]	['tæːl̩]	speak
ville:	['vilə]	['vill̩]	would
komme:	['kʌmə]	['kʌmm̩]	come
penge:	['pɛŋə]	['pɛŋŋ̩]	speak

After voiceless consonants/obstruents (less common)
This is possible, but it is clearly indicative of a slightly more casual speaking style.

huse:	['huːsə]	['huːs]	houses
løbe:	['løːbə]	['løːb]	run
hoppe:	['hʌbə]	['hʌb]	hop

Most of these examples involve words with the simple structure CVCə, that is ə after one consonant. If ə follows consonant clusters we see a slightly different picture.

After two semi-vowels (very common)
This is (almost) always a combination of [ʌ̯] and [w].

værge:	['væʌ̯wə]	['væʌ̯w̩]	guardian

After semi-vowel plus sonorant (less common)

tegne:	['tɑjnə]	['tɑjn̩]	draw
vidne:	['viðnə]	['viðn̩]	witness

After two sonorants (less common)
This is clearly more marked, that is mostly found in a more relaxed speaking style.

gamle:	['gɑmlə]	['gɑml̩]	old
emne:	['ɛmnə]	['ɛmnə]?	topic

Note the potential merging of *gamle* (pl.) and *gammel* (sg.). ə-assimilation sounds extremely marked or even impossible in the word *emne*, which may have to do with the two adjoining nasals.

After sonorant plus obstruent (not common)
ə is sometimes assimilated in this context, but it does not seem equally acceptable in all words.

hjælpe:	['jɛlbə]	['jɛlb]	help
danse:	['dansə]	['dans]	dance

Both are natural, but mostly found in casual speech, whereas

femte:	['fɛmdə]	['fɛmd]*	fifth

does not seem possible, although there is nothing in the phonological structure which separates it from the other two examples. The constraint must therefore be lexical or semantic.

2.2.2. ə *between two sonorant consonants*

An interesting question arises when we have ə between two sonorant consonants, namely: which of them will become syllabic in words like *kanden*, *føden*, *foldet*, *manden*: ['kanən, 'føːðən, 'fʌləð, 'manˀən] , "the pot, the food, folded, the man"?

The governing principle is that the syllabicity goes to the most sonorant of the two consonants, according to this sonority scale:

w j (>) ð > l > m n ŋ > consonants with *stød*

If the two consonants are equally sonorant the second one becomes syllabic, according to [4]. Thus

kanden:	['kanən]	['kann̩ː]	the pot
føden:	['føːðən]	['føːð̩n]	the food
foldet:	['fʌləð]	['fʌlð̩]	folded
manden	['manˀən]	['manˀn̩]	the man

2.3. Complications

Those were the simple and regular examples. In slightly more complicated structures things become less clear. If, for example, there is more than one ə in a word the second ə cannot assimilate to a preceding sonorant:

malende:	['mæːlənə]	['mæːlnə]	graphic
malerne:	['mæːlʌnə]	['mæːlʌnə]	the painters

Not [mˈæːlˌn]* and ['mæːlʌn]*. Note that this is also true when the first (underlying) schwa has merged with /r/ as in the second example above.

There are certain exceptions: For instance after [ð] as in

ammede:	['ɑməðə]	['ɑmð̩ð̩]	breast-fed

where [ð] can take over the syllabicity of both ə's. And in a word like *yngelen*: ['øŋˀələn] "brood" there is some disagreement about how the second schwa is assimilated. According to [4] it assimilates to the following [n]: ['øŋˀl̩n] whereas [5] claims that both ə's assimilate to (the more sonorant) [l]: ['øŋˀl̩l̩n]. This can be very difficult to hear, and is unlikely to have any significance for our synthetic speech.

