Lecturing in L1 Dutch and in L2 English. A pairwise comparison of speech samples of three HE lecturers

Anja Schüppert[*], Penny Heisterkamp

Center for Language and Cognition Groningen, University of Groningen, the Netherlands

ABSTRACT

This paper investigates how the language of instruction in Dutch higher education (Dutch versus English) affects speech production by L1 Dutch-speaking lecturers. In a pairwise design, three young lecturers that were highly proficient in English gave two comparable lectures each (L1 Dutch and L2 English). Results show that the L1 Dutch lectures were consistently given at slightly higher syllabic speech and articulation rates and that filled pauses were shorter and occurred less often in Dutch than in English lectures. In addition, L1 Dutch lectures contained a more diverse vocabulary and showed pitch patterns which have been shown to be associated with greater liveliness and higher perceived charisma of the speakers. We discuss possible reasons for the observed acoustic differences and the potential impact of our findings in the light of the ongoing transition from Dutch-medium instruction to English-medium instruction in Dutch higher education.

© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The number of English-medium instruction programmes is increasing rapidly all over the world. In Europe, the Netherlands is among the countries with the highest proportion of EMI programmes (Maiworm & Wächter, 2002; Wächter & Maiworm 2008, 2014). In many cases, this means that the number of Dutch-medium instruction (DMI) programmes is decreasing as the continuation of entirely mirrored programmes is considered inefficient and financially unprofitable (Wilkinson, 2013, p. 10). In practice, this does not only mean that many students have fewer programmes to choose from if they wish to study in their native language, but also that higher education (HE) lecturers increasingly teach in a language that is not their native language. What are the implications of this for the lectures that students attend in this setting?

Previous research into the nature of EMI lectures suggests that EMI speech is slower (Hincks, 2010; Thøgersen & Airey, 2011; Vinke, 1995; Vinke, Snippe, & Jochems, 1998) and more formal (Thøgersen & Airey, 2011) than lecturers' native-language speech. Some investigations report more disfluencies in EMI speech (Thøgersen & Airey, 2011), others fewer (Vinke et al., 1998).

In a pioneering study, Vinke (1995:108) reported that L1 Dutch-speaking lecturers produced 13% fewer words per minute when they were teaching in L2 English than when they were teaching in Dutch. However, in an analysis based on words rather than concepts, differences in orthography may bias results in one or the other direction. Therefore, generally, lexical speech and lexical articulation rates seem less suitable for a cross-linguistic comparison.

---

* Corresponding author. Center for Language and Cognition Groningen, University of Groningen, Box 716, 9700, AS, Groningen, the Netherlands.
E-mail address: a.schueppert@rug.nl (A. Schüppert).

Hincks (2010) investigated the speech of 14 L1 Swedish-speaking Master students, who each gave the same presentation twice: once in L2 English and once in L1 Swedish. Her results reveal that the presentations were significantly longer (about 11%) and that the speakers' mean length of runs (MLR, or number of syllables between pauses) was significantly shorter in the speakers' L2 English than in their L1 Swedish. In line with this, speaking rate (here as defined by the number of syllables produced per second) was significantly lower in L2 English than in L1 Swedish, with syllabic speech rate in L2 English being 23% slower than in L1 Swedish. The results also revealed that, when the presentation content is analysed in relation to the length of the presentation, speakers include significantly fewer pieces of information in their L2 than in their L1.

In a comparable setup, Thøgersen and Airey (2011) analysed speech from one L1 Danish-speaking university lecturer who gave three lectures in Danish and two in English. In line with Hincks's (2010) results, they report that the lecturer talked on average 21.5% longer in L2 English than in L1 Danish and that his speech rate was 23% lower in L2 English than in L1 Danish. In addition, L1 Danish segments contained 7% more syllables than L2 English segments, and L1 Danish lectures were more informally taught than L2 English lectures.

The current paper takes the aforementioned studies by Hincks (2010) and Thøgersen and Airey (2011) as points of departure to explore how language of instruction is linked to differences in L1 Dutch-speaking lecturers' EMI and DMI speech. The project partly replicates previous studies by analysing audio recordings of teachers' EMI and DMI speech samples and extends the investigation with additional analyses that explore a number of speech features, such as pitch movements. As it has been shown that L2 speech typically employs a smaller pitch range and less distinct pitch peaks than L1 speech (Pickering, 2004), which reduces the distinctiveness of prominent syllables and thereby hampers intelligibility, we investigate the speakers' pitch in the two conditions using a number of measures.

