Syllable deletion in contemporary Danish

Anja Schüppert[1], Nanna Haug Hilton[1], Charlotte Gooskens[1] & Vincent J. van Heuven[2]

[1]University of Groningen, [2]Leiden University

Abstract

Danish has been described as a language exhibiting particularly many reduction phenomena, a development which might lead to impaired intelligibility. This paper quantifies syllable deletion and articulation rate in spoken Danish and investigates their effects on the intelligibility of Danish to native speakers. In a crossed-design, sentences in four experimental conditions were presented to native speakers of Danish in a translation task. The conditions were (i) quick and accurate (few syllable deletions) speech, (ii) quick and inaccurate speech, (iii) slow and accurate speech, and (iv) slow and inaccurate speech. The results reveal that slow and accurate speech is most intelligible, while quick and accurate speech is least intelligible to native speakers. The findings are discussed in the light of auditory and articulatory theoretical frameworks of speech production and perception.

1 Introduction

In a document called *Sprogpolitik for DR* ("Language policies for the Danish Broadcasting Corporation"), the Danish state-owned radio broadcast Danmarks Radio (DR) declares that "in DR programmes, the Danish languages should be comprehensible without subtitles" (Danmarks Radio 2009). This was stated as a reply to a statement supposedly made by Christoph Bartmann, the former head of the German Goethe-Institut in Copenhagen, who claimed that Danes frequently asked each other "What did you say?". Danmarks Radio gives a precise description of pronunciation which is to be used on the broadcasts and which should ensure that "syllable cannibalism, wrong vowel colours and muddy consonants" were avoided. This is a rather concrete statement compared to the guidelines in other European countries. Norsk Rikskringkasting (NRK) in Norway recommends "clear [language] with natural stress and colloquial sentence structure" (Språkrådet 2007), Sveriges Radio (SR) in Sweden strives for "comprehensible, living and correct" language (Göransson & Lundin 2012), and the British Broadcasting Corporation (BBC) in the UK instructs their speakers to use "clear, precise language" (British Broadcasting Corporation 2012). So why does Danmarks Radio bother to define preferable speech used by news readers to such a degree?

Danish has been described as a language particularly difficult to decode for speakers of its two most closely related languages Norwegian and Swedish (Maurud 1976, Bø 1978, Delsing & Lundin Åkesson 2005, Schüppert & Gooskens 2011, Schüppert 2011), as well as hard to learn for adults (Grønnum 2003) and even challenging to acquire for young children (Bleses et al. 2008). Bleses et al. (2008) suggested that the fact that Danish consonants are frequently vocalised, which results in long vocalic stretches boosts or causes the observed delay in language production in children. In line with Grønnum (2007), they suggested that a particularly high number of reduction and assimilation processes in Danish together with schwa-deletion processes make the Danish sound structure unclear with weak, or even no, cues for word and syllable boundaries. The assumption is that this large number of reduction and deletion processes makes it more difficult for Danish children to discover where a word or a syllable ends and where the next begins, and that this makes it more difficult for them to acquire the language compared to, for example, Norwegian and Swedish children.

Speech reduction is closely linked to an increased speaking tempo (Engstrand & Krull 2001). It is difficult to establish, however, which of the two variables is cause and which is effect. An increased speaking tempo demands more articulatory activation and therefore, a high speaking tempo generally leads to less accurate speech, while the shortening or deletion of segments and syllables allows the speaker to complete an utterance in a shorter time span. Both variables (reduction and speaking tempo) have been shown to influence intelligibility. A high speaking tempo impairs intelligibility (Fairbanks & Kodman 1957, Fairbanks et al. 1957, Foulke & Sticht 1969, Vaughan & Letowski 1997, Krause & Braida 2002, Gordon-Salant et al. 2007, Jones et al. 2007) and a high number of syllable deletions (as one of several possible manifestations of reduced speech) have a detrimental effect on intelligibility (Ernestus et al. 2002, Janse 2004; Janse & Ernestus 2011).

In this paper, we measure syllable deletion in spoken Danish and investigate the influence of syllable deletion and articulation rate on the intelligibility of spoken language. More specifically, we are interested in the relative influence of syllable deletions and increased articulation rate for the intelligibility of Danish.

2 Research question

The aim of this paper is threefold. Firstly, we aim at quantifying one particular feature of reduction in spoken language, namely syllable deletion. In order to do so, we measure the number of phonetic syllables, here operationalized as the number of sonority peaks in speech signals from Danish news broadcasts. This number is compared to the number of canonical syllables. These measurements are reported in section 3.

Secondly, we aim at investigating the influence of the two factors syllable deletion and articulation rate on the intelligibility of spoken Danish to native speakers. We hypothesise that an increased articulation rate and/or a higher number of deleted syllables result in less accurate speech, and thereby hamper intelligibility. This hypothesis is tested experimentally by comparing intelligibility of slowly and accurately pronounced stimuli with the intelligibility of quickly and inaccurately pronounced stimuli.

Thirdly, we investigate the relative importance of these two factors for the impairment of intelligibility of unclear speech. In order to do so, we analyse intelligibility of speech across four different conditions, namely slow and accurate speech, slow and inaccurate speech, quick and accurate speech, and quick and inaccurate speech. The latter two parts are reported in section 4.

3 Acoustic analysis - Radio news measurements

3.1 Material and speakers

Our materials consist of radio news broadcasts read by professional news readers. They were recorded in a highly controlled setting. Our findings will therefore not necessarily reflect reduction and articulation rates found in more colloquial speech. An advantage of using these data is, however, that recordings from professional news readers are produced in a way that is intelligible to a large community of listeners. This is supported by the language policy guidelines cited in section 1.

The corpus was compiled of recordings made by the state-owned Danish radio station DR on stations P1 and P4. A total of 10 minutes and 51 seconds of fluent speech was used for the analysis. The length of the recordings varied between 29 and 45 seconds. The data were produced by 19 speakers, 10 of whom were male. The news broadcasts were all aired in the spring of 2010 and had been recorded with speakers who use a standard-like accent – the Copenhagen regiolect.

3.2 Measurements

While speech rate is defined as the number of entities produced per time unit including pauses, articulation rate is based on the number of entities produced per time unit without pauses. A pause, again, is often defined as a silent interval in the speech signal which lasts at least 150 (Tsao and Weismer 1997) or 200 milliseconds (Campione and Véronis 2002). However, Kendall (2009) defines a pause as a silent interval with a duration of at least 60 milliseconds. Silent intervals preceding a noise burst during production of a plosive are excluded by this definition. This means that while pauses in speech are considered as being part of the signal for speech rate, and therefore included in the measurement, articulation rate is a measure of the amount of articulatory activity within a time frame. In this study, we are concerned with articulation rate and follow Tsao and Weismer's (1997) suggestion with minimum duration of 150ms for pauses. This definition is motivated by their claim that 150ms is longer than the typical silent interval during production of plosives, while probably being at the lower end of a range of what constitutes a meaningful pause (Tsao and Weismer 1997:861).