Now let us examine an example similar to *malende* above, namely *hundene*. Not only is the second ə "protected" against assimilation to the preceding [n]. This [n] also seems to be protected against taking over the syllable from the first ə. According to one of the earlier rules ə assimilates to the most sonorant of the adjoining consonants, or to the following consonant if they are equally sonorant, as in

hundene:	['hunənə]	dogs

but it is still my intuition (confirmed by students and phoneticians at various lectures and phonetics classes) that the first [n] becomes syllabic: ['hunn̩nə]. And if we choose a word like

kommende:	['kʌmənə]	['kʌmm̩nə]	coming

the different places of articulation of the two nasals makes this impression much stronger. In fact, the principle even holds in a word like

forstemmende: [fʌ'sdɛmʔənə] [fʌ'sdɛmˀn̩nə] depressing

despite the fact that the [m] has *stød*, and therefore is less sonorant than the [n]. This implies that there is some kind of boundary before the [nə] ending over which ə-assimilation cannot apply.

The nature of this boundary is unclear, however. It does not match the most commonly accepted phonological syllable boundary, which would be after the second [n] (see [2]), nor does it coincide with morphological boundaries. The three relevant suffixes are *ende* (pres. part.), *ne* (def. plural) and *ene* (def.+plural), which have (two) different phonological structures. Postulating some kind of phonetic syllable boundary as a help in formulating the assimilation rules would not be helpful either. First of all, no syllable boundaries are defined in the DST project, and trying to introduce them would be to create massive problems, and secondly, there are plenty of examples where ə-assimilation does apply across syllable boundaries. In fact, in normal speech it can apply across word boundaries ([1:192]), but this is not implemented in DST.

There are even situations where the boundary before the [nə] ending can be crossed, cf.

kattene: ['kadənə] ['kadn̩(n)ə] the cats

Here the boundary is transgressed, and a very interesting phenomenon arises where the n-sound is both clearly syllabic because of the ə-assimilation and initial in the following syllable.

2.4. Sporadic exceptions

While the above complications at least seem to be governed by some sort of regularity, we also find many words where ə-assimilation is not easily acceptable even though the phonological context meets the criteria, for example

Hanne: ['hanə] ['hann̩]? girl's name

is very problematic, whereas

kande: ['kanə] ['kann̩] (tea) pot

is very natural. This might be a general resistance towards ə-assimilation in certain word types, e.g. proper names, but then how do we explain

Gilleleje: [gilə'lajə] [gil̩'laj̩j̩] name of town

where ə-assimilation is completely natural?

The word *Gilleleje* is interesting for another reason: it has ə before the stressed syllable, that is, pretonic ə. Many words with this structure seem to resist ə-assimilation, for example *general, generel* [gɛnə'ʁɑːˀl, gɛnə'ʁalˀ] "general (n.), general (adj.)". A closer inspection of words with pretonic ə indicated the following possibilities for ə.

Can assimilate to a following sonorant in the same syllable

hottentot: [hʌdən'tʌd] [hʌdn̩'tʌd] Hottentot
appelsin: [abəl'siːˀn] [abl̩'siːˀn] orange (n.)

Can assimilate between two sonorants when the second sonorant is followed by a clear syllable boundary

angelsaksisk: [aŋəl'sagsisg] [aŋl̩'sagsisg] Anglo-Saxon
lidenskabelig [liːðən'sgæːˀbəli] [liːðn̩'sgæːˀbəli] passionate

ə *will not normally assimilate to a sonorant which is initial in the following syllable*

avenue: [avə'ny] [avn̩'ny]* avenue

bakelit: [bagə'lid] [bagl̩'lid]* Bakelite

In some words with similar structure it is possible, though

lutenist: [ludə'nisd] [ludn̩'nisd] lute player

The most likely explanation why it works better in this word is that the two surrounding consonants are homorganic, which makes dropping the vowel easier. Most, if not all, other words with that structure in our lexicon can have ə-assimilation, such as *fastelavn, kotelet, satellit, tartelet* "Shrovetide, cutlet, satellite, patty shell", and those that seem more dubious are very infrequent words, like *neurasteni, kandelaber* "neurasthenia, candelabrum", which in itself makes them much less prone to reduction.

