Why is Danish so difficult to understand for fellow Scandinavians?☆

Anja Schüppert*, Nanna H. Hilton, Charlotte Gooskens

University of Groningen, Center for Language and Cognition Groningen, Postbus 716, 9700AS Groningen, Netherlands

Abstract

It has consistently been shown that among the three mainland Scandinavian languages, Danish is most difficult to understand for fellow Scandinavians. Recent research suggests that Danish is spoken significantly faster than Norwegian and Swedish. This finding might partly explain the asymmetric intelligibility among Scandinavian languages. However, since fast speech goes hand in hand with a high amount of speech reduction, the question arises whether the high speech rate as such impairs intelligibility, or the high amount of reduction. In this paper we tear apart these two factors by auditorily presenting 168 Norwegian- and Swedish-speaking participants with 50 monotonised nonsense sentences in four conditions (quick and unclear, slow and clear, quick and clear, slow and unclear) in a translation task. Our results suggest that speech rate has a larger impact on the intelligibility of monotonised speech than naturally occurring reduction.
© 2016 Elsevier B.V. All rights reserved.

1. Introduction

In Scandinavia, it has long been the tradition to communicate by relying on mutual intelligibility, i.e. by using one's own native Scandinavian language with speakers of other Scandinavian languages. That means that a speaker of Danish might speak Danish to Norwegians who then reply in Norwegian. Several studies have shown, however, that the three mainland Scandinavian languages are not mutually intelligible to the same extent. Norwegian is the language which is generally the most intelligible to Scandinavians, and Norwegians do better in comprehending their neighbouring languages as well (cf. Delsing and Lundin-Åkesson, 2005). One of the central explanations has been the fact that the Norwegian lexicon is very similar to the Danish lexicon, a result of Norway having been part of the Danish empire between 1380 and 1814, while Norwegian pronunciation is similar to Swedish pronunciation both on a segmental and on a prosodic level (Gooskens, 2007; Haugen, 1966).

Lower intelligibility scores are generally found for Danish-Swedish communication, and particularly so for Swedes listening to Danish. Several factors have been suggested to cause this asymmetry. Research by Delsing and Lundin Åkesson (2005), Maurud (1976), Schüppert and Gooskens (2011) and Schüppert et al. (2015) suggest that Danes hold a more positive attitude towards Swedish than vice versa. A widespread belief is that Danes therefore might make a greater effort understanding Swedish, which results in higher intelligibility scores. However, Gooskens (2006) points out that the causal relationship between a positive attitude and higher intelligibility scores is hard to establish. It might also be the case that participants who have fewer difficulties understanding the neighbouring language have a more positive attitude towards this language.

Other suggested explanations for the variation in intelligibility scores within and between Scandinavian countries have been that of geographic proximity and contact frequency. Gooskens and Hilton (2013) find no differences in intelligibility of Danish between Norwegian teenagers living 2000 km from Denmark and those who live close enough (300 km) to frequently visit the country. Nor does Gooskens (2006) report significant correlation coefficients for the amount of personal contact or visits, or contact with the language via television or newspapers, with intelligibility. However, this missing

correlation might be due to the fact that the contact index was generally very low and thereby little variance was observed.

Other factors that previous research has considered in order to establish to which extent they influence mutual intelligibility in Scandinavia are linguistic: Kürschner et al. (2008) indicate that lexico-phonological factors such as word length and neighbourhood density might play a role for successful intelligibility of Danish by Swedes. They also show a significant correlation between phonetic (Levenshtein) distances and intelligibility, earlier established by Gooskens (2007). Hilton et al. (2013) indicate that word order differences between the languages can influence intelligibility levels negatively, but conclude that phonological factors are more crucial to successful comprehension between speakers of Scandinavian languages.

One such phonological factor could be articulation rate (i.e. the number of linguistic entities per time unit such as phonemes, syllables, or words, excluding pauses; cf. Jacewicz et al., 2009; Tsao et al., 2006). Hilton et al. (2011) report that Danish newsreaders speak significantly faster than their colleagues in Norway and Sweden do if syllables per second are measured. Schüppert et al. (2012) confirmed this finding for Danish and Swedish with a different measure, namely words per second. In both studies, the same material from the non-commercial public service radio stations Danmarks Radio (DR), Sveriges Radio (SR) and, for Hilton et al. (2011), Norsk Rikskringkasting (NRK) was used.[1]

The findings that Danish is spoken more quickly than Norwegian and Swedish when it is read by professional news readers to a broad public suggests that an increased tempo impairs native speakers' intelligibility of Danish to a lesser extent than native speakers' intelligibility of Norwegian and Swedish. Janse (2004) and Vaughan and Letowski (1997) showed that the process of time-compressing a given speech sample generally impairs intelligibility more than the process of time-extending a given speech sample. Therefore, it seems reasonable to assume that this difference in articulation rate is at least part of the reason why spoken Danish is so difficult to understand for Norwegians and Swedes.

If we have a closer look at what makes fast speech less intelligible, we can identify at least two different factors which are both inter-correlated with a high articulation rate (Bradlow et al., 2003; Ferguson et al., 2010; Ferguson and Quené, 2014; Lam et al., 2012; Picheny et al., 1986; Rosen et al., 2011; Smiljanic and Bradlow, 2005; Smiljanic and Bradlow, 2008). The first factor concerns the speakers: speaking quickly increases the demands on the articulatory apparatus. Hence, the faster the speech, the more likely the speaker is to reduce

specific sound entities such as phonemes or syllables. The second factor is located in the listener: namely the shorter time frame for the decoding of linguistic units, and hence the higher demands on the decoding process.

Firstly, when listening to fast speech, we need to decompose and process the stream of speech sounds more quickly. Several studies have investigated the effect of presentation rate (usually defined as the number of items presented visually or auditorily per minute) in recalling tasks, where participants are confronted with a sequence of words and are asked to recall this sequence as accurately as possible. While Lilienthal et al. (2014), Mackworth (1962) and Tan and Ward (2008) reported that decreasing the presentation rate, and thereby giving participants additional time to rehearse between the presentations of items, improves memory performance, Conrad and Hille (1958) and Posner (1964) showed that memory performance decreases with an increase in presentation rate. Baddeley et al. (1975) report a systematic relationship between memory span (the number of words a person can recall immediately after hearing them) and the duration of the words, such that memory span is equivalent to the number of words which can be read out in approximately two seconds when read at a normal rate. Their data suggest that the articulatory system has a temporally limited capacity. Part of the explanation why a high articulation rate is linked to poor intelligibility might thus be the fact that speech processing partly relies on the working memory capacities and that the demand on the working memory is higher when the information is dense.

Secondly, the role of reduction on intelligibility of speech has been investigated. By reduced speech we mean abbreviated durations of long sounds, use of a smaller vowel space, as well as elision of entire segments (Gahl et al., 2012). Reduction has shown to cause intelligibility difficulties in subjects listening to their native language (e.g. Bond and Moore, 1994; Hazan and Markham, 2004). In his H&H ('hyper'- and 'hypo'-articulation) theory, Lindblom (1990) argues that speakers of any language are constantly balancing between 'hyperspeech', i.e. clear articulation to maximise intelligibility in the listener, and 'hypospeech', i.e. unclear speech to minimise the articulatory effort for the speaker. Generally, these two opposing efforts lead to speech which contains a certain amount of reduction phenomena but is still fairly intelligible to the listener. The amount of reduction in speech depends on factors such as age of the speaker (Guy 1992), gender (Neu, 1980; Wolfram, 1969; Zue and Laferriere, 1979), speaking situation or style (Coupland, 1980; Ernestus et al., 2015; Labov, 1966; Picheny et al., 1986), and also on the rate at which the speech is produced (Ernestus et al., 2015; Fosler-Lussier and Morgan, 1999; Fourakis, 1991; Guy, 1980; Jurafsky et al., 2001; Labov and Cohen, 1967; Labov et al., 1968; Raymond et al., 2006; Wolfram 1969). Due to articulatory restrictions, fast speech generally is less accurately articulated than slow speech.

The aim of the present paper if twofold: (1) We investigate whether the reported difference in articulation rate can partly account for the fact that spoken Danish is so difficult

---

[1] Since newscasters are often trained in specific professional styles, we do not know whether these findings can be extended conversational speech or other spontaneous speech tasks. However, we are not aware of any cross-linguistic investigation of spontaneous Danish, Norwegian and Swedish. One clear advantage of news readings is that the setting is roughly comparable across all three languages, which can be assumed to be very difficult to achieve for spontaneous speech even if speech samples are recorded specifically for this purpose.

to understand for Norwegians and Swedes, and (2) we tear apart the two intertwined factors *duration* and *reduction* and shed light on whether one of these factors impairs the intelligibility of Danish for fellow Scandinavians more, and if this is the case, which one.