When calculating articulation rate, the produced units (such as words, syllables or phonemes) are counted and their number is divided by the duration of the utterance which was analysed. The most common way to determine articulation rate is calculating the number of syllables produced per second (e.g. Kowal et al. 1983; Den Os 1988; Almberg, 2000; Verhoeven et al. 2004). To enable comparison of our results with previous studies, we measure articulation rate in the same way here. Importantly, however, we make a distinction between phonetic (actually produced and acoustically measurable) syllables and phonological (canonical syllables). This is explained in greater detail below. All sound recordings analysed in this investigation were transcribed orthographically in Praat (Boersma & Weenink 2009).

3.2.1. Canonical articulation rate

We assume that the number of canonical syllables is basically reflected in orthography. For example, the Danish word *lærere* ('teachers'), which has three orthographic syllables and three canonical syllables according to Grønnum 2007, is pronounced disyllabically /lɛːʌ/ in normal speech (Molbæk Hansen 1990, Hjorth & Kristensen 2003). Allegedly, the word had three syllables when borrowed from Low German, as still reflected in contemporary Danish orthography, and as still observed in the two most closely related languages Norwegian and Swedish. In East Norwegian (a language which has not been standardised), the word *lærere* can be pronounced /lærərə/, and in Swedish, the word *lärare* is pronounced /lærarə/ (Hedelin 1997). Both in East Norwegian and in Swedish, these pronunciations can be reduced to /læɾːə/ or a similar pronunciation, but /lærərə/ and /lærarə/ are completely unmarked forms.

Obviously, canonical syllables are frequently reflected in orthographic rules and norms. However, native speaker intuitions about what constitutes a canonical syllable might again also be influenced by knowledge of the orthographic system. This means that Danish orthography is not only likely to reflect ancient pronunciation, but might also shape the idea in literate Danes that the word *lærere* /lɛːʌ/ has three underlying (or canonical) syllables, although all three are unlikely to be pronounced in colloquial speech. In careful speech, however, the word is likely to be produced with three syllables.

The duration of the recordings was established for each individual speaker, and canonical syllables were counted manually by a native speaker. Any pauses in the speech signal with durations of more than 150 ms were removed after the transcriptions and prior to counting the syllables. The number of canonical syllables was subsequently divided by the utterance duration to obtain articulation rate of canonical syllables, or canonical articulation rate.

3.2.2. Phonetic articulation rate

To calculate the articulation rate of phonetic syllables, the number of acoustic sonority peaks actually produced in the speech signal was determined automatically. We follow De Jong & Wempe's (2009) definition of a phonetic syllable, which is defined as an intensity peak (having an intensity of at least 2 dB higher than the surrounding signal) in the voiced part of the speech signal, i.e. where F_0 can be measured. To obtain individual phonetic articulation rates per speaker, the number of syllables produced by each speaker is divided by the duration of the analysed sample for this speaker. The automated count of acoustically realised syllables makes it possible to establish the number of syllables without human interference. Human knowledge of phonology and underlying syllable structure is likely to influence the ability to identify phonetic syllables in a speech signal. More specifically, humans are prone to detect syllables that are not actually produced and therefore count more phonetic syllables than a computer algorithm does (De Jong & Wempe 2009). However, De Jong and Wempe (2009) also report that automatic and human syllable counts correlate significantly ($.71 < r < .88$). This indicates that human and automatic syllable detection are highly congruent processes, but it remains unclear whether the algorithm detects slightly too few syllables, or humans detect too many syllables as they are subconsciously influenced by their phonological and/or orthographic knowledge.

An example of the output of the automatic procedure is shown in the upper three panels in Figure 1. The top panel of the figure shows the oscillogram of the fragment *hvad synes du om forældrenes initiativ i Hørsholm* ('what do you think about the parents' initiative in Hørsholm'). The central panel shows the cochleagram with intensity (black line) and pitch (white line). The panel underneath shows a grid that with automatically marked sonority peaks. The bottom panel shows an manually added orthographic transcription for each detected peak.

It can be seen that the fragment contains 17 canonical syllables (hvad.sy.nes.du.om.for.æl.dre.nes.i.ni.ti.a.tiv.i.Hørs.holm), but only 12 intensity peaks are detected in those parts of the signal that have voicing, as can be seen in the syllable tier displayed underneath the cochleagram.

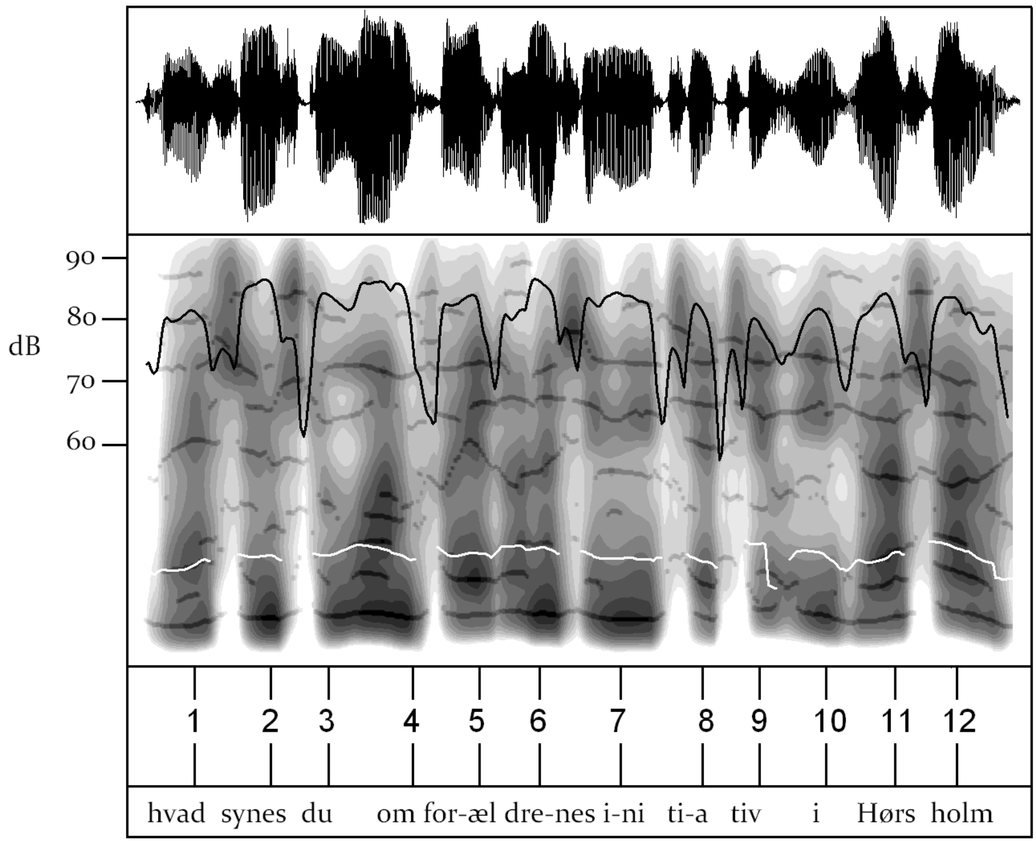

FIGURE 1. Oscillogram (upper panel), cochleagram (central panel) with intensity curve (black line) and pitch contour (white line) as well as syllable tier (lower panel) for the fragment *hvad synes du om forældrenes initiativ i Hørsholm* ('what do you think about the parents' initiative in Hørsholm') as read by a female news reader.

