Why is Danish so difficult to understand for fellow Scandinavians?\textsuperscript{a}

Anja Schüppert\textsuperscript{b}, Nanna H. Hilton, Charlotte Gooskens

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

Received 10 February 2015; received in revised form 9 February 2016; accepted 10 February 2016
Available online 24 February 2016

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.

Keywords: Danish; Norwegian; Swedish; Intelligibility; Speech reduction; Articulation rate.

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 lexi-co-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 Riksrådskasting (NRK) was used.1

The findings that Danish is spoken more quickly than Norwegian and Swedish when it is read by professional newsreaders 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 (‘hypo’- and ‘hypoa-rticulation’) 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
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 intelli-
gibility of Danish for fellow Scandinavians more, and if this
is the case, which one.

2. Method

The stimulus material consisted of 50 semantically unpre-
dictable sentences (henceforth SUS) that were read aloud by
a native speaker of Danish in two different, yet natural, con-
ditions: (i) at a slow speaking rate with a deliberately accurate
pronunciation, and (ii) at a high speaking rate with less ac-
curate 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 arti-
culated speech. The two additional conditions and the dif-
ferent 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
Benoit et al. (1996). These sentences were originally devel-
oped 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 syntacti-
cally 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 ba-
sic 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 ar-
ticulation 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 DanskOrd-
bogen, 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 indi-
cations given by DanskOrdbogen represent phonological sy-
lables, and use the term ‘canonical syllables’ for these. For
example, the word kirke (Engl. ‘church’) consists of two syl-
lables split at one point, namely kirke both following Dan-
skOrdbogen as well as Basbøll’s principles (2005:257). In ad-
tion 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, con-
strained 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 politicy 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 (con)
- 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 (in-
cluding 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
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...
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 F₀ 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 stimuli sentences sound unnatural and restricts their general intelligibility. McClory 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

<table>
<thead>
<tr>
<th>Condition</th>
<th>Duration (s)</th>
<th>No. of canonical syll.</th>
<th>No. of sonority peaks</th>
<th>Articulation rate</th>
<th>No. of spectral changes/s</th>
<th>F1 (Hz)</th>
<th>F2 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) Quick and unclear</td>
<td>1.8</td>
<td>13.1</td>
<td>8.7</td>
<td>4.8</td>
<td>1532</td>
<td>929</td>
<td>272</td>
</tr>
<tr>
<td>(ii) Slow and clear</td>
<td>3.0</td>
<td>13.1</td>
<td>13.0</td>
<td>4.3</td>
<td>1389</td>
<td>917</td>
<td>262</td>
</tr>
</tbody>
</table>
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>
<thead>
<tr>
<th>Phoneme</th>
<th>Danish realisation</th>
<th>Norwegian realisation</th>
<th>Swedish realisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>b</td>
<td>p</td>
<td>p</td>
</tr>
<tr>
<td>t</td>
<td>d</td>
<td>t</td>
<td>t</td>
</tr>
<tr>
<td>k</td>
<td>ɐ</td>
<td>k</td>
<td>k</td>
</tr>
<tr>
<td>b</td>
<td>ɐ</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>d</td>
<td>ɑ̃</td>
<td>d</td>
<td>d</td>
</tr>
<tr>
<td>g</td>
<td>j/ ʊ</td>
<td>q</td>
<td>q</td>
</tr>
<tr>
<td>v</td>
<td>ʊ</td>
<td>v</td>
<td>v</td>
</tr>
<tr>
<td>j</td>
<td>ɐ̃</td>
<td>j</td>
<td>j</td>
</tr>
<tr>
<td>r</td>
<td>ɐ̃</td>
<td>ɐ̃</td>
<td>ɐ̃</td>
</tr>
</tbody>
</table>

* Realised by a dental approximant in medial and final position.

Table 2 gives an example of phonetic realisation differences in the consonantal phonemes /p, t, k, b, d, g, v, j, t/ 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
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 Weenink, 2013). In step 1, the excitation levels for each frequency band are differentialised 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 differentialised 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

### Table 3

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

Fig. 4. Mean F1 and F2 values plotted against each other for the two originally recorded conditions (i) and (ii.)
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 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
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 *kvina* was counted as correct, as were the slightly misspelt translations (*kvina*) or typos such as (*kv:ina*). 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 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. Fig. 7 visualises these results.

![Intelligibility results split up per L1 and per condition.](image)

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.

<table>
<thead>
<tr>
<th>L1</th>
<th>Condition</th>
<th>(i)</th>
<th>(ii)</th>
<th>(iii)</th>
<th>(iv)</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Danish (control)</td>
<td>Slowly and clearly</td>
<td>94.6</td>
<td>85.2</td>
<td>83.8</td>
<td>88.8</td>
<td>88.1</td>
</tr>
<tr>
<td>Norwegian</td>
<td>Quickly and unclearly</td>
<td>62.0</td>
<td>39.3</td>
<td>38.8</td>
<td>45.1</td>
<td>46.3</td>
</tr>
<tr>
<td>Swedish</td>
<td>Quickly and clearly</td>
<td>44.2</td>
<td>25.4</td>
<td>21.7</td>
<td>30.2</td>
<td>30.4</td>
</tr>
<tr>
<td>Mean</td>
<td>Slowly and unclearly</td>
<td>66.9</td>
<td>49.9</td>
<td>48.1</td>
<td>54.7</td>
<td>54.9</td>
</tr>
</tbody>
</table>

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.

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

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-
speeching 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 (Lilenthal 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 monotomisation 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

This work was supported by the Netherlands Organisation for Scientific Research (NWO) [grant number 276-75-005]. The NWO was neither involved in the development of the study design, the collection, the analysis, the interpretation of data, nor in the writing process nor the decision to submit the article for publication. We thank two anonymous reviewers for their valuable comments, which helped us to improve this manuscript significantly.

Appendix

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

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