2. Method

The stimulus material consisted of 50 semantically unpredictable sentences (henceforth SUS) that were read aloud by a native speaker of Danish in two different, yet natural, conditions: (i) at a slow speaking rate with a deliberately accurate pronunciation, and (ii) at a high speaking rate with less accurate pronunciation.

These two conditions were manipulated so as to form two additional ones, namely (iii) slowly paced and inaccurately articulated speech and (iv) quickly paced and accurately articulated speech. The two additional conditions and the different steps of deriving them will be explained in detail in the following section.

2.1. Material

2.1.1. Compilation of material

The SUS were generated by the method developed by Benoît et al. (1996). These sentences were originally developed to assess the intelligibility of text-to-speech synthesis, but have also been used for testing intelligibility of natural language (cf. Gooskens et al., 2010). The SUS are syntactically correct sentences but consist of phrases with concepts that are not likely to be semantically related to each other. 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 consist of six or seven words.

Instead of using words as the linguistic entity to define articulation rate, we used syllables. When defining syllables, we followed the traditional approach that assumes that the centre of a syllable is a vocoid (for Danish cf. Basbøll, 2005:180f). More specifically, we established two measures: The number of canonical syllables and the number of phonetically realised syllables.

For our material, we established the number of canonical syllables per sentence using the Danish dictionary *DanskOrdbogen*, which indicates the possible ways to split a particular word at the end of a line for each entry. By comparing our division to the principles of phonological syllabification laid out in Basbøll (2005:252–258) we concluded that the indications given by *DanskOrdbogen* represent phonological syllables, and use the term 'canonical syllables' for these. For example, the word *kirke* (Engl. 'church') consists of two syllables split at one point, namely *kir|ke* both following *DanskOrdbogen* as well as Basbøll's principles (2005:257). In addition to canonical syllables we define phonetically realised syllables. These are syllables which are measureable as sonor-

ity peaks in the acoustic signal. These measurements will be explained in greater detail in Section 2.1.5.1. Figs. 1 and 2 illustrate that when speaking slowly, the word *kirke* is indeed pronounced with two sonority peaks by our speaker, while only one sonority peak was detected in fast speech. In other words, in contrast to canonical syllables, phonetically realised syllables show great inter- and intra-individual variation, constrained by factors such as the familiarity with the content (Goldman-Eisler, 1968), the mental health of the speaker (Cannizzaro et al., 2004; Darby and Hollien, 1977; Teasdale et al., 1980), the speakers' regional origin (Jacewicz et al., 2010; Robb et al., 2004), and gender (Jacewicz et al., 2009; Van Borsel and De Maesschalck, 2008), to name just a few.

The SUS varied in length between 10 and 18 canonical syllables with a mean of 13.1 syllables. 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 politik hjælper på en sikker kirke.*
English 'A policy helps against a secure church.'

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 were special restrictions, the most important of which were the following:

○ verbs: no auxiliaries and reflexives, only present tense (including the imperative)
○ 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

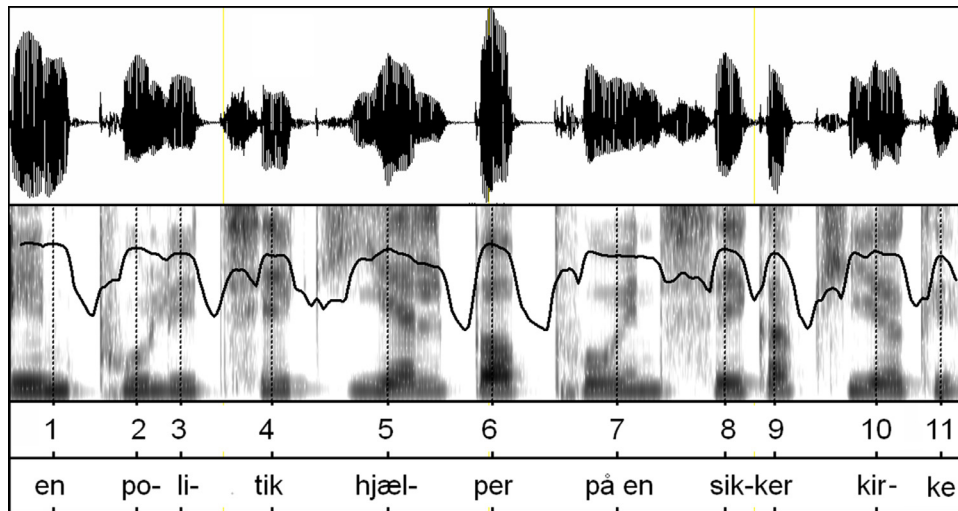All words were selected randomly from the thousand most frequent Danish words in their lexical category using the

Fig. 1. Oscillogram (upper panel), spectrogram (central panel) with intensity curve (black line), and syllable tier (lower panel) for the SUS En politik hjælper på en sikker kirke in condition (i), slowly and clearly produced, with a total duration of 3163 ms. This version was linearly time-compressed to 2080 ms to form condition (iii).
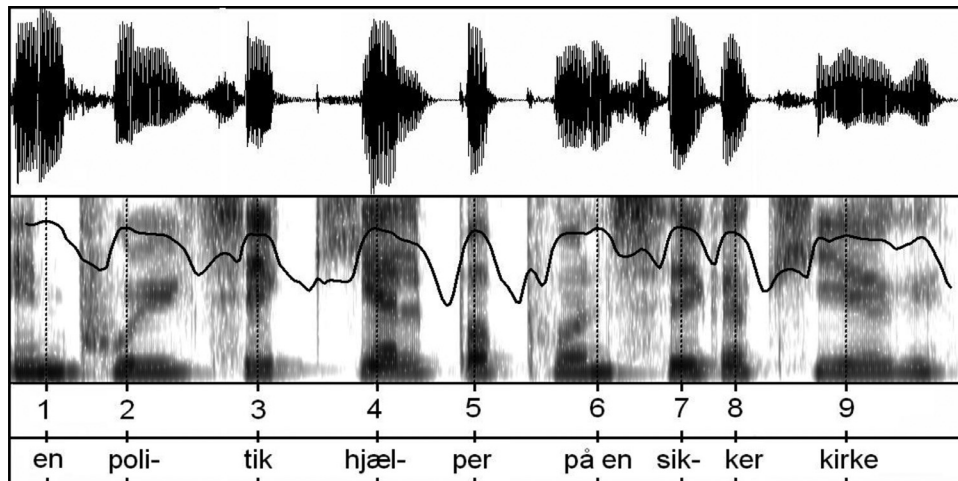

Fig. 2. Oscillogram (upper panel), spectrogram (central panel) with intensity curve (black line), and syllable tier (lower panel) for the SUS En politik hjælper på en sikker kirke in condition (ii), quickly and unclearly produced, with a total duration of 2080 ms. This version was linearly time-extended to 3163 ms to form condition (iv).

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). Crucially, only cognate words were included, i.e. all Danish content words and all Danish function words in the material shared etymology with their Swedish and Norwegian counterparts.

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* (conjunction 'and') and *som* (relative pronoun) were allowed to occur more often.

2.1.2. Speech material

Once the material was compiled in written form, the stimulus sentences were read aloud by a female native speaker

of Danish and recorded in a sound-attenuated room at the University of Groningen at a sampling rate of 44.1 kHz. Two conditions were recorded, namely (i) slowly and clearly and (ii) quickly and less accurately. The speaker was instructed to produce the quick and the slow sentences without sentence-internal prosodic boundaries, i.e. without any pauses.

2.1.3. Manipulation

The material was manipulated in two ways. First, each slowly produced sentence was time-compressed linearly by reducing the total duration to the duration of the same sentence produced quickly. In a similar manner, each quickly produced sentence was 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

Table 1
Mean duration, mean number of canonical syllables and phonetically realised syllables, mean articulation rate and mean number of spectral changes per second for sentences in conditions (i) and (ii).