The syllable count of an utterance is equal to the number of sonority peaks, i.e. phonetic syllables detected in the utterance. The number of phonetic syllables was subsequently divided by the utterance duration to obtain articulation rate of phonetic syllables, or phonetic articulation rate.

3.2.3. Syllable deletion rate and syllable deletion percentage

The syllable deletion rate (SDR) is based on the two measures number of canonical syllables (N_c) and number of phonetic syllables (N_p). It is calculated using the equation SDR = 1 - (N_p / N_c), which yields a value between 0 and 1, since generally, $N_p < N_c$. By multiplying the SDR value by 100, a syllable deletion percentage (SDP) is obtained. The SDP of the fragment displayed in Figure 1 is 29% as 5 out of 17 canonical syllables are not measurable in the acoustic signal.

3.3 Results

Figure 2 displays phonetic and canonical articulation rate for the analysed 19 news readers in a box plot. The line in the middle of the boxes is the median speech rate while the boxes represent the middle two quartiles. The whiskers above and below the boxes cover 95% of the values. It can be seen that the number of sonority peaks (or phonetic syllables as detected by the automatic analysis) is smaller than the number of canonical syllables as counted by the native speaker. This strongly suggests that some of the canonical syllables are reduced in actual speech to such a degree that the automatic analysis cannot detect them anymore. Consequently, we regard these syllables as being deleted from the actual pronunciation. Importantly, this does not always mean that the complete segmental and suprasegmental information in this syllable has been dropped. Often, two or more canonical syllables are merged into one longer syllable. An example of this is the word *initiativ* ('initiative'), which contains five canonical syllables, which are merged into three phonetic syllables (syllables nr. 7, 8 and 9 in Figure 1). It can be seen that the first two canonical syllables are merged into one long phonetic syllable.

The mean number of phonetic syllables produced per second in our material is 4.4 (*sd* = 0.3), while the same material contains 6.2 (*sd* = 0.4) canonical syllables per second. This means that, on average, 1.8 of the canonical syllables per second were not produced. This corresponds to an SDP of 29%. Interestingly, this is a much higher SDP than found for the two languages that are the most closely related to Danish, viz. Norwegian and Swedish (Hilton et al. 2011). Hilton et al. (2011) reported that

Norwegian and Swedish news readers delete no more than one canonical syllable per second on average.

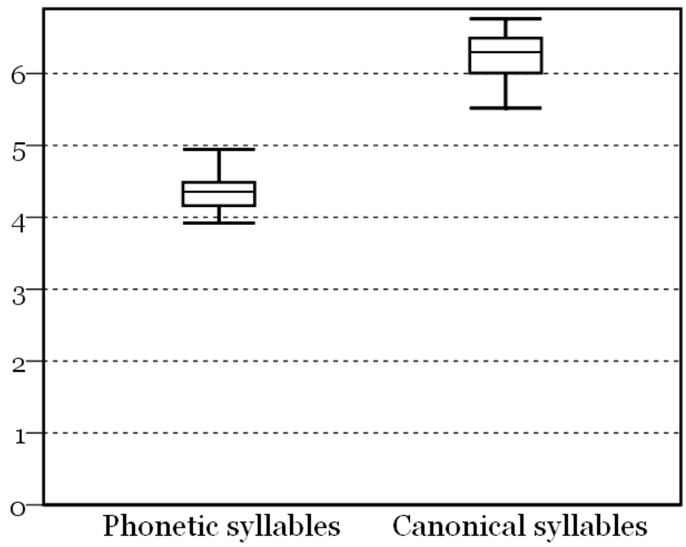

FIGURE 2. Mean phonetic and canonical articulation rate (number of acoustically realised and orthographically defined syllables per second, respectively) for 19 Danish news readers.

The large discrepancy between the number of phonetic and the number of canonical syllables produced per second indicates that contemporary Danish is characterized by a large degree of syllable deletion – much more so than the neighbouring languages Norwegian and Swedish (Hilton et al. 2011). Crucially, the high number of reduction and deletion processes is linked to an increased articulation rate. It is unclear which of the two factors articulation rate and syllable deletion is cause and which is effect, as a high degree of deletion makes it possible to pronounce more canonical syllables per second, while at the same time, this high articulation rate might lead to the high number of deletions. We were interested in whether syllable deletions and a high articulation rate impairs intelligibility and whether one of these two factors is linked more strongly to intelligibility than the other, i.e. whether increased intelligibility can be achieved by slow speech, or by accurate speech, or only by slow *and* accurate speech. To tease the two factors articulation rate and syllable deletion apart, we conducted an intelligibility experiment. This is reported in the following section.

4 Intelligibility experiment

4.1. Stimulus material

The stimulus material consisted of 55 semantically unpredictable sentences (henceforth SUS) that were read aloud by a native speaker of Danish in three different conditions: (i) at a slow speaking rate with a deliberately accurate pronunciation, (ii) at a 'normal' rate, as well as (iii) at a high speaking rate with less accurate pronunciation. Versions (i) and (iii) were manipulated so as to form two additional conditions (iv) and (v). These will be explained in greater detail in the following section.

The SUS were generated by the method developed by Benoît et al. (1996). These sentences are generally used in sentence intelligibility experiments because it has been shown that they yield very accurately differentiating intelligibility scores. The SUS are syntactically correct sentences but consist of phrases with concepts that are not likely to be semantically related to each other (cf. Gooskens et al. 2010 for a more detailed description of the material). Sentences consisting of semantically unrelated concepts can be assumed to measure intelligibility more reliably, as every word has to be decoded separately and cannot be derived from the context.

SUS can be automatically generated using basic syntactic structures and a number of frequently occurring short words. The syntactic structures are simple and vary in length having between 10 and 18 canonical syllables (either 6 or 7 words). The sentence length does not exceed seven words in order to avoid saturation of the listeners' short-term memory. An example of a SUS is given in (1). The entire set of sentences can be found in the Appendix.