2. Research question

In this paper, we investigate whether language of instruction affects (1) speech and articulation rate, (2) disfluencies, (3) lexical diversity, and (4) liveliness of speech. The four measures will be introduced in depth in the next section. The lecturers are three native speakers of Dutch who each give the same lecture twice: once in L1 Dutch and once in L2 English. Lecture hall, audience and technical equipment were kept constant for both lectures. This design enabled us to conduct a direct pairwise comparison of three lecturers in highly comparable settings. The next section summarises previous findings about differences between L1 and L2 speech for features (1)–(4) among highly proficient L2 speakers.

3. L1 and L2 speech

Since the 1970s, a large number of studies within the field of Second Language Acquisition research have investigated the mechanisms behind L2 speech production (for an overview, see e.g. Archibald, 1998). However, the focus of many of these studies has been on beginning or intermediate learners' speech. Structural investigations of L2 speech in highly proficient L2 speakers were not conducted on a large scale until the 1990s (Hyltenstam, Bartning, & Fant, 2018), and insights into L2 acoustic features are still limited overall. This section provides a brief overview of findings concerning L1 and L2 speech features with a focus on highly proficient L2 speakers — the group of speakers that the vast majority of university lecturers belong to in the Netherlands.

3.1. Speech and articulation rates

Traditionally, speech rate has been defined as the number of linguistic units per time unit including pauses, while articulation rate has been defined as the number of linguistic units per time unit excluding any pauses (Tsao, Weismer, & Iqbal, 2006). A number of studies have investigated differences in L1 versus L2 speech rate. They almost unanimously report a higher speech rate for L1 than for L2 samples.

In a crossed design, Raupach (1980) explored how L1 German- and L1 French-speaking students described a cartoon in their L1 compared to their L2 French or German. Their findings reveal that the participants had lower speech and articulation rates when they spoke in their L2 than their L1. In a comparable crossed setup, Trouvain and Möbius (2014) analysed the number of L1 and L2 French and German syllables produced per second as well as phones per second and words per minute in L1 speakers of German and French. Their data confirmed that speech rates were higher in the speakers' L1 than in their L2. In a follow-up study, Trouvain, Fauth, and Möbius (2016) showed that L2 speech contains more pauses and more disfluencies than L1 speech.

These results are in line with findings by De Jong, Groenhout, Schoonen, and Hulstijn (2015), who reported that syllables produced by L1 English and L1 Turkish learners of L2 Dutch had a longer duration in L2 Dutch than in the speakers' L1. However, De Jong et al.'s (2015) study also revealed that L1 and L2 speech rate correlate significantly, as speakers with a higher L1 speech rate generally speak faster in their L2 as well. This points to important interpersonal differences that are maintained in L2 speech or even transferred from L1 to L2 speech. Overall, however, L2 speech can be described as being slower than L1 speech.

3.2. Disfluencies

Speech fluency is generally regarded as the ability to transfer a message with as few filled pauses, repetitions, repairs, and corrections as possible (Chambers, 1997). Consequently, chunks such as pauses, repetitions, repairs and corrections are called disfluencies.

In a pairwise design, Riazantseva (2001) investigated L1 Russian and highly vs. intermediate L2 English pausing patterns of native speakers of Russian. Her results showed that those speakers who were highly proficient in L2 English paused less often than those who spoke L2 English at an intermediate level. However, both Russian groups paused more often in English than a group of native English speakers.

Decreased fluency in a second language has also been reported in other studies. Tavakoli (2011) found that L2 English speakers with various L1 backgrounds paused more often and longer in English than L1 speakers of English. Similarly, a study by De Jong et al. (2015) revealed that pauses made by L1 Turkish and L1 English-speaking participants were longer and more frequent when they spoke in L2 Dutch than when they spoke in their native language. Apart from pauses, speech may be disrupted by other types of disfluencies, such as repetitions, repairs and corrections. For instance, De Jong et al. (2015) reported that their participants' speech contained significantly more repetitions and corrections in L2 Dutch than in L1 Turkish and L1 English.