Condition	Duration (s)	No. of canonical syll.	No. of sonority peaks	Articulation rate	No. of spectral changes/s	F1 (Hz)			F2 (Hz)		
						/a/	/i/	/u/	/a/	/i/	/u/
(i) Quick and unclear	1.8	13.1	8.7	4.8	1532	929	272	284	1523	2431	1010
(ii) Slow and clear	3.0	13.1	13.0	4.3	1389	917	262	262	1521	2509	948

was compressed from 3.0 to 1.8 s to create quick and yet accurately articulated sentences that form condition (iii). Likewise, sentence duration was extended from 1.8 to 3.0 s to create slow and inaccurately articulated sentences that form condition (iv). The mean factors for duration manipulation were 1.67 and 0.6, respectively. As mentioned above, time compression leads to larger intelligibility issues than time expansion. The compressed speech samples (condition iii) are therefore likely to be more difficult for the listener to decode than the original (condition i). As one of the main questions we address in this paper is whether the effect of reduction is more or less powerful than timing for intelligibility of a closely related language, this approach is still chosen as the most functional, despite the confounding effect of compression on intelligibility.

It is unknown for Danish how a speaker adjusts 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 and Van Heuven, 1993; Ladd, 1996). To ensure that the manipulated sentences would not sound less natural than the unmanipulated sentences, sentences in all four conditions were monotonised with a fixed F0 of 213 Hz, which was the mean F0 employed in the original recordings. Arguably, the monotonisation makes all stimulus sentences sound unnatural and restricts their general intelligibility. McCloy et al. (2015:382) find a larger overall pitch range in speakers who are more intelligible in dialect comprehension tasks. Speech at higher tempos generally employs a smaller pitch range than normal or slow speech (e.g. Caspers, 1994). It can therefore be argued that the intelligibility of quick speech (conditions ii and iv) in our material is impaired slightly less by monotonisation than the intelligibility of slow speech (conditions i and iii). This needs to be kept in mind when interpreting the results.

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.

2.1.4. Acoustic differences across conditions

We measured mean duration, mean number of canonical syllables, mean number of phonetically realised syllables, mean number of sonority peaks, mean number of spectral changes per second, and mean frequency of the first two formants in the stressed vowels /a, i, u/ for sentences in conditions (i) and (ii). Results are given in Table 1.

The number of sonority peaks was measured using the automatic script developed by De Jong and Wempe (2009) as explained in Section 2.1.5.1. In the slow mode, the sentences were produced with 10 to 18 sonority peaks (mean = 13.0), while the same sentences read in the quick mode had 5 to 12 sonority peaks (mean = 8.7). Arguably, a sonority peak generally indicates the presence of a phonetic syllable, so we interpret this difference as reflecting a difference in the mean number of phonetic syllables.

When comparing the number of phonetic syllables to the number of canonical syllables it can be concluded that the slow recordings were indeed produced very accurately, as hardly any canonical syllables were deleted in actual pronunciation. This seems to suggest that syllable deletions in Danish are not phonological, i.e. that in very careful speech, less than 1% of the canonical syllables are deleted. 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 the newsreaders' corpus reported in Schüppert et al. (2012) (SDP 29%, see Section 3.3).

One of the aims of this paper is to compare intelligibility of four different types of speech. To be able to interpret the intelligibility scores, it is necessary to investigate and describe the material thoroughly. This will be done in this section by means of two measurements. Firstly, we identified the number of sonority peaks in every sentence (cf. Section 2.1.4.1). Secondly, we quantify the number of spectral changes per second in the acoustic signal (cf. Section 2.1.4.3). We interpret both measures as acoustic indications of the accuracy of the articulation. However, longer utterances generally contain more sonority peaks and more spectral change. Therefore, both measures are normalised for the duration of the utterance by dividing through utterance duration. Details of the algorithms used for detecting sonority peaks and measuring the amount of spectral change are given in this section.

2.1.4.1. Number of sonority peaks measured per second. In order to detect sonority peaks in the material, we employed an algorithm developed by De Jong and Wempe (2009). We defined a sonority peak as having an intensity of at least 2 dB higher than the surrounding signal in the voiced part of the speech signal. An example of the output of the automatic procedure is shown in the upper three panels in Figs. 1 and 2. The top panel of the figures shows the oscillogram of the SUS *En politik hjælper på en sikker kirke* ('A policy helps on a secure church'). The mid-upper panel shows the spectrogram with the measured intensity (black line). The mid-lower panel
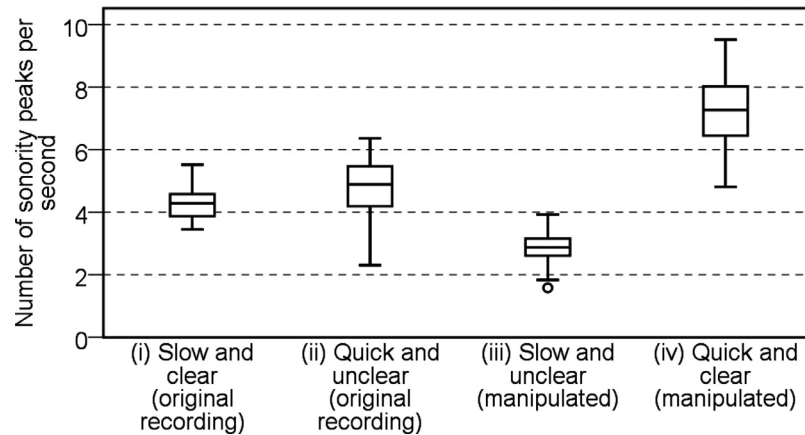
Fig. 3. The phonetic articulation rate (number of sonority peaks per second) for all four conditions.

shows a grid with detected sonority peaks according to our definition above. The bottom panel contains a manually added orthographic transcription for each detected peak. This panel is not part of the automatic output and has only been added here for illustration purposes. If we compare Figs. 1 and 2, it can be seen that the SUS *En politik hjælper på en sikker kirke* contains 11 sonority peaks if produced slowly and clearly, but only eight if produced quickly and unclearly.

In order to normalise the number of sonority peaks per sentence for differences in sentence length, the number of sonority peaks was subsequently divided by the sentence duration, which results in a measure of phonetic articulation rate. Fig. 3 shows the phonetic articulation rate per condition in a box plot. It can be seen that the sentences in condition (i), slowly and clearly produced, have a lower phonetic articulation rate ($m = 4.26$ syll/s) than the sentences in condition (ii), quickly and unclearly produced ($m = 4.95$ syll/s). The difference is significant ($t(98) = 5.34$, $p < 0.001$), which indicates that the sentences in condition (i), although pronounced carefully, contain fewer sonority peaks per second than the sentences in condition (ii), since they are produced more slowly.

Logically, the two conditions containing manipulated sentences show artificially high (condition iii) and artificially low (condition iv) phonetic articulation rates. It can also clearly be seen that the articulation rates in the original slow recordings are less variable than in the original quick recordings. Since we manipulated the duration of every original sentence to the corresponding sentence in the opposite original recording, this difference in variances is also found in condition (i) compared to in condition (iv).

To our knowledge, the algorithm used to quantify sonority peaks in the speech signal has not been verified for Danish specifically and therefore, our measurements have to be interpreted carefully. However, since our measurements are not comparing sonority peaks cross-linguistically, we assume that they are reliable.

2.1.4.2. Segmental reduction in Danish. In contrast to Norwegian and Swedish, Danish is characterised by radical reduction processes which are manifested in consonant gra-

Table 2
The realisation of the phonemes /p, t, k, b, d, g, v, j, r/ in in word-medial and word-final position Danish, East Norwegian, and Swedish.

Phoneme	Danish realisation	Norwegian realisation	Swedish realisation
p	b̥	p	p
t	d̥	t	t
k	g̊	k	k
b	b̥	b	b
d	ð[a]	d	d
g	ɪ / ʊ	g	g
v	ʊ	ʋ	v
j	ɪ	j	j
r	ɐ̯	ɾ	ɾ/r

[a] Realised by a dental approximant in medial and final position.

dation phenomena, and schwa-assimilation. For an overview of the underlying processes and a cross-linguistic comparison between Danish and Swedish, see Basbøll (2005:258ff.). Table 2 gives an example of phonetic realisation differences in the consonantal phonemes /p, t, k, b, d, g, v, j, r/ in word-medial and word-final post-vocalic position in Danish, Swedish and East Norwegian, as based on Basbøll (2005), Engstrand (2004) Kristoffersen (2000).