(1)	Danish	*En*	*amerikansk*	*regering*	*studerer*	*et*	*ansvar.*
	English	'An	American	government	studies	a	responsibility.'

The following lexical categories were used to construct the sentences:

- nouns
- transitive verbs (trans. verb)
- intransitive verbs (intrans. verb)
- adjectives (adj)
- relative pronouns (rel pron)
- prepositions (prep)
- conjunctions (conj)
- question-words (quest)
- determiners (det)

These word classes were used to implement the following sentence types:

- Intransitive structure: det + noun + intrans. verb + prep + det + adj + noun
- Transitive structure: det + adj + noun + trans. verb + det + noun
- Interrogative structure: quest + trans. verb + det + noun + det + adj + noun
- Relative structure: det + noun + trans. verb + det + noun + rel pron + intr. Verb

For each lexical category, there are special restrictions. The most important restrictions are the following:

- verbs: no auxiliaries and reflexives, only present tense (including the imperative in S3)
- nouns: only singular forms
- adjectives: only forms which can be used attributively, no comparative and superlative forms
- prepositions: only single-word prepositions
- determiners: only indefinite forms

All words were selected randomly from the thousand most frequent words in their lexical category using the published database Korpus90, which lists words in terms of their token frequency in a text corpus of 28 million words from various kinds of written texts (available at http://korpus.dsl.dk/e-resurser/k90_info.php?lang=dk).

To preclude any repetition priming, each content word appeared just once in the whole set of stimulus sentences used, although some lexemes appeared in different word classes. Function words such as *en* (indefinite article common gender), *et* (indefinite article neuter gender), *og* ('and') and *som* (relative pronoun) were allowed to occur more often.

4.1.1. Recordings

The stimulus sentences were read aloud in three versions (see above) by a female native speaker of Danish and recorded in a sound-attenuated room at the University of Groningen. The three versions were (i) slowly and accurately, (ii) 'normal', and (iii) quickly and less accurately. The speaker was instructed to produce the quick and the slow sentences without sentence-intern prosodic boundaries, i.e. without any pauses. Mean and range of sentence length for quickly and slowly produced sentences are visualized in Figure 3. It turned out that the slowly produced sentences were 1.67 times as long as the quickly produced ones. The normal sentences were faster than the slowly produced ones but slightly slower than the quickly produced sentences.

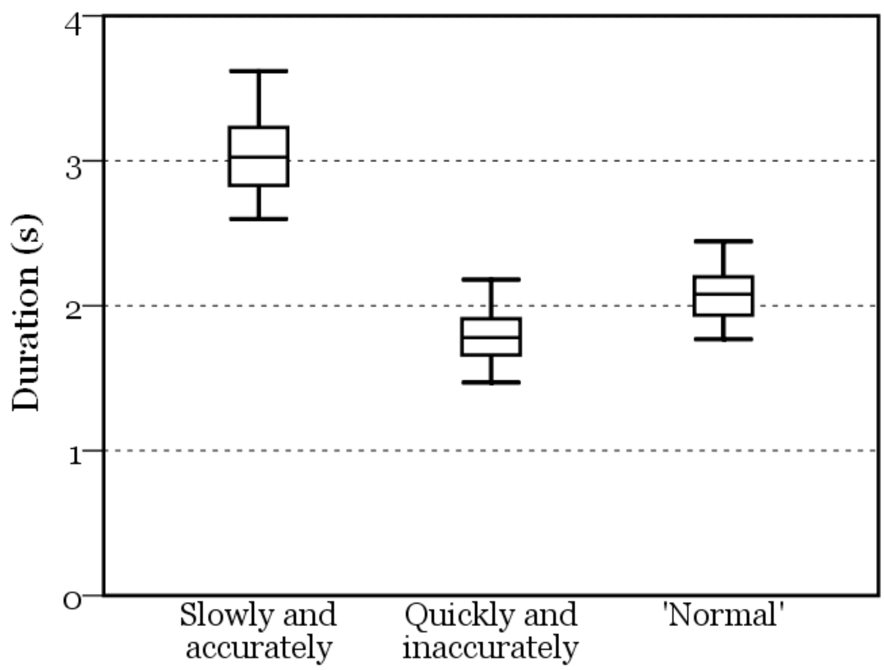

FIGURE 3. Duration (seconds) for slowly and accurately, quickly and inaccurately, as well as normally produced sentences.

The sentences had between 10 and 18 canonical syllables (mean = 13.1). In the slow mode, the sentences were produced with 10 to 18 phonetic syllables (mean = 13.0), while the same sentences read in the quick mode had 5 to 12 phonetic syllables (mean = 8.7). The slow recordings were indeed produced very accurately, as hardly any canonical syllables were deleted in actual pronunciation, which confirms that syllable deletions in Danish are not phonological. In the quick mode, one third (namely 33.6%) of the syllables were deleted. Not surprisingly, as this represents a particularly high tempo, this is an even higher percentage than in our news readers' corpus (SDP 29%, see section 3.3).

FIGURE 4. Phonetic and canonical articulation rate for the slowly and quickly produced recordings of the stimulus material.