However, there are also indications that L2 fluency is linked to L1 fluency. In their longitudinal study, Derwing, Munro, Thomson, and Rossiter (2009) found a correlation between L1 fluency and L2 fluency of Mandarin and Slavic learners of English. This correlation, however, was only significant for both L1 groups at the beginning of the study. These findings suggest that, for some more proficient L2 speakers, L1 fluency has little influence on L2 fluency.

3.3. Lexical diversity

Lexical diversity generally refers to the number of unique words ('types') in relation to the total number of words ('tokens') that a speaker uses (e.g., Dewaele & Pavlenko, 2003). A number of studies in the field of L2 development point to differences in lexical diversity between L1 and L2 speech. Bulté, Housen, Pierrard, and Van Daele (2008) studied various lexical aspects of the speech of children living in Brussels. In a French retelling task, the L1 Dutch-speaking learners of French scored significantly lower on lexical diversity measures than their L1 French-speaking peers. However, their lexical diversity showed significant improvements after one additional year of French classes, probably due to the learners' improved proficiency in French. De Jong, Steinel, Florijn, Schoonen, and Hulstijn (2012) analysed the performance of L1 and L2 speakers of Dutch in various speaking tasks. In line with results reported by Dewaele and Pavlenko (2003) and Bulté et al. (2008), they found that the speech fragments of the L1 speakers were significantly more diverse than those produced by L2 speakers, irrespective of the Dutch proficiency of the latter group.

However, data collected by Tavakoli and Foster (2008) and Foster and Tavakoli (2009) revealed that the relation between L2 proficiency and lexical diversity is not as straightforward as it may seem. In their study, intermediate learners of English from Teheran created less diverse stories than native speakers of English did, while narratives produced by intermediate learners in London were as diverse as the stories of the L1 English speakers. This indicates that L2 learners with similar levels of proficiency can relate differently to the native-speaker level for certain linguistic domains. Together, these studies suggest that L2 speakers often, but not consistently show lower levels of lexical diversity than L1 speakers.

3.4. Pitch patterns

L1 and L2 communication might also differ in the perceived liveliness of speech. Arguably, factors such as gestures, facial expressions, and interactions with the interlocutors affect the perceived liveliness, but these aspects are not analysed here due to the limited scope of the study. When it comes to acoustic traits, Traunmüller and Eriksson (1995) showed that speech rate affects the perceived liveliness (as assessed by human raters), with a higher speech rate being linked to higher liveliness ratings. This result, however, was not confirmed by Hincks (2005). Instead, Hincks (2005) reported that liveliness of speech significantly correlates with the standard deviation of pitch in relation to mean pitch, which is quantified by the Pitch Variation Quotient (PVQ). The PVQ is calculated by dividing the standard deviation of the pitch (Hz) by the mean pitch (Hz). In Hincks' study, the liveliness scores were obtained from 8 judges who rated the speech of 18 speakers. Hincks (2005) reported a significant and strong positive correlation between PVQ and liveliness scores. Finally, a recent study by Niebuhr and Skarnitzl (2019) revealed that mean and median pitch correlate positively, and pitch kurtosis correlates negatively with perceived charisma (again, as assessed by human raters).

Importantly, however, language-specific differences need to be taken into account when analysing these data. In Mennen, Schaeffler and Docherty's study (2012), for instance, L1 English-speaking participants spoke with a significantly wider pitch range than L1 German-speaking participants. Thus, a reduction in the pitch range of an L2 learner could be due to a lower proficiency in the L2, but also to different language-specific pitch patterns of the second language - or both. To address this concern, Zimmerer, Jügler, Andreeva, Möbius, and Trouvain (2014) analysed sentences and narratives read aloud by L1 French learners of German and L1 German learners of French. Their data suggested that their participants' pitch range was significantly smaller in the L2 than in the L1, regardless of language. Although this study only compared L1 and L2 performance in two languages, it supports the claim that L1-L2 differences in pitch range can remain relatively constant across languages.