It has been shown that differences in articulation rate or speech rate are linked to vowel quality. More specifically, vowels tend to be more centralised in reduced speech, than in normal or clear speech (Fourakis, 1991; Hirsch et al., 2008; Lindblom, 1990; Wright, 1997). To quantify differences in vowel quality in the two original conditions, we measured the frequency of the two first formants and conducted pairwise comparisons. Results are shown in Table 3 and visualised in Fig. 4. Our measurements confirm that the vowels in the quickly produced sentences are produced slightly more centralised than the vowels in the slowly produced sentences. However, the frequency of only one of the six formant comparisons (three vowels * two formants) differs significantly, as can be seen in Table 3. This suggests that reduction in these particular speech samples is not manifested as differences in vowel space.

2.1.4.3. Number of spectral changes per second. To quantify the amount of spectral changes per second in any given sound

Table 3
Pairwise comparisons of F1 and F2 per vowel across conditions (i) and (ii). The mean differences across conditions represent the frequencies of the slowly produced condition minus the frequencies of the quickly produced condition (larger values indicate stronger centralisation).

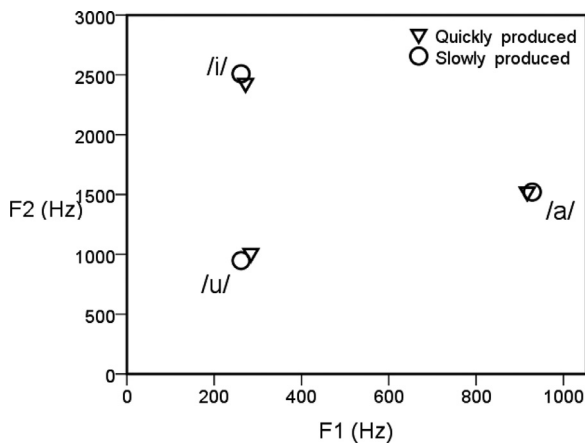| Vowel	Formant	Mean difference (Hz)	Std. deviation	Std. error mean	t	df	Sig. (2-tailed)
/a/	F1	11.50	52.89	12.47	0.92	17	0.37
/a/	F2	−1.61	68.64	16.18	−0.10	17	0.92
/i/	F1	−10.38	24.86	6.89	−1.51	12	0.16
/i/	F2	77.92	122.32	33.93	2.30	12	0.04
/u/	F1	−22.33	24.06	9.82	−2.27	5	0.07
/u/	F2	−61.33	88.60	36.17	−1.70	5	0.15

Fig. 4. Mean F1 and F2 values plotted against each other for the two originally recorded conditions (i) and (ii).

file automatically, we analyse the acoustic information of the speech signal. The algorithm we used has been written for the phonetic software Praat (Boersma and Wenink, 2013). In step 1, the excitation levels for each frequency band are differentiated in time, and then, for each moment (frame), these differences are integrated (summed) over all frequency bands. This yields both the spectral changes and the overall excitation levels differences of the signal as a function of time. In step 2, the order of these two operations is reversed. Now, first the excitation levels are summed over all frequencies - which yields the total excitation level (think of it in terms of the loudness of the signal) and then this result is differentiated in time. This renders excitation level differences as a function of time, over the complete spectral band, ignoring any spectral issues. In step 3, the difference is taken from the results of the two former operations. By doing this, we aim to get a measurement which represents the amount of spectral changes in time, not being affected by overall excitation level (intensity or loudness) changes. In other words, a measurement which indicates timbre differences as a function of time, independently of intensity/loudness differences. Finally, for each sentence the average value of this measurement (Phones*Bark/s) is calculated.

Mean and standard deviation of Phones*Bark/s were obtained for all of the 200 stimuli sentences using a Praat

script to quantify speech reduction (Schüppert et al., in prep.). Fig. 5 shows a box plot of the number of spectral changes per second per condition. It can be seen that the sentences in condition (i), slowly and clearly produced ($m = 1162$ Phones*Bark/s) contain fewer spectral changes per second than the sentences in condition (ii), quickly and unclearly produced ($m = 1426$ Phones*Bark/s). The difference is significant ($t(98) = 4.8$, $p < 0.001$) which indicates that the sentences in condition (i), although pronounced carefully, contain less acoustic information per second than the sentences in condition (ii).

A repeated-measures ANOVA with the factor duration (two levels: long and short) and articulation accuracy (two levels: clear and unclear) revealed that there is a main effect of both factors on the mean number of spectral changes per second. More specifically, the speech samples containing fast speech contain significantly more spectral changes per second than those containing slow speech. Likewise, the speech samples containing clear speech show significantly more spectral changes per second than those containing unclear speech.

As expected, the number of spectral changes in the two conditions derived through manipulation of the natural recordings deviate strongly from the mean number of spectral changes per second in the two naturally recorded conditions, while sentences in the two naturally produced conditions show less extreme numbers of spectral changes per second.

In the original sentences, the mean number of spectral changes per second correlates significantly with the phonetic articulation rate in the same signal ($r = 0.28$, $p = 0.003$). Note that this correlation coefficient is only based on the original recordings, because including the values for the manipulated sentences, where the same factor has been applied on the two measures, would distort the results and render artificially high correlation coefficients.

2.2. Design and task

The participants for the intelligibility experiment were asked to fill in a short questionnaire providing information about their background, such as age, sex, place of residence and which language(s) they spoke with their parents at home.

After questionnaire completion the experiment started. Every sentence was presented auditorily twice to every participant. The participants' task was to translate the Danish sentences as accurately as they could into their native language, i.e. Norwegian or Swedish for the experimental groups. The task of the Danish-speaking control group was to write down the sentences in Danish (no translation involved).

Prior to the actual experimental session, the listeners were presented with five training sentences to get used to the task. These five sentences were not analysed. After the training session, the experiment started. Each participant listened to 50 sentences in different conditions. Sentences were blocked by condition and rotated over listener groups according to a complete Latin-square design (cf. Box et al., 1978). Four different experiment versions were created to present all 200 sentences to the participants to ensure that the same sentence
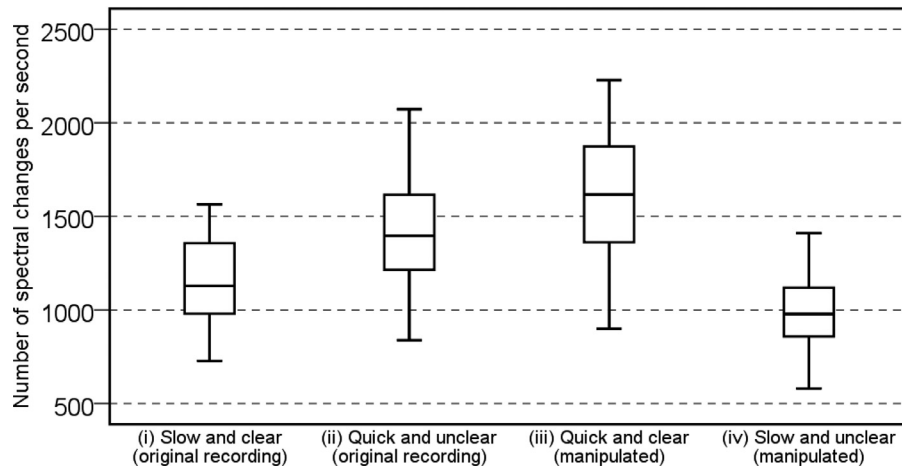
Fig. 5. Box plot of mean spectral changes per second across all four conditions.
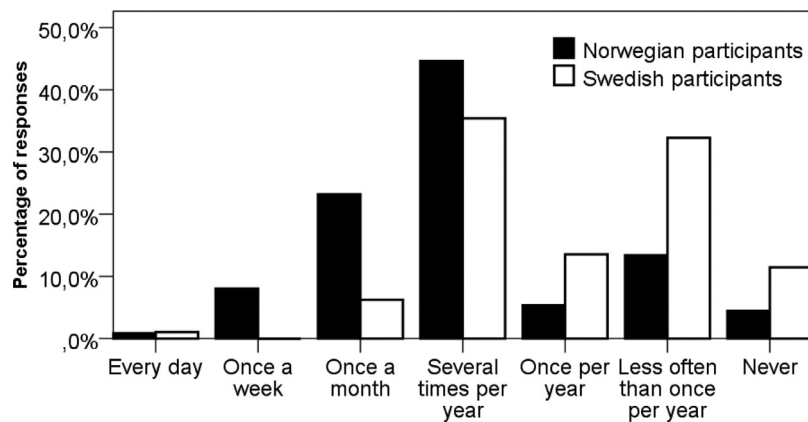

Fig. 6. Norwegian and Swedish participants' answers on the question 'How often do you hear Danish?' (1 – every day, 2 – every week, 3 – every month, 4 – several times per year, 5 – once per year, 6 – less often than once per year, 7 – never).

was only presented once to each listener. The order of the sentences was randomised across conditions, but kept constant across participants within one group.