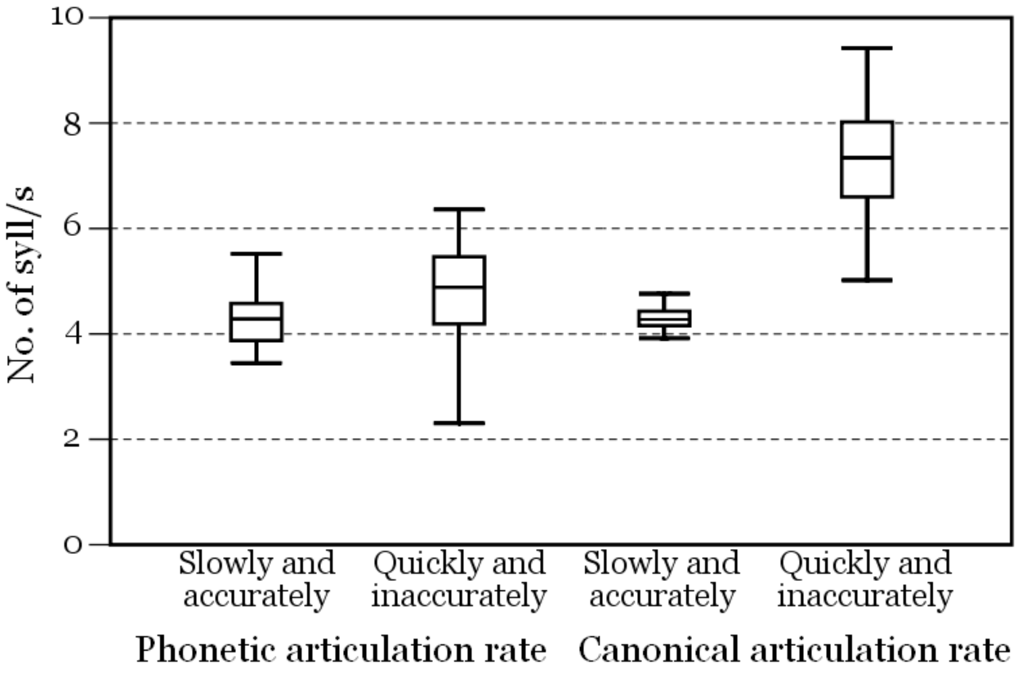

Figure 4 displays phonetic and canonical articulation rate for the material used in the experiment. It can be seen that the phonetic articulation rate differs less across the two modes than the canonical articulation rate. The speaker produced 4.3 (slow mode) and 4.8 (quick mode) phonetic syllables (i.e. sonority peaks) per second. The difference in phonetic articulation rate across the two modes is significant ($t(54) = 4.4$, $p < .001$), but less striking than the difference in canonical articulation rate. As the number of canonical syllables is constant across the two modes, canonical articulation rate in the quick mode is naturally much faster (mean = 7.3) than in the slow mode (mean = 4.3). This difference is highly significant ($t(54) = 26.8$, $p < .001$). Both measures together suggest that the speaker tended to adjust the number of intensity peaks and, thereby, cues for word recognition to the total duration that she had at her disposal. More time to produce an utterance leads to more accurate pronunciation, while a reduced amount of time results in a large number of deleted syllables. The SDP for the material used is 0.1% for the slowly produced sentences and 33.2% for the quickly produced sentences.

4.1.2. Manipulation

The material was manipulated in two ways. First, the slowly produced sentences were time-compressed linearly by reducing the total duration to the duration of the same sentence produced quickly. Similarly, the quickly produced sentences were time-expanded by increasing the total duration to the duration of the same sentence produced slowly. That means that duration manipulation was performed on each sentence individually. On average, sentence duration was expanded from 1.8 to 3.0 seconds to create slow but inaccurately articulated sentences, and compressed from 3.0 to 1.8 seconds to create quick and yet accurately articulated sentences. The factors for duration manipulation were on average 1.67 and 0.6, respectively.

It is unknown (for any language and certainly for Danish) how a human speaker adapts the sentence melody when speaking rate is increased or decreased. Pitch movements may either be time-compressed (faster rate of F_0 change), or reduced in excursion size, or they may be incompletely realized (either through truncation or through gestural overlap, see Caspers & Van Heuven 1993, Ladd 1996). To ensure that the manipulated sentences would not sound less natural than the unmanipulated sentences, all sentences were monotonised with a fixed F_0 of 213 Hz, which was the mean F_0 employed in the original recordings. This procedure makes all stimulus sentences equally unnatural, which restricts their general intelligibility, but, importantly, it keeps the prosody *equally* unnatural across all five conditions. Duration and pitch manipulations were performed by the PSOLA (Pitch Synchronous Overlay and Add) analysis-resynthesis technique (e.g. Moulines and Verhelst 1995), as implemented in Praat.

4.2 Design and task

The participants were asked to fill in a short questionnaire providing information about their background, age, sex and place of residence. They were also asked to indicate which language they spoke with their parents at home. The purpose of this was to be able to include only the results of listeners who had a native level of language proficiency of Danish.

After questionnaire completion the intelligibility experiment started. Prior to the actual experimental session, the listeners were presented with five training sentences to get used to the task. The five sentences were not analysed. After the training session the experiment started. Each participant listened to 50 sentences, 10 in each condition. The order of the sentences was randomised across conditions, but kept constant across participants within one group. The same sentence was only presented once to each listener. Sentences were blocked by condition and rotated over listener groups according to a complete Latin-square design (cf. Box, Hunter & Hunter 1978).

4.3 Participants

The 42 participants were divided into five groups of either eight or nine participants. Each group took part in one of the versions of the experiment. All participants were from Odense on the island of Funen in the central part of Denmark. They all attended secondary school at a level that would admit entrance to university after completion. Listeners who did not speak Danish with at least one of their parents were excluded from the analysis. The mean age of the remaining participants was 18.7 years and their age ranged between 16 and 22 years. Sixteen participants were male. None of the subjects reported hearing problems.

4.4 Results and discussion

The intelligibility score per participant was determined by calculating the percentage of correctly translated content words. The mean scores for all five conditions are given in Table 1. The two conditions with short duration (quick and accurate, and quick and inaccurate speech) are the least intelligible to the participants. It is not surprising that the slow and accurate conditions are the most intelligible. In other words, the intelligibility of normal speech is increased by speaking slowly and accurately, while speaking quickly and/or inaccurately impairs intelligibility.

TABLE 1. Percentage range and mean of intelligibility scores per condition.

	Minimum (%)	Maximum (%)	Mean (%)	Std. Deviation
Short and inaccurate	50.0	100.0	85.2	11.4
Short and accurate	48.6	100.0	83.8	11.8
Long and inaccurate	53.6	100.0	88.8	11.5
Long and accurate	82.1	100.0	94.6	5.1
Normal	57.0	100.0	89.6	10.7

Obviously, all five conditions were highly intelligible to the participants. There is a clear ceiling effect in the data, as the maximum score is 100 percent in all five conditions, while the minimum score is not even close to 0 percent. The mean scores range between 80 and 95 percent. Therefore, we applied the arcsine transform (Studebaker 1985) on our data, which yields intelligibility scores that are numerically close to the original percentages over most of the percentage range while retaining all of the desirable statistical properties. Subsequently, we ran a one-way ANOVA with the transformed intelligibility scores as dependent variable and condition as factor. It revealed that condition had a significant effect (*F* (4)

= 8.3, $p < .001$) on intelligibility. A Bonferroni post-hoc test showed that the intelligibility of the 'normal' condition did not differ significantly from the four experimental conditions (all $p > .05$). This is not surprising, given the fact that the 'normal' condition lies somewhere in between the four experimental conditions both with respect to duration and with respect to pronunciation accuracy. Intelligibility scores of only two of the ten combinations showed significant differences: The slowly and accurately produced sentences are more intelligible than the quickly and accurately produced sentences, as well as the quickly and inaccurate produced ones (both $p = .001$).

Subsequently, we ran a repeated-measures ANOVA with duration (short or long) and pronunciation accuracy (accurate or inaccurate) as factors. The results are shown in Table 2. The normal condition is disregarded in this section, as we were interested in the relative influence of reduction and duration on intelligibility. The analysis revealed that there is a main effect of duration as well as an interaction between duration and accurateness, but no main effect of accurateness. The mean intelligibility scores as a function of these two factors are displayed in Figure 5.

TABLE 2. Summary of Repeated Measures Analysis of Variance.