Between two sonorants, when the second sonorant is initial in the following (stressed) syllable

Rågeleje: [ʁåːwə'lajə] [ʁåːw̩'laj̩j̩] name of town
bimmelim: [bemə'lem] [bemm̩'lem]* crazy
Annelise: [anə'liːsə] [ann̩'liːsə](?) girl's name

The first one is unproblematic. ə assimilates to the preceding sonorant, which is the more sonorant of the two. In the second example ə-assimilation would be unusual although the word is not all that infrequent. And in the last example − *Annelise* − there is some uncertainty about which of the sonorants will become syllabic. According to the principle of higher sonority, the syllabicity should go to the [l], and this is indeed how the word is listed in [5], but this is counterintuitive to all the phoneticians that were polled on the question in an informal investigation. So if we accept the above notation as the "correct" or normal one, it seems that prosodic factors − here the position of the stressed syllable − influences ə-assimilation: if the following syllable is stressed, ə-assimilation becomes predominantly regressive.

3. Implementation

It should be clear by now that the principles that govern ə-assimilation in Danish are very complex. Not only can it sometimes be difficult to know which sound ə is assimilated to, but the question of whether the assimilation will take place at all is determined by many factors: *speaking style, tempo, phonetic structure, prosodic structure, frequency of occurrence* and possibly various semantic consideration. The exact nature of the interplay of these features is not very well known, and when you add idiosyncratic preferences to the equation you will probably find that no two people would produce the same assimilations in even a fairly short stretch of speech.

Implementing such a complex mechanism in a text-to-speech system is of course not a viable option, and in the DST project we have set a much more modest goal. We have decided on three different speaking styles, or *levels of distinctness*, from *very distinct*, which is in effect without any assimilations except the almost obligatory assimilation to a preceding long vowel, *distinct*, which includes a couple of the more common ə-assimilations, and *less distinct*, with a fuller set of assimilation processes, although still much more restricted than most natural speech. The three levels simply consist of three different sets of phonetic rules, and the level can be selected by the user of the program. Below, I will give a brief sketch of the implementation of the *less distinct* level.

3.1. Coding

The DST prosody module, that is, duration and F0 assignment and any modifications to the phonetic transcription, is

programmed in SPL (Synthesis Programming Language [6]). This language allows for SPE-style formulations of rules based on feature definitions of phones, making it easy to write phonetic/phonological rules. It does, however, lack many of the more common features of general programming/scripting languages, such as the more conventional *if...then* conditional statements. All rules in an SPL program are sequentially ordered, which means that the output of one rule "feeds" the next rule. Since there are no ordinary conditional functions, all restrictions in the application of a rule are accomplished by giving a precise feature specification of the phone which is to be changed and of the context in which the change is to apply. All ə-assimilation rules are paired: first the relevant sonorant consonant is marked as [+syllabic] (or, more precisely, the scalar value *syllabic* is incremented) and then the ə triggering this is deleted, thus preventing it from serving as input to a succeeding ə-assimilation rule. The compensatory lengthening of the syllabic sonorant is handled later in the rule complex and will not be described here.

Because of the sequential/ordered nature of the rules it is convenient and sometimes necessary to place the more specific rules first and the more general rules later. All the rules below are written in a shorthand format and not in formal SPL (or SPE). They are meant for illustration only.

The sonority level of all phones is set using a *scalar* value *sonor*. Sonorant consonants have values between 6 and 8. Syllabicity is assigned through the scalar value *syllabic*, initially set to "1". A scalar, and not a binary feature, is used because the consonant [ð] can assimilate two schwas and thus carry two syllables. Further conventions:

V = any vowel	C = any consonant
* = any number	→ ∅ = is deleted
? = Danish *stød*	() = optional

First the second ə in a word must be marked, so that it does not get assimilated:

(1) ə → [+flag] / ə ([-syl]*) ___ [-syl] #

Before a [nə] *ending, assimilate to preceding sonorant:*
(2a) [+cons, sonor>5] → syllabic+=1 / ___ (?) ə n [ə, +flag]
(2b) ə → ∅ / [+cons, sonor>5, syllabic>1] ___ n [ə, +flag]

This ensures that ə will assimilate correctly, namely to the first consonant, in words such as *tilkommende* ['telkʌmˀənə] ['telkʌmˀn̩nə] "due (adj.)" where a later, more general, rule would assimilate ə to the following, more sonorant consonant.