2.3. Participants

Participants were 103 Norwegian and 66 Swedish adolescents aged 15 to 22 years, as well as a comparable Danish control group consisting of 42 adolescents. The Danish control group came from the Odense area, the Norwegian group from the larger Oslo area and the Swedish group from the Stockholm area. At the time of testing, all participants attended secondary school at a level that would admit entrance to university after completion. For the Norwegian and Swedish participants, apart from eliciting their age and education, some information concerning their attitude towards and contact with Danish was elicited. More specifically, the participants were asked to indicate on a seven-point menu how often they hear Danish (1 – every day, 2 – every week, 3 – every month, 4 – several times per year, 5 – once per year, 6 – less often than once per year, 7 – never), and on three five-point Semantic Differential Scales (Osgood et al., 1957) how much they liked Danish, whether they find Danish difficult to

Table 4
Information provided by the participants: Mean age, percentage of women, attitude towards Danish, and difficulties understanding Danish split up per language group.

L1	Danish (control group)	Norwegian	Swedish
N	42	103	66
Mean age	18.7	17.8	15.7
Percentage women	61.9	58.3	71.2
Attitude towards sound of Danish	–	3.9	4.0
Attitude towards learning Danish	–	4.0	3.8
Difficulties understanding Danish	–	3.8	4.0

understand, and whether they would like to learn Danish (all scales: 1 – very much, 5 – not at all). Table 4 summarises the information on the participants and Fig. 6 shows a bar graph of the responses to the question 'How often do you hear Danish?'

2.4. Analysis

Every sentence contained exactly four content words (nouns, verbs and adjectives). Only the translations (for

Table 5
Mean intelligibility in the four conditions: (i) slowly and clearly, (ii) quickly and unclearly, (iii) quickly and clearly, and (iv) slowly and unclearly for the participants from all three language groups.

L1	Condition				
	(i) Slowly and clearly	(ii) Quickly and unclearly	(iii) Quickly and clearly	(iv) Slowly and unclearly	*Mean*
Danish (control)	94.6	85.2	83.8	88.8	*88.1*
Norwegian	62.0	39.3	38.8	45.1	*46.3*
Swedish	44.2	25.4	21.7	30.2	*30.4*
Mean	*66.9*	*49.9*	*48.1*	*54.7*	*54.9*

Danish participants: written answers) of the content words were analysed. For every correct translation one point was given, so maximally, every participant could score four points per sentence and 200 points in total. A translation was counted as correct if there was not more than one spelling mistake which did not result in a new existing word. For example, a translation of the target word *kvinde* ('woman') with the Swedish word *kvinna* was counted as correct, as were the slightly misspelt translations ⟨kvina⟩ or typos such as ⟨kv;inna⟩. A translation with the word *vinna* ('to win') was counted as incorrect, however, as the missing letter results in the forming of a new word.

3. Results and discussion

3.1. Main effects of L1, duration and reduction on intelligibility

Table 5 shows the mean intelligibility scores per language group per condition. It can clearly be seen that the scores for the Danish control group show a ceiling effect. Since the task was not designed for native speakers in the first place, the Danish speakers' overall results are very high, which weakens the analysis of the effect of duration and reduction. A mixed-effects ANOVA was used on the data with L1 and condition as factors. As the design of the study is a Latin-square crossed design experiment all participants are presented with all sentences and so listener and sentence are not included as random factors in the model.

The model reveals that the three groups of L1 speakers' (Danish, Norwegian, Swedish) scores are significantly different ($F(2,147) = 548.83$, $p < 0.001$), with the native-speaker group of Danish participants correctly writing down 88.1% of the content words, the Norwegian-speaking listeners correctly translating 46.3% and the Swedish-speaking listeners scoring lowest at 30.4%. A post-hoc test shows that all three means are significantly different from each other. This confirms earlier research by Bø (1978), Delsing and Lundin Åkesson (2005), Gooskens (2006) and Maurud (1976) who reported that Norwegian-speaking participants have fewer difficulties decoding spoken Danish than Swedish-speaking participants. It should be noted that the Swedish participants have a lower mean age than the Norwegian informants, and that this
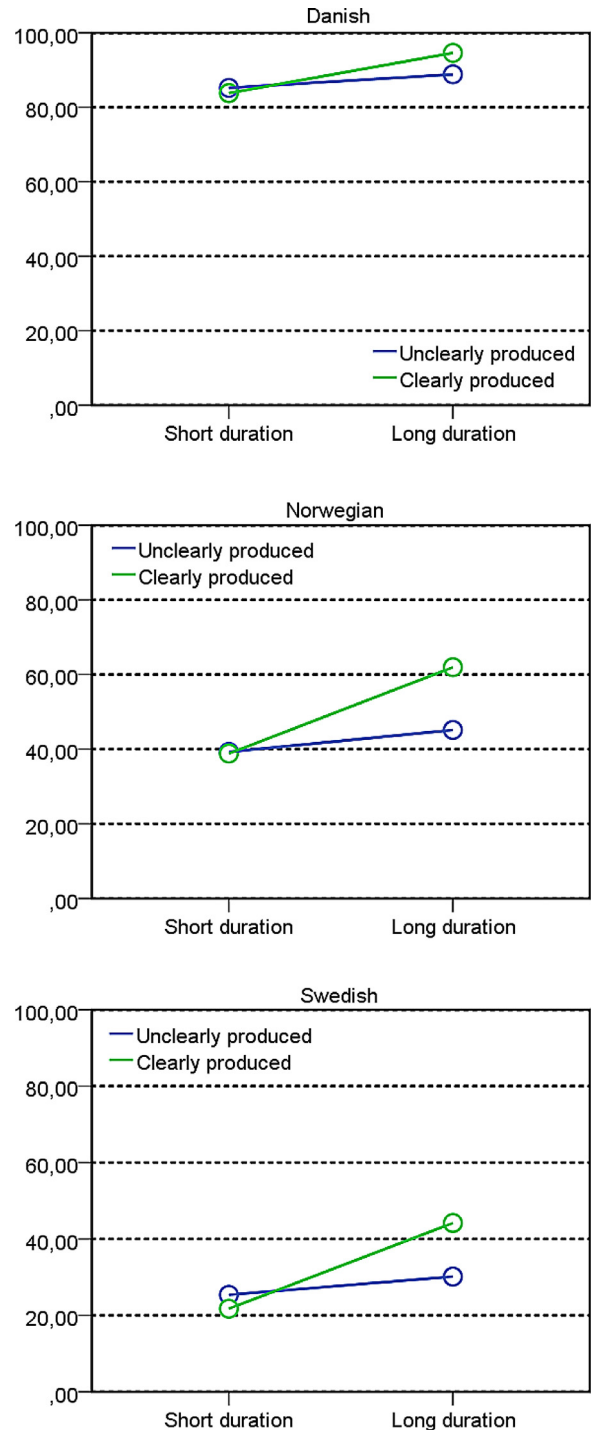

Fig. 7. Intelligibility results split up per L1 and per condition.

factor may increase the difference in intelligibility between these groups somewhat. Table 5 gives mean intelligibility scores for all three groups of listeners split up per condition.[2] Fig. 7 visualises these results.

---

[2] Considering the constraints of this study due to the monotonisation of the material (see Section 2.1.4), it is possible that our results are slightly skewed, as monotonising clear speech is likely to impair the intelligibility to a larger extent than monotonising unclear speech. This might therefore have resulted in artificially low intelligibility scores for conditions (i) and (iii). This would

Table 6

Pairwise comparisons of the intelligibility in the four conditions: (i) slowly and clearly, (ii) quickly and unclearly, (iii) quickly and clearly, and (iv) slowly and unclearly. Note that correcting for multiple analyses using the Bonferroni correction renders an alpha-level of 0.0042. Significant differences are shaded in grey.