	F	***df1***	***df2***	***p***
Duration	19.4	1	49	< .001
Accuracy	1.7	1	49	.20
Duration * accuracy	5.3	1	49	.03

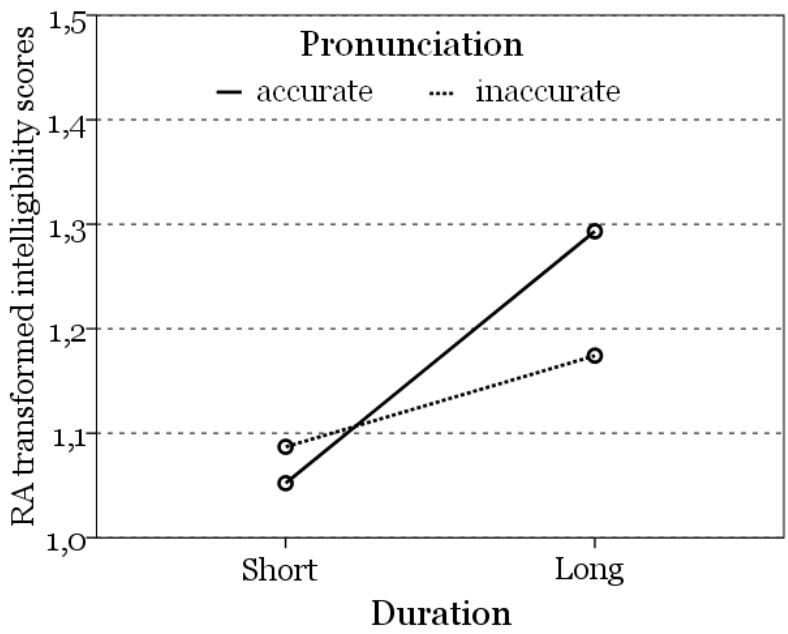

FIGURE 5. Intelligibility (RA transformed percentage of recognised content words) for slow and fast utterances with and without syllable deletions.

Together with Table 2, Figure 5 shows that speaking accurately only improves intelligibility if the duration of the utterance is expanded. An accurate pronunciation without practically any syllable deletion (see section 4.1.1.) with a short duration, on the other hand, impairs intelligibility. Interestingly, this type of speech was significantly less intelligible in our experiment than inaccurate pronunciation in a short utterance. In other words, speaking quickly and accurately is not more intelligible than speaking quickly and inaccurately.

5 Conclusions

5.1 Acoustic analysis - Radio news measurements

In section 3, we quantified the amount of syllable deletion in Danish news readers' speech by measuring the number of canonical syllables and the number of sonority peaks in read speech. It turned out that more than a quarter of the canonical syllables, namely 29% were not detected as sonority peaks in the speech signal, which we interpret as syllable deletions. Compared with its neighbouring languages Norwegian and

Swedish, Danish deletes more canonical syllables from the actual speech signal. Hilton et al. (2011) report that 1.0 phonetic syllables are deleted by East Norwegian news readers, and 0.9 by Swedish news readers. The fact that almost twice as many syllables are deleted in a comparable corpus of Danish news readers might be one of the reasons why Danish is the language which is generally most difficult to understand for Scandinavians (Maurud 1976, Bø 1978, Delsing & Lundin Åkesson 2005, Schüppert 2011). It is also likely that this large number of syllable deletions in colloquial Danish is associated with the delayed language development reported by Bleses et al. (2008).

5.2 Intelligibility experiment

In section 4, we investigated whether an increased articulation rate, a high number of deleted syllables, or both result in less intelligible speech. To study this, we compared the intelligibility of slow and accurate, quick and accurate, slow and inaccurate and quick and inaccurate speech. By this, we could also draw conclusions as to whether one of the two variables accurateness and duration plays a larger role for the impairment of intelligibility of unclear speech.

Generally, the intelligibility scores for all five conditions were rather high. Recalling that our compressed speech was only 1.67 times as fast as the slow speech, this is not surprising, given the finding by Zemlin et al. (1969), who reported that intelligibility of time-compressed speech is hardly hampered as long as the speaking tempo is twice the normal tempo at the most. This finding is also in line with Fairbanks et al. (1957), who found that increasing the speaking tempo to double rate reduced the intelligibility score to only 90% of uncompressed speech.

However, it is a well-documented phenomenon that fast speech is less accurate than slow speech. Gay (1978) found that stressed vowels were shortened when articulated at a high speaking rate, while unstressed vowels were not only shortened, but also produced with a lowered intensity, a lowered F_0 and, partly, within a smaller vowel space. Engstrand (1988) reported that vowels and consonants were produced with more co-articulation when produced at a higher speaking rate than when produced at a lower rate. Crystal & House (1990) showed that the main predictors of syllable duration are the number of phones per syllable and the proportion of stressed phones. Increased co-articulation in fast speech is thus a form of reduction needed by the articulatory organs to produce speech at this pace.

The Hyper- and Hypospeech Theory (Lindblom 1990) states that a speaker producing fast speech preserves those parts of the speech signal that are most important for the intelligibility of the utterance. For example, stressed syllables are reduced less than unstressed syllables (Peterson & Lehiste 1960, Port 1981). The H&H theory supposes that the preservation of the most important parts in the speech signal is to be seen as a compromise between articulatory and communicative constraints: the amount of articulatory activity in humans is restricted, which is why an utterance is produced with the greatest articulatory effort as possible articulatorily, but with the lowest as necessary for intelligibility. Our data indicated that a long duration enhances intelligibility, while accurate pronunciation only increases intelligibility if the duration is prolonged. In other words, speaking accurately (i.e. pronouncing all syllables) but very quickly does not improve intelligibility. It is likely that this effect is due to the increased phonetic articulation rate. In this condition, there are the most phonetic syllables produced per second, while the quick and inaccurate condition is likely to omit exactly those syllables that tend to be redundant in casual speech. Our results contradict findings by Janse (2004), who reported that normal speech that was linearly compressed to a rate of 8.5 syll/s was associated with significantly faster reaction times in a phoneme detection task compared to naturally fast speech. More specifically, our results seem to suggest that the economising of articulatory activity in fast speech, manifested through a large number of syllable deletions, not only is due to articulatory constraints, but might as well increase intelligibility of the utterance. If less important parts of the signal are deleted in fast speech, this yields a phonetic articulation rate which is closer to more accurately produced speech at a normal tempo. This might have a beneficial effect on intelligibility.

To sum up, our data suggest that articulation rate and reduction processes are closely connected not only for articulatory reasons, as it is easier to produce many syllables if there is more time than if the utterance has to be compressed into a short duration, but also for auditory reasons – it is easier to decode speech if a specific phonetic articulation rate is not exceeded. If duration is short, it is easier to decode inaccurate speech with syllable deletions, at least if the deletions occur at natural places, i.e. if they concern syllables that carry less information than the preserved syllables.