4. Method

4.1. Speakers and materials

In order to investigate the speech rate, the articulation rate, disfluencies, vocabulary size and pitch movements, spontaneous speech from three native speakers of Dutch was collected in two conditions: L1 Dutch and L2 English. For reasons of feasibility, we opted for short lectures of 7–10 min rather than the traditional 45- or 90-minute lectures. It has to be kept in mind that our data therefore might not be entirely representative of regular university lectures. In total, six lectures were recorded, resulting in about 48 min of speech. Importantly, the topics of the lectures were the same across conditions (although not across speakers).

All three lecturers were L1 Dutch-speaking beginning academic scholars from the University of Groningen (UG) in the Netherlands. The speech samples in this study were collected in the context of a larger project on the influence of EMI on Dutch students' content learning. The topics of the lectures were language and aging, the use of cochlear implants, and linguistic salience. The speakers were highly proficient in English and had all completed at least part of their tertiary studies in English (see Table 1). Given the high number of international staff at the UG, the speakers could be expected to make active use of English at work on a daily basis. All three speakers were preparing their PhD theses at the time of the recordings.

About two weeks prior to the recordings, the speakers were approached with the question whether they could give mirrored presentations in two languages for the project and prepare powerpoint slides in advance. They were asked to discuss the same content in the Dutch and English version of their talks. When the recordings were made, all speakers started with the lecture in L1 Dutch, followed by a short break, and then continued with the same lecture in L2 English. As the recordings were part of a larger research project, the lectures were video-taped as well. The audio files for the current recordings were recorded at 48 kHz using a Sennheiser clip-on microphone.

Subsequent inspection of the data showed that all speakers had closely aligned the English lecture with the Dutch lecture or vice versa. However, the three speakers took a different approach in the organisation of the lectures. The lectures taught by speaker A and speaker B mainly provided an introduction to a broader topic. Speaker A focused on the cognitive development of elderly and the activities that could prevent cognitive decline. Speaker B talked about the anatomy of the ear and the functioning of cochlear implants. Lecturer C, however, mainly talked about a study he conducted himself. Consequently, he provided less background information about the subject area in favour of a detailed discussion of the set-up and outcomes of the study. His lecture could therefore potentially be assumed to resemble a conference talk rather than a lecture.

The speakers gave their lectures in a lecture room in front of a small audience, which consisted of the same people for all six recordings. Using this paired setting, in which every speaker presented the content of their lecture twice, it was possible to investigate the influence of the language of instruction on speech in a semi-naturalistic setting while controlling for the potentially confounding effects of topic complexity.

4.2. Analyses

We analysed the speakers' lectures with regard to speech and articulation rates, speech fluency, lexical diversity, and liveliness of speech. This section describes how these various measures were quantified. Prior to the analyses, the six sound files were exported from the video recordings in wav-format and annotated in the phonetics software Praat version 6.0.46 (Boersma & Weenink, 2019) using TextGrid files.

Four tiers were created during the segmentation, but only two of them served as input tiers for the analyses reported in this paper:

(1) a 'word' tier, in which complete words were segmented and orthographically transcribed. This tier excluded any disfluencies. The words in this tier served as input for the calculation of speech and articulation rates as well as lexical diversity;

(2) a 'disfluency' tier, which categorised the different disfluencies (i.e. silent pauses, filled pauses, repairs, repetition, correction; see section 4.2.2). The categorisations in this tier served as input for the disfluency measures.

The two additional tiers were a 'sentence' tier, which rather arbitrarily divided the lecture into shorter fragments to facilitate the segmentation process, and a 'disfluency content' tier, which specified for some of the disfluencies what was produced in this segment. The four tiers are visualised in Fig. 1. The calculate-segment-durations Praat script (Lennes, 2017) was used to export the annotations as text files for the annotation-based analyses.

Table 1
The speakers' gender, age, educational background and lecture topic.