Paired differences									
L1	Condition		Std.	Std. Error	95% CI		*t*	df	*p*
		Mean	Deviation	Mean	Lower	Upper			
Danish	(ii)–(iii)	1.40	14.03	1.98	−2.59	5.38	0.70	49	0.49
	(i)–(iv)	5.77	13.05	1.85	2.06	9.48	3.13	49	0.003
	(ii)–(iv)	−3.63	16.56	2.34	−8.34	1.07	−1.55	49	0.13
	(i)–(iii)	10.80	13.40	1.89	6.99	14.60	5.70	49	>0.001
Norwegian	(ii)–(iii)	0.46	24.90	3.52	−6.62	7.54	0.13	49	0.90
	(i)–(iv)	16.82	23.84	3.37	10.04	23.59	4.99	49	>0.001
	(ii)–(iv)	−5.87	22.23	3.14	−12.18	0.45	−1.87	49	0.07
	(i)–(iii)	23.14	21.78	3.08	16.96	29.33	7.51	49	>0.001
Swedish	(ii)–(iii)	3.64	19.02	2.69	−1.77	9.04	1.35	49	0.18
	(i)–(iv)	14.02	24.56	3.47	7.04	21.00	4.04	49	>0.001
	(ii)–(iv)	−4.77	24.00	3.39	−11.59	2.05	−1.41	49	0.17
	(i)–(iii)	22.43	21.69	3.07	16.26	28.59	7.31	49	>0.001

Another result from the mixed-effects ANOVA is that for all three L1 groups of participants duration as well as reduction have a significant main effect on intelligibility: Content words in slowly produced sentences are significantly more intelligible than in quickly produced sentences ($F(1,147) = 92.58$, $p < 0.001$), and content words in clearly produced sentences are significantly more intelligible than unclearly produced sentences ($F(1,147) = 17.92$, $p < 0.001$).

Finally, another trend becomes clear from Table 5: While the slowly and clearly produced sentences are most intelligible to all three groups of listeners, the quickly and unclearly produced sentences are not the least intelligible. Rather, the quickly and clearly produced sentences are slightly less intelligible. This suggests that all three groups of listeners are able to compensate better for reduction phenomena when the amount of reduction fitted to the articulation rate than when there was actually enough time for a clear production, but the pronunciation remained unclear. This phenomenon has been described and discussed by Schüppert et al. (2012) for native speakers of Danish listening to spoken Danish. The results reported here extend these findings to native speakers of Norwegian and Swedish listening to spoken Danish.

3.2. Interaction effects between L1, duration and reduction on intelligibility

The mixed-effects ANOVA further reveals that there is a significant interaction effect of duration and L1 ($F(2,147) = 3.52$, $p = 0.03$). We assume, however, that the significance of this effect is mainly due to the fact that the Danish control group scores near the ceiling. This can also be seen from Fig. 7. Due to this ceiling effect, the effect of duration and reduction are artificially reduced for the native speaker

control group, which is likely to result in an underestimation of the effects that would be measurable in an experiment explicitly designed for native speakers of Danish. In other words, the factor 'L1' also contains the confounding factor of 'nativeness versus non-nativeness.'

The interaction effect between reduction and L1 is not significant ($F(2,147) = 1.93$, $p = 0.15$). This suggests that the effect of duration differs significantly across the three language groups, but the effect of reduction does not. To scrutinise this result, we conducted pairwise comparisons within each level of the factor. The results are given in Table 6. Note that correcting for multiple analyses using the Bonferroni correction renders an alpha-level of $p = 0.0042$ (0.05 divided by 12). Significant differences are shaded in grey.

Crucially, the interaction effect between duration and reduction is highly significant ($F(1,147) = 39.52$, $p < 0.001$), suggesting that the difference in the intelligibility of clear and unclear speech is significantly modulated by the duration of the sentences. Contrasts showed that slowly produced sentences are more intelligible if they are clearly pronounced ($M = 66.9\%$) than if they are unclearly pronounced ($M = 54.7\%$), while quickly produced sentences are slightly more intelligible if they are unclearly pronounced ($M = 49.9\%$) than if they are clearly pronounced ($M = 48.1\%$). This interaction effect is found in all three groups of listeners, as shown by a non-significant 3-way interaction-effect of L1, reduction and duration ($F(2,147) = 2.372$, $p = 0.1$). This finding suggests that slow speech is more intelligible when produced clearly, while fast speech is more intelligible when produced unclearly.

3.3. Relative contribution of L1, duration and reduction on intelligibility

Our second research question is which of the two intertwined factors *duration* and *reduction* has the largest effect on the intelligibility of Danish in Norwegian- and Swedish-

mean that the figures we report in this study might slightly underestimate the effect of reduction.

Table 7
Results from a multiple linear regression, predicting intelligibility of spoken Danish in native speakers of Norwegian and Swedish, with the factors duration and reduction using forced entry.

	B	SE B	Beta	t	p	Partial correlations	R^2
Constant	7.22	4.30		6.34	>0.001		
Duration	14.05	1.97	0.33	7.12	>0.001	0.34	0.11
Reduction	−6.68	1.97	0.16	−3.39	0.001	−0.17	0.03

Note model $R^2 = 0.14$.

speaking listeners. To answer this question, we conduct a multiple linear regression analysis on the intelligibility data from Norwegian- and Swedish-speaking participants only. Using forced-entry, the factors *reduction* and *duration* are entered into one model. This analysis reveals that, taken together, both factors explain about 13.5% of the variance in the intelligibility results.

The estimate for the factor *duration* is positive ($B = 14.05$), while the estimate for the factor *reduction* is negative ($B = −6.68$), which suggests that a longer duration is associated with higher intelligibility and a larger degree of reduction is associated with lower intelligibility. The estimate for the factor *duration* is also larger than the estimate for the factor *reduction*, which suggests that *duration* is the factor which has a larger impact on intelligibility, although both factors contribute significantly to the model. Table 7 summarises the results from the linear regression.

4. Conclusion

This paper took as a starting point the differences in articulation rate across the three mainland Scandinavian languages reported by Hilton et al. (2011). That paper showed that Danish is spoken significantly faster than Norwegian and Swedish. Their results were discussed in light of the finding that the intelligibility of spoken Danish by fellow Scandinavians, and in particular to Swedish-speaking listeners, is lower than other cases of mutual intelligibility in Scandinavia (Bø, 1978; Delsing and Lundin Åkesson, 2005; Maurud, 1976). The question that arose from Hilton et al. (2011) was whether this low intelligibility is due to the high articulation rate as such, which is assumed to result in higher demands on the working memory (Lilienthal et al., 2014; Mackworth, 1962; Tan and Ward, 2008), or whether it is mainly caused by the higher number of reduction phenomena which is linked to this high articulation rate.

The first step we took in the current paper to address these questions was to investigate whether the reported difference in speech tempo is in fact part of the reason why spoken Danish is so difficult to understand for Norwegians and Swedes. Our data show that spoken Danish is more intelligible when produced slowly and clearly than when produced quickly and unclearly, which suggests that either a high articulation rate by Danish-speaking news readers, or the large number of reductions that is associated with a high articulation rate, or both, are likely to impede the intelligibility for Norwegian- and

Swedish-speaking listeners. In other words, speaking slowly and clearly should improve intelligibility of spoken Danish to native speakers of Norwegian and Swedish. It also improves the intelligibility of spoken Danish to Danish-speaking listeners, but to a much lesser degree.

To investigate whether this improvement of intelligibility is due to a slow articulation rate making lower demands on the working memory, or whether the clear pronunciation made possible by the slow rate is the main cause for the improved intelligibility, the second research question aimed at tearing apart the two intertwined factors *duration* and *reduction* and shed light on whether these factors make Danish so hard to understand on individual bases. Our data suggest that *duration* has a larger impact in the intelligibility of spoken Danish than *reduction*. That means that speaking slowly increases intelligibility to a greater extent than speaking clearly does, although not surprisingly, the most efficient way of improving intelligibility is to speak slowly *and* clearly. This finding then, if taken to be universal for perception of speech in a L2, or a closely related speech variety, supports the approach that most speakers take intuitively when speaking to someone with a different L1 phonology. There are some possible applications of this finding. Professional fields such as intercultural communication and speech synthesis development can benefit from the iteration that, especially in multilingual contexts, a slow speed is crucial to intelligibility, and in a relative sense more important than articulating extremely clearly.

Our data show that sentences produced quickly and clearly are slightly less intelligible than sentences produced quickly and unclearly. This could be due to the effect of compression on the speech signal being larger than expansion for intelligibility, as pointed out by Vaughan and Letowski (1997), The finding confirms previous observations by Schüppert et al. (2012) for native speakers of Danish listening to spoken Danish. The observations thus far are in conflict with those made by Janse (2004), however, who found that artificially time-compressed Dutch speech is more intelligible than normally fast Dutch speech to Dutch listeners in a phoneme-detection task. Recent work indicates that the differences in findings between our study, and that of Janse (2004) has to do with the difference in language background of the listeners, rather than a difference in methodology. Rosink et al. (2014) conducted a translation task with 50 semantically unpredictable sentences in Dutch for Dutch speakers in a comparable study to ours. They confirm the findings reported by Janse (2004), calling to question whether the effects of compression on speech are language-specific or brought about by reduced recall capacities when L2, or dialect, speech is presented. We suggest future studies address this question, and call for studies using naturalistic stimuli so as to avoid the confounding effects of compression and monotonisation of manipulation on the speech signal. This answer needs to be addressed to come to a comprehensive account of phonology and speech perception in a non-native language.