The obvious paradox which our data constitutes in the light of Janse's (2004) findings should be further investigated by future research. It might be the case that the translation task employed in our study yields slightly different results than the phoneme detection task used by Janse (2004). It is desirable that the role of the higher phonetic articulation rate, which artificially time-compressed speech yields compared to natural fast speech is studied thoroughly. It might be the case that this higher phonetic articulation rate increases the processing load.

Also, the role of orthographic rules during spoken word recognition of reduced forms should be investigated more closely. There is evidence that literate adults who have mastered an alphabetic writing system outperform illiterate and preliterate participants in phoneme detection, phoneme deletion and phoneme insertion tasks (Morais, Cary, Alegria & Bertelson 1979, Adrian, Alegria & Morais 1995). It has also been reported that native speakers of Danish use their L1 orthographic knowledge as an additional cue during spoken word recognition of spoken Swedish (Schüppert et al. submitted). It is likely this activation of L1 orthography also has a beneficial effect on L1 spoken wird recognition, especially when word forms are highly reduced.

Acknowledgements

Kirsten Kolstrup invested much of her private time in order to lend us her voice and read aloud the stimulus material. Without her, this experiment would not have been possible. Virkelig tusen tak for hjælpen, Kirsten!

Referenser

Adrian, José A., Jesús Alegria & José Morais. 1995. Metaphonological abilities of Spanish illiterate adults. *International Journal of Psychology* 3: 329–353.

Almberg, Jørn. 2000: Kor fort snakker vi eigentleg? *Nordlyd* 28. 60-73.

Benoît, Christian, Martine Grice & Valerie Hazan. 1996. The SUS test: a method for the assessment of text-to-speech synthesis intelligibility using semantically unpredictable sentences. *Speech Communication* 18(4): 381-392.

Bleses, Dorthe, Werner.Vach, Malene Slott, Sonja Wehberg, Pia Thomsen, Thomas Madsen & Hans Basbøll. 2008. Early vocabulary development in Danish and other languages: A CDI-based comparison. *Journal of Child Language* 35(3): 619-650.

Bø, Inge. 1978. *Ungdom og naboland. En undersøkelse av skolens og fjernsynets betydning for nabospråksforståelsen.* Stavanger: Rogalandsforskning.

Boersma, Paul &David Weenink. 2009. Praat: doing phonetics by computer Version 5.1.12. [Computer program] http://www.praat.org. Accessed 16 April 2012.

Box, George E. P., William G. Hunter & Stuart J. Hunter. 1978. *Statistics for Experimenters*. New York: John Wiley and Sons.

British Broadcasting Corporation 2012. *Editorial Guidelines: Section 3. Accuracy Principles*. [Online version] http://www.bbc.co.uk/ guidelines/editorialguidelines/page/guidelines-accuracy-principles. Accessed 16 April 2012.

Campione, Estelle & Jean Véronis. 2002. A large-scale multilingual study of silent pause duration. B. Bel & I. Marlien (red.) *Proceedings of the Speech Prosody 2002 Conference*. Aix-en-Provence: Laboratoire Parole et Langage. 199-202.

Caspers, Johanneke & Vincent J. van Heuven 1993. Effects of Time Pressure on the Phonetic Realization of the Dutch Accent-Lending Pitch Rise and Fall. *Phonetica* 50(3): 161-171.

Crystal, Thomas H. & Arthus S. House. 1990. Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America* 88(1): 101-112.

Danmarks Radio. 2009. *Sprogpolitik for DR*. [Online version]. http://www.dr.dk/NR/rdonlyres/D0F84992-F0E6-4107-A2B272B6

F35B4 2D4/1199706/SprogpolitikFebruar20091.pdf. Accessed 16 April 2012.

De Jong, Nivja & Ton Wempe. 2009. Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods* 41(2): 385-390.

Delsing, Lars-Olof & Katarina Lundin Åkesson. 2005. *Håller språket ihop Norden? En forskningsrapport om ungdomars förståelse av danska, svenska och norska.* Copenhagen: Nordiska ministerrådet.

Engstrand, Olle. 1988. Articulatory correlates of stress and speaking rate in Swedish VCV utterances. *Journal of the Acoustical Society of America* 83(5): 1863-1875.

Engstrand, Olle & Diana Krull. 2001. Segment and syllable reduction: preliminary observations. *Working papers in Linguistics* 49. Lund: Lund University Department of Linguistics. 26-29.

Ernestus, Mirjam, Harald Baayen & Rob Schreuder. 2002. The recognition of reduced word forms. *Brain and Language* 81(1): 162-173.

Fairbanks, Grant & Frank Kodman. 1957. Word intelligibility as a function of time compression. *The Journal of the Acoustical Society of America* 29(5): 636-641.

Fairbanks, Grant, Newman Guttman & Murray S. Miron. 1957. Effects of time-compression upon the comprehension of connected speech. *Journal of Speech and Hearing Disorders* 22(1): 10-19.

Foulke, Emerson & Thomas G. Sticht. 1969. Review of research on the intelligibility and comprehension of accelerated speech. *Psychological Bulletin* 72(1): 50-62.

Gay, Thomas. 1978. Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America* 63(2-3): 223-230.

Gooskens, Charlotte, Vincent J. van Heuven, Renée Van Bezooijen & Jos J. A. Pacilly. 2010. Is spoken Danish less intelligible than Swedish? *Speech Communication* 52(11-12): 1022-1037.

Göransson, Eva & Stefan Lundin. 2012. *Om Sveriges Radio: Språkvård*. [Online version] http://sverigesradio.se/sida/artikel.aspx?programid=3113och artikel=2170109. Accessed 16 April 2012.

Gordon-Salant, Sandra, Peter J. Fitzgibbons & Sarah A. Friedman. 2007. Recognition of time-compressed and natural speech with selective temporal enhancements by young and elderly listeners. *Journal of Speech, Language, and Hearing Research* 50(5): 1181-1193.

Grønnum, Nina. 1998. *Fonetik og fonologi. Almen og dansk*. Copenhagen: Akademisk Forlag.

Grønnum, Nina. 2003. *Why are the Danes so hard to understand?* Henrik Galberg Jacobsen Dorthe Bleses, Thomas O. Madsen & Pia Thomsen (red.) *Take Danish – for instance: linguistic studies in honour of Hans Basbøll presented on the occasion of his 60th birthday 12 July 2003*. Odense: University Press of Southern Denmark. 119-130.

Hedelin, Per. 1997. *Norsteds svenska uttalslexikon*. Stockholm: Norstedt.

Hilton, Nanna Haug, Charlotte Gooskens & Anja Schüppert. 2011. Syllable reduction and articulation rates in Danish, Norwegian and Swedish. *Nordic Journal of Linguistics* 34. 215-237.