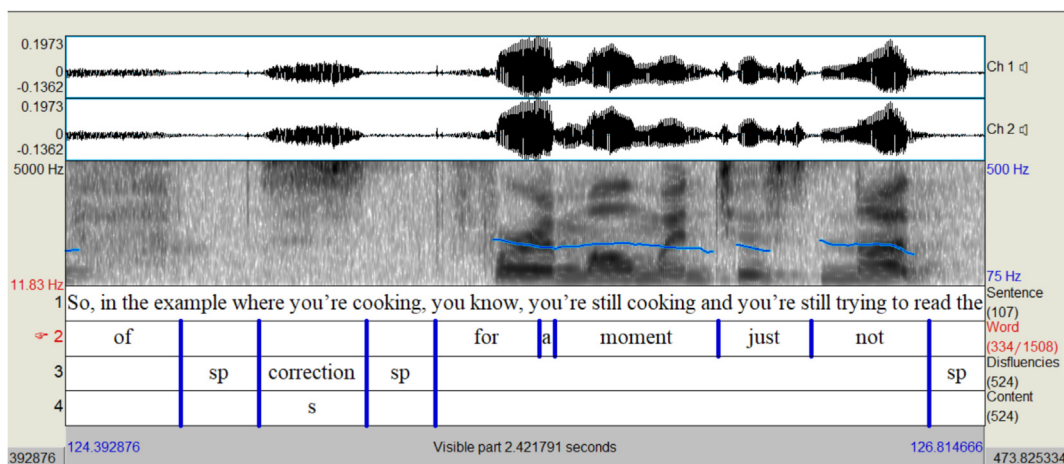| Speaker	Gender	Age	HE completed in English	Year PhD	Topic
A	Female	25	Master	1	Language and aging
B	Female	27	Bachelor, Master	2	Cochlear implants
C	Male	25	Master	2	Linguistic salience

Fig. 1. Oscillogramme, spectrogramme and TextGrid for a fragment of 2 s, illustrating the segmentation, transcription and annotation process.

4.2.1. Speech and articulation rates

We used different measures to explore how quickly a message is transferred by our three speakers and, in particular, across conditions. Firstly, following the approach by Hincks (2010) and Thøgersen and Airey (2011), we quantified the speakers' canonical speech rate in both conditions, which is defined as the number of canonical syllables per second including pauses. The *Cambridge English pronouncing dictionary* (Roach, Hartman, Setter, & Jones, 2006) was used to establish the number of canonical syllables per word, which were added up and divided by the total length of the lecture to establish the canonical speech rate per lecture. Repetitions, repairs, corrections, and filled pauses (see section 4.2.2 for more information) were excluded from the syllable count. For the calculation of the canonical articulation rate, the number of canonical syllables was divided by the length of the fragment excluding any silent and filled pauses.

Next, unlike Hincks (2010) and Thøgersen and Airey (2011), we also identified the speakers' phonetic speech rate per condition. This measure is defined as the number of phonetically detectable syllables per second and tells us more about how many phonetic, rather than underlying phonologically syllables were actually articulated per time unit. The phonetic speech rate is independent of phonology and orthography and is solely based on the acoustic information in the speech signal. Therefore, it might be particularly informative for cross-linguistic studies such as the current study, as cross-linguistic differences in spelling norms are eliminated. Phonetic speech rate was established using a Praat script written by De Jong and Wempe (2009). The script detects and counts all intensity peaks (defined as drops in intensity of 2 dB before and after the peak) in the parts of the speech signal that have voicing (defined as measurable F_0). A drawback of this fully automatic method is that some of the filled pauses were counted as phonetic syllables. Therefore, the number of phonetic syllables was manually corrected for this measurement error, and, subsequently, phonetic speech rate and reduction ratio were calculated based on the corrected figures. To establish phonetic articulation rate, the number of syllables was divided by the duration of the analysed sample without silent and filled pauses.

Naturally, due to reduction processes in practically any type of human speech, the number of phonetically detectable syllables is lower than the number of canonical syllables. Therefore, the syllabic speech rate based on canonical syllables is generally higher than the phonetic speech rate. In other words, the ratio of canonical syllables per phonetic syllable is generally above 1. The higher this syllable reduction ratio is, the more a speaker reduces syllables. We include this ratio in our analyses. Importantly, however, this ratio is highly language-specific (Robb, Maclagan, & Chen, 2004). Therefore, the degree of reduction has to be interpreted with great caution if several languages are involved.