Acknowledgments

Appendix

Sentence	No. of syllables		
	Canonical	Phonetic	
		Quick & unclear	Slow & clear
En regel synger efter et økonomisk udtryk.	14	8	12
En forskning indtræffer bag en europæisk karakter.	15	10	15
Et program udgår på en effektiv procent.	12	9	15
En person kommer i en religiøs nation.	12	6	11
En sommer flytter under en sikker handling.	12	8	10
En nyhed rejser over et demokratisk arbejde.	15	9	15
En turist regner efter en intellektuel effekt.	15	9	12
En fordel forsvinder bag en gammel forklaring.	13	10	11
En militær kunstner vinder en myndighed.	12	8	11
En effektiv amerikaner afgør et møde.	14	7	12
En mulig baggrund mærker en indsats.	10	8	10
En personlig samling hænger en generation.	13	10	14
En aktuel ekspert består en meter.	11	5	13
En historisk professor præsenterer et studium.	15	10	14
Et væsentligt behov kalder en roman.	11	6	10
Et centralt samarbejde udgør en mulighed.	13	10	13
En billig situation modsvarer en oplevelse.	15	9	13
En amerikansk regering studerer et ansvar.	14	9	14
En praktisk direktør forklarer en time.	12	9	11
En nordisk generation ordner en anledning.	13	7	12
Et socialt system skyder en gade.	10	7	11

Sentence	No. of syllables		
	Canonical	Phonetic	
		Quick & unclear	Slow & clear
Et ydre resultat køber en oplevelse.	13	9	13
Hvor forstår en litteratur en svensk indstilling.	13	10	15
Hvorfor glemmer et hoved et litterært parti.	13	8	13
Hvor påpeger en general en engelsk udvikling.	14	10	14
Hvorfor måler en radio et færdigt bidrag.	13	10	14
Hvor lægger et besøg en normal tradition.	12	9	14
Hvorfor henter en årsag en vigtig revolution.	14	10	15
Hvor vækker en kvinde en alvorlig ungdom.	12	8	14
Hvorfor foreslår en historie en rigtig glæde.	15	9	13
Hvor beslutter en institution en speciel produktion.	15	9	18
Hvorfor løser en præsident en ensom krone.	13	10	17
Hvor følger en undervisning en politisk befolkning.	15	11	15
Hvorfor behøver et bibliotek et nuværende museum.	18	8	16
Hvor kræver et nummer et teknisk forslag.	11	7	10
Hvorfor synger en forestilling en berømt diskussion.	15	10	15
En behandling skaffer en rejse som læser.	12	9	12
En kontakt vælger en forudsætning som vover.	13	10	13
En kritik træffer en frihed som eksisterer.	13	9	11
En linje behandler en organisation som sidder.	15	7	13
En patient rækker en arkitekt som accepterer.	14	10	14
En skole maler et område som venter.	12	8	12
En magt skaber en virksomhed som ligger.	11	8	10
En mening konstaterer en retning som forsøger.	14	10	14
En tanke anvender et middel som tænker.	12	7	13
Et forsøg udvikler en musik som hænder.	12	8	11
En mandag viser en vare som betaler.	12	10	13
En artikel flytter en side som spiller.	12	8	12
Et indtryk bygger en debat som fortsætter.	12	11	12
Et universitet udnytter en ordning som arbejder.	16	10	15

References

Baddeley, A.D., Thomson, N., Buchanan, M., 1975. Word length and the structure of short-term memory. J. Verbal Learn. Verbal Behav. 14, 574–589.

Basbøll, H., 2005. The Phonology of Danish. University Press, Oxford.

Benoît, C., Grice, M., Hazan, V., 1996. The SUS test: A method for the assessment of text-to-speech synthesis intelligibility using semantically unpredictable sentences. Speech Commun. 18, 381–392.

Bø, I., 1978. Ungdom og naboland. En undersøkelse av skolens og fjernsynets betydning for nabospråksforståelsen. [Youth and neighbouring country. An investigation of the influence of school and TV on inter-Scandinavian comprehension.] Stavanger: Rogalandsforskning.

Boersma, P., Weenink, D. 2013, Praat: doing phonetics by computer [Computer program]. Version 5.3.51, retrieved 30 September 2013 from http://www.praat.org/, (accessed 15.11.15).

Bond, Z.S., Moore, T.J., 1994. A note on the acoustic-phonetic characteristics of inadvertently clear speech. Speech Commun. 14 (4), 325–337.

Box, G.E.P., Hunter, W.G., Hunter, J.S., 1978. Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building. John Wiley and Sons, New York.

Bradlow, A.R., Kraus, N., Hayes, E., 2003. Speaking clearly for children with learning disabilities: Sentence perception in noise. J. Speech Lang. Hear. Research 46, 80–97.

Cannizzaro, M., Harel, B., Reilly, N., Chappell, P., Snyder, P.J., 2004. Voice acoustical measurement of the severity of major depression. Brain Cognition 56, 30–35.

Caspers, J., 1994. Pitch Movements Under Time Pressure. Holland Academic Graphics, Den Haag.

Caspers, J., van Heuven, V.J., 1993. Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. Phonetica 50, 161–171.

Conrad, R., Hille, B.A., 1958. The decay theory of immediate memory and paced recall. Can. J. Psychol. 12, 1–6.

Coupland, N., 1980. Style-shifting in a Cardiff work-setting. Lang. Soc. 9 (01), 1–12.

Darby, J.K., Hollien, H., 1977. Vocal and Speech Patterns of Depressive Patients. Folia Phoniatrica et Logopaedica 29, 279–291.

De Jong, N., Wempe, T., 2009. Praat script to detect syllable nuclei and measure speech rate automatically. Behav. Res. Methods 41 (2), 385–390.

Delsing, L.-O. Lundin Åkesson, K., 2005. Håller språket ihop Norden? En forskningsrapport om ungdomars förståelse av danska, svenska och norska [Does the language keep together the Nordic countries? A research report of mutual comprehension between Young Danes, Swedes and Norwegians]. Copenhagen: Nordiska ministerrådet.

Engstrand, O., 2004. Fonetikens Grunder. Studentlitteratur.

Ernestus, M., Hanique, I., Verboom, E., 2015. The effect of speech situation on the occurrence of reduced word pronunciation variants. J. Phon. 48, 60–75.

Ferguson, S., Quené, H., 2014. Acoustic correlates of vowel intelligibility in clear and conversational speech for young normal-hearing and elderly hearing-impaired listeners. J. Acoust. Soc. Am. 135, 3570–3584.

Ferguson, S.H., Poore, M.A., Shrivastav, R., 2010. Acoustic correlates of reported clear speech strategies. J. Acad. Rehabil. Audiol. 43, 45–64.

Fosler-Lussier, E., Morgan, N., 1999. Effects of speaking rate and word frequency on pronunciations in conversational speech. Speech Commun. 29, 137–158.

Fourakis, M., 1991. Tempo, stress, and vowel reduction in American English. J. Acoust. Soc. Am. 90, 1816–1827.

Gahl, S., Yao, Y., Johnson, K., 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. J. Memory Lang. 66 (4), 789–806.

Goldman-Eisler, F., 1968. Psycholinguistics: Experiments in Spontaneous Speech. Academic Press, London/ New York.

Gooskens, C., 2006. Linguistic and extra-linguistic predictors of Inter-Scandinavian intelligibility. In: van de Weijer, J., Los, B. (Eds.), Linguistics in the Netherlands, 23. John Benjamins, Amsterdam, pp. 101–113.

Gooskens, C., 2007. The contribution of linguistic factors to the intelligibility of closely related languages. J. Multiling. Multicult. Dev. 6, 445–467.

Gooskens, C., Hilton, N.H., 2013. The effect of social factors on the comprehension of a closely related language. In: Tirkkonen, J.-M., Anttikoski, E. (Eds.), Proceedings of the 24th Scandinavian Conference of Linguistics, 5. Publications of the University of Eastern Finland, pp. 201–210 Reports and Studies in Education, Humanities, and Theology.