After an obstruent, assimilate to a following sonorant:
(3a) [+cons, sonor>5] → syllabic+=1 / [+obs] ə ___
(3b) ə → ∅ / [+obs] ___ [+cons, sonor>5, syllabic>1]

This takes care of examples such as *bussen, katten*. They could also have been handled by the following rule, which assigns syllabicity to the most sonorant of two consonants surrounding a ə (but separating them makes the rule complex more flexible):

(4a) [+cons, sonor>5 && sonor>sonor(+2)] →
syllabic+=1 / V (:) ([+voc, +cons]) (?)
__ [+cons, sonor>5]
(4b) [+cons, sonor>5 && sonor>=sonor(-2)] →
syllabic+=1 / V (:) ([+voc, +cons]) (?)
[+cons, sonor>5] ə __
(4c) ə → ∅ / [+cons, sonor>5] __ [+cons, sonor>5]

The expressions (+2) in rule (4a) and (-2) in (4b) refer to the position of another phone, so (+2) means the phone two places to the right. The first part of rule (4a) then reads: *A sonorant consonant which is more sonorant than the consonant two places to the right (following the ə) becomes syllabic in the context....* This is a good demonstration of how to make conditional statements in SPL.

Make sonorants syllabic between a (full) vowel and ə:
(5a) [+cons, sonor>5] → syllabic+=1 / V (:) (?) __ ə
(5b) ə → ∅ / V (:) (?) [+cons, sonor>5, syllabic>1]

There are separate rules to handle assimilation to *stød*-bearing sonorants, and finally there are rules to handle assimilation of the second schwa in a word (which was "protected" in rule (1)) in the places where assimilation should take place.

Assimilate "protected" schwa to a preceding ð *or any following sonorant:*
(6a) ð → syllabic+= / __ [ə, +flag]
(6b) [ə, +flag] → ∅ / [ð syllabic>1] __
(7) [+cons, sonor>5] → syllabic+=1 / [ə, +flag]
(7b) [ə, +flag] → ∅ / [+cons, sonor>5, syllabic>1]

Implementing all these rules in DST yields very good overall results. A list of words, in their assimilated and unassimilated forms, plus accompanying sound files can be found on the Internet address:
http://www.cphling.dk/pers/chrjen/eurospeech2001.

Some things cannot be handled through phonetic rules, namely the, lexically and/or semantically based exceptions to the general principles. At the moment, this problem is partially solved by marking the relevant words in our lexical database with a special symbol for schwas that resist assimilation, for example in the personal names *Anne, Hanne, Line*, and (some) words with pretonic schwa such as *generel*. Exceptions that are based on semantic/pragmatic considerations or even on frequency of occurrence cannot be handled within the present framework.

The general level of acceptability to the listener of the assimilation mechanism has not yet been tested formally, but will be evaluated together with the overall performance of the system. These tests will give us an indication of the acceptability and desirability of reductions and assimilations in general, but the implementation of ə-assimilation into our system has revealed a need for two further investigations: 1) a diagnostic listening test of the acceptability of ə-assimilation in selected Danish words in different contexts, and 2) a large empirical investigation of the occurrence of ə-assimilation in natural, scripted and unscripted speech.

4. References

[1] Brink, L. and J. Lund, *Dansk rigsmål. Lydudviklingen siden 1840 med særligt henblik på sociolekterne i København*, Gyldendal, 1975.

[2] Grønnum, N., *Fonetik og fonologi. Almen og dansk*, Akademisk Forlag, 1998.

[3] Thorsen (Grønnum), N., "Selected problems in the tonal manifestation of words containing assimilated or elided schwa," *ARIPUC* 16, pp. 37-100, 1982.

[4] Thorsen (Grønnum), N. and O. Thorsen, *Fonetik for sprogstuderende*, Institut for fonetik, Københavns Universitet, 1988.

[5] Brink, L., J. Lund, S. Heger, and J. N. Jørgensen (eds.), *Den Store Danske Udtaleordbog*, Munksgaard, 1991.

[6] Holtse, P. and A. Olsen, "SPL: a speech synthesis programming language," *ARIPUC* 19, pp. 1-42, 1985.