Hjort, Ebba & Kjeld Kristensen. 2003. *Den Danske Ordbog*. Copenhagen: Det Danske Sprog- og Litteraturselskab/Gyldendal 2003. Accessed via http://ordnet.dk/ddo on 26 march 2012.

Janse, Esther. 2004. Word perception in fast speech: artificially time-compressed vs. naturally produced fast speech. *Speech Communication* 42(2): 155-173.

Janse, Esther & Mirjam Ernestus. 2011. The roles of bottom-up and top-down information of reduced speech: Evidence from listeners with normal and impaired hearing. *Journal of Phonetics* 39(3): 330-343.

Jones, Caroline, Lynn Berry & Catherine Stevens. 2007. Synthesized Speech intelligibility and persuasion: Speech rate and non-native listeners. *Computer Speech and Language* 21(3): 641-651.

Kendall, Tyler. 2009. *Speech Rate, Pause, and Linguistic Variation: An Examination through the Sociolinguistic Archive and Analysis Project*. PhD thesis. Durham, NC: Duke University.

Kowal, Sabine, Richard Wiese & Daniel C.O'Connell. 1983. The use of time in storytelling. *Language and Speech* 26(4): 377-392.

Krause, Jean C. & Louis D. Braida. 2002. Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *Journal of the Acoustical Society of America* 112(5): 2165-2172.

Ladd, Robert. 1996. *Intonational Phonology*. [Cambridge Studies in Linguistics 79] Cambridge: Cambridge University Press.

Lindblom, Björn. 1990. Explaining phonetic variation: a sketch of the H&H theory. W. J. Hardcastle & A. Marchal (red.) *Speech production and speech modelling*. Dordrecht: Kluwer. 403-439

Maurud, Øivind. 1976. *Nabospråksforståelse i Skandinavia. En undersøkelse om gjensidig forståelse av tale- og skriftspråk i Danmark, Norge og Sverige*. Stockholm: Nordiska rådet.

Molbæk Hansen, Peter. 1990. *Udtaleordbog*. Copenhagen: Gyldendal.

Morais, José, Luz Cary, Jesús Alegria & Paul Bertelson. 1979. Does awareness of speech as a sequence of phones arise spontaneously?, *Cognition* 7: 323–331.

Moulines, Eric & Werner Verhelst. 1995. Time-domain and frequency-domain techniques for prosodic modification of speech. W.B. Kleijn & K.K. Paliwal (red.) *Speech Coding and Synthesis.* Amsterdam: Elsevier Science. 519-555.

Os, Els A Den. 1988. *Rhythm and tempo of Dutch and Italian: A contrastive study*. PhD thesis.Utrecht University.

Peterson, Gordon E. & Ilse Lehiste. 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32(6): 693-703.

Port, Robert F. 1981. Linguistic timing factors in combination. *Journal of the Acoustical Society of America* 69(1): 262-274.

Schüppert, Anja & Charlotte Gooskens. 2011. Investigating the role of language attitudes for perception abilities using reaction time. *Dialectologia* II: 119-140.

Schüppert, Anja. 2011. *Origin of Asymmetry. Mutual intelligibility of spoken Danish and Swedish*. Groningen: Grodil.

Schüppert, Anja, Johannes C. Ziegler, Kristina Borgström, Kenneth Holmqvist, Holger Juul & Charlotte Gooskens. Submitted: On-line activation of L1 orthography enhances L2 spoken word recognition. Evidence from ERP.

Språkrådet, 2007. *Språkregler for NRK*. [Online version] http://www.sprakradet.no/nb-no/Politikk-Fakta/Spraakpolitikk/NRK. Accessed 16 April 2012.

Studebaker, Gerald A. 1985. A "Rationalized" Arcsine Transform. *Journal of Speech and Hearing Research* 28(3), 455-462.

Tsao, Ying-Chiao & Gary Weismer. 1997. Interspeaker variation in habitual speaking rate: Evidence for a neuromuscular component. *Journal of Speech, Language and Hearing Research* 40(4): 858-866.

Vaughan, Nancy E. & Tomasz Letowski. 1997. Effects of age, speech rate, and type of test on temporal auditory processing. *Journal of Speech, Language, and Hearing Research* 40(5): 1192-1200.

Verhoeven, Jo, Guy De Pauw & Hanne Kloots. 2004. Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and Speech* 47(3): 297-308.

Zemlin, Willard R., Raymond G. Daniloff & Thomas H. Shriner. 1968. The difficulty of listening to time compressed speech. *Journal of Speech and Hearing Research* 11(4): 875-881.

Appendix
Stimulus material

En regel synger efter et økonomisk udtryk.
En forskning indtræffer bag en europæisk karakter.
Et program udgår på en effektiv procent.
En person kommer i en religiøs nation.
En sommer flytter under en sikker handling.
En nyhed rejser over et demokratisk arbejde.
En turist regner efter en intellektuel effekt.
En fordel forsvinder bag en gammel forklaring.
En militær kunstner vinder en myndighed.
En effektiv amerikaner afgør et møde.
En mulig baggrund mærker en indsats.
En personlig samling hænger en generation.
En aktuel ekspert består en meter.
En historisk professor præsenterer et studium.
Et væsentligt behov kalder en roman.
Et centralt samarbejde udgør en mulighed.
En billig situation modsvarer en oplevelse.
En amerikansk regering studerer et ansvar.
En praktisk direktør forklarer en time.
En nordisk generation ordner en anledning.
Et socialt system skyder en gade.
Et ydre resultat køber en oplevelse.
Hvor forstår en litteratur en svensk indstilling?
Hvorfor glemmer et hoved et litterært parti?
Hvor påpeger en general en engelsk udvikling?
Hvorfor måler en radio et færdigt bidrag?
Hvor lægger et besøg en normal tradition?
Hvorfor henter en årsag en vigtig revolution?
Hvor vækker en kvinde en alvorlig ungdom?
Hvorfor foreslår en historie en rigtig glæde?
Hvor beslutter en institution en speciel produktion?
Hvorfor løser en præsident en ensom krone?
Hvor følger en undervisning en politisk befolkning?
Hvorfor behøver et bibliotek et nuværende museum?
Hvor kræver et nummer et teknisk forslag?

Hvorfor synger en forestilling en berømt diskussion?
En behandling skaffer en rejse som læser.
En kontakt vælger en forudsætning som vover.
En kritik træffer en frihed som eksisterer.
En linje behandler en organisation som sidder.
En patient rækker en arkitekt som accepterer.
En skole maler et område som venter.
En magt skaber en virksomhed som ligger.
En mening konstaterer en retning som forsøger.
En tanke anvender et middel som tænker.
Et forsøg udvikler en musik som hænder.
En mandag viser en vare som betaler.
En artikel flytter en side som spiller.
Et indtryk bygger en debat som fortsætter.
Et universitet udnytter en ordning som arbejder.