Gooskens, C., Van Heuven, V.J., van Bezooijen, R., Pacilly, J., 2010. Is spoken Danish less intelligible than Swedish. Speech Commun. 52, 1022–1037.

Guy, G.R., 1980. Variation in the group and the individual. In: Labov, W. (Ed.), Locating language in time and space. Academic Press, New York, pp. 1–36.

Guy, G.R., 1992. Explanation in variable phonology: an exponential model of morphological constraints. Lang. Var. Change 3, 1–22.

Haugen, E., 1966. Semicommunication: the language gap in Scandinavia. Sociol. Inq. 36, 280–297.

Hazan, V., Markham, D., 2004. Acoustic-phonetic correlates of talker intelligibility for adults and children. J. Acoust. Soc. Am. 116, 3108–3118.

Hilton, N.H., Gooskens, C., Schüppert, A., 2011. Syllable reduction and articulation rates in Danish, Norwegian and Swedish. Nordic J. Linguist. 34, 215–237.

Hilton, N.H., Gooskens, C., Schüppert, A., 2013. The influence of non-native morphosyntax on the intelligibility of a closely related language. Lingua 137, 1–18.

Hirsch, F., Bouarourou, F., Vaxelaire, B., Monfrais-Pfauwadel, M.C., Bechet, M., Sturm, J., Sock, R., 2008. Formant structures of vowels produced by stutterers in normal and fast speech rates. Proceedings of the 8th International Seminar On Speech Production.

Jacewicz, E., Fox, R.A., O'Neill, C., Salmons, J., 2009. Articulation rate across dialect, age, and gender. Lang. Var. Change 21 (2), 233–256.

Janse, E., 2004. Word perception in fast speech: artificially time-compressed vs. naturally produced fast speech. Speech Commun. 42, 155–173.

Jurafsky, D., Bell, A., Gregory, M., Raymond, W.D., 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In: Bybee, J., Hopper, P. (Eds.), Frequency and the Emergence of Linguistic Structure, 45. Benjamins, Philadelphia, pp. 229–254.

Kristoffersen, G., 2000. The Phonology of Norwegian. Oxford University Press on Demand.

Kürschner, S., Gooskens, C., van Bezooijen, R., 2008. Linguistic determinants of the intelligibility of Swedish words among Danes. Int. J. Humanit. Arts Comput. 83–100.

Labov, W., 1966. The Social Stratification of English in New York City. Cambridge University Press, Cambridge.

Labov, W., Cohen, P., 1967. Systematic Relations of Standard and Non-Standard Rules in the Grammars of Negro Speakers Project literacy reports 8. Cornell University, Ithaca, NY.

Labov, W., Cohen, P., Robins, C., Lewis, J., 1968. A Study of the Non-Standard English of Negro and Puerto Rican Speakers in New York City, vol. 1. US Office of Education, Washington, DC Final research report, Cooperative Research Project No. 3288.

Ladd, R., 1996. Intonational Phonology(Cambridge Studies in Linguistics 79). Cambridge University Press, Cambridge.

Lam, J., Tjaden, K., Wilding, G., 2012. Acoustics of clear speech: effect of instruction. J. Speech Lang. Hear. Res. 55, 1807–1821.

Lilienthal, L., Hale, S., Myerson, J., 2014. The effects of environmental support and secondary tasks on visuospatial working memory. Mem. Cognition 42, 1118–1129.

Lindblom, B., 1990. Explaining phonetic variation: A sketch of the H&H theory. In: Hardcastle, W., Marchal, A. (Eds.), Speech Production and Speech Modeling. Kluwer, Dordrecht, pp. 403–439.

Mackworth, J.F., 1962. Presentation rate and immediate memory. Can. J. Psychol. 16, 42–47.

Maurud, Ø., 1976. Nabospråksforståelse i Skandinavia. En undersøkelse om gjensidig forståelse av tale- og skriftspråk i Danmark, Norge og Sverige [Neighboring language comprehension in Scandinavia. An investigation of mutual comprehension of written and spoken language in Denmark, Norway, and Sweden]. Stockholm: Nordiska rådet.

McCloy, D.R., Wright, R.A., Souza, P., 2015. Talker versus dialect effects on speech intelligibility: a symmetrical study. Lang. Speech 58, 371–386.

Moulines, E., Verhelst, W., 1995. Time-domain and frequency-domain techniques for prosodic modification of speech. Speech Coding Synth. 519–555.

Neu, H., 1980. Ranking of constraints on t,d deletion in American English: a statistical analysis. In: Labov, W. (Ed.), Locating Language in Time and Space. Academic Press, New York, pp. 37–54.

Osgood, C.E., Suci, G.J., Tannenbaum, P.H., 1957. The Measurement of Meaning. University of Illinois Press, Urbana, IL.

Picheny, M.A., Durlach, N.I., Braida, L.D., 1986. Speaking clearly for the hard of hearing II: acoustic characteristics of clear and conversational speech. J. Speech Hear. Res. 29, 434–446.

Posner, M.I., 1964. Rate of presentation and order of recall in immediate memory. Br. J. Psychol. 55, 303–306.

Raymond, W.D., Dautricourt, R., Hume, E., 2006. Word-internal/t, d/deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. Lang. Var. Change 18, 55–97.

Robb, M., Maclagan, M., Chen, Y., 2004. Speaking rates of American and New Zealand varieties of English. Clin. Ling. Phon. 18, 1–15.

Rosen, K.M., Folker, J.E., Murdoch, B.E., Vogel, A.P., Cahill, L.M., Delatycki, M.B., Corben, L.A., 2011. Measures of spectral change and their application to habitual, slow, and clear speaking modes. Int. J. Speech Lang. Pathol. 13, 165–173.

Rosink, S., Kroon, M., Heeswijk, L., Schüppert, A., 2014. De verstaanbaarheid van natuurlijk snelle en kunstmatig versnelde spraak in het Nederlands. Dutch J. Appl. Ling. 3 (2), 197–210.

Schüppert, A., Gooskens, C., 2011. Investigating the role of language attitudes for perception abilities using reaction time. Dialectologia 2, 119–140.

Schüppert, A., Hilton, N.H., Gooskens, C., 2015. Swedish is beautiful, Danish is ugly? Investigating the link between language attitudes and intelligibility. Linguistics 53 (2), 275–304.

Schüppert, A., Hilton, N.H., Gooskens, C., Van Heuven, V.J., 2012. Syllable deletion in contemporary Danish. Copenhagen Studies in Language. In: Heegård, J., Henrichsen, P.J. (Eds.), Speech in Action. Proceedings of the 1st SJUSK Conference on Contemporary Speech Habits. Samfundslitteratur, Copenhagen, pp. 73–97.

Schüppert, A., Pacilly, J., Gooskens, C., and Van Heuven, V.J., [in prep.]. Praat script to quantify speech reduction automatically.

Smiljanic, R., Bradlow, A.R., 2005. Production and perception of clear speech in Croatian and English. J. Acoust. Soc. Am. 118, 1677–1688.

Smiljanic, R., Bradlow, A.R., 2008. Temporal organization of English clear and conversational speech. J. Acoust. Soc. Am. 124, 3171–3182.

Tan, L., Ward, G., 2008. Rehearsal in immediate serial recall. Psychon. Bull. Rev. 15, 535–542.

Teasdale, J.D., Fogarty, S.J., Williams, J.M.G., 1980. Speech rate as a measure of short-term variation in depression. Br. J. Soc. Clin. Psychol. 19, 271–278.

Tsao, Y.C., Weismer, G., Iqbal, K., 2006. Interspeaker variation in habitual speaking rate: additional evidence. J. Speech Lang. Hear. Res. 49 (5), 1156–1164.

Van Borsel, J., De Maesschalck, D., 2008. Speech rate in males females, and male-to-female transsexuals. Clin. Ling. Phon. 22, 679–685.

Vaughan, N.E., Letowski, T., 1997. Effects of age, speech rate, and type of test on temporal auditory processing. J. Speech Lang. Hear. Res. 40, 1192–1200.

Wolfram, W., 1969. A Sociolinguistic Description of Detroit Negro Speech.

Wright, R., 1997. Lexical Competition and Reduction in Speech: A Preliminary Report. Indiana University, Bloomington, IN, pp. 1996–1997 Progress Report No. 21.

Zue, V.W., Laferriere, M., 1979. Acoustic study of internal t,d in American English. J. Acoust. Soc. Am. 66, 1039–1050.