4.2.2. Speech fluency

Speech fluency was explored by establishing the number and length of silent and filled pauses, repetitions, repairs and corrections during the whole lecture. These disfluencies were noted on a separate disfluency tier to prevent incomplete words (i.e. when corrections or repairs occur) from increasing lexical diversity artificially, and, likewise, to prevent repetitions from artificially reducing lexical diversity (see section 4.2.3). All silent intervals longer than 150 ms were interpreted as silent pauses. Filled pauses were specified as intervals in which the speaker produced a sound that did not carry lexical weight, such as "uh" or "ehm", irrespective of their length. Repetitions were defined as instances in which there was a literal repetition of a word or word group. For this disfluency, the second instance of the word, compound or phrase was marked as the repetition. In the case of repairs and corrections, the (first) item that was repaired or corrected was seen as the disfluency. We differentiated between repairs and corrections on the basis of the similarity between the initial and corrected word, compound or phrase. If the modification concerned the form or structure of the initial word, compound or phrase, the disfluency was classified as a repair. Repairs also included modifications that consisted of the repetition of a large part of the initial word,

compound or phrase and a modification. These cases were classified as repairs because the modified word groups were still very similar to the original word groups. If the modification involved the selection of a new lexical item or structure without meaningful repetition, it was seen as a correction. Examples (a), (b), (c) and (d) illustrate the distinction between the latter three types of disfluencies:

(a) and the functions of the <u>of the</u> ear (Speaker B)
(b) there is a relation with <u>cogne</u> eh cognitive processes (Speaker C)
(c) hoe <u>normale mensen</u> normaalhorende mensen (Speaker B)
(d) <u>and that you can</u> it literally means without nourishment (Speaker A)

Example (a) was classified as a repetition because it involves the literal repetition of 'of the'. In (b), 'cogne' is repaired to the related word 'cognitive'. Example (c) was categorised as a repair because the word 'mensen' and the structure of the modified item were repeated in the modification. The structure in (d) was considered to be a correction because it involved the selection of a new structure without meaningful repetition of elements from the modified chunk.

4.2.3. Lexical diversity

The 'word' tiers for all six recordings served as input for the lexical diversity analyses. As indicated in section 4.2.2, this tier did not contain any repairs, repetitions, or corrections in order to prevent distortions of lexical diversity.

Lexical diversity has been analysed with a variety of measures. The robustness of some of these measures, including the type-token ratio (TTR), has been found to be unreliable as they are affected by the length of a sample: a large number of tokens deflates the TTR (Van Hout & Vermeer, 2007), which makes the TTR less suitable for the analysis of speech fragments of different lengths. We therefore opted for two other measures: the Guiraud's index (Van Hout & Vermeer, 2007), in which the number of word types is divided by the square root of the number of tokens, and the *D* statistic using the programme D_Tools (Meara & Miralpeix, 2007). This programme calculates the TTR for 1600 random sub-samples of the fragment in question, each of which are between 35 and 50 words long. D_Tools then uses the formula proposed by Malvern, Richards, Chipere, and Duran (2004) to find a value of *D* that best matches this data set.

Both statistics are based on lemmas (types) and rather than words (tokens). SketchEngine was used to determine lemmas to all words. The lemmatising sets that were used were developed by Marcus, Santorini, and Marcinkiewicz (1993) for English, and by Schmid (1994; 1995) for Dutch. All six corpora were inspected after the automatic lemmatising and manually corrected for items that had not been identified correctly.

4.2.4. Pitch patterns

Finally, the liveliness or charisma of the speakers in both conditions was quantified by analysing the fundamental frequency (F_0). Pitch was extracted from the speech signal using Praat. The range was set to 75−500 Hz, and autocorrelation was selected as method of analysis. The silence threshold was set to 0.03 dB and the voicing threshold to 0.45.

In addition to analysing pitch mean and median (Hz), we followed Hincks's (2005) approach of calculating the PVQ. As a higher overall pitch (e.g. in female speakers) is often linked to a higher standard deviation (Haan & Van Heuven, 1999; Traunmüller & Eriksson, 1995), dividing the standard deviation of pitch (Hz) in every speaker by the mean pitch of that particular speaker can normalise for these differences. In contrast to Hincks (2005), however, we based our analysis on the mean and standard deviation for the two whole audio files for every speaker instead of separate shorter fragments of these files.

Furthermore, based on Niebuhr and Skarnitzl's (2019) findings, we explored the Pearson's kurtosis of the pitch distribution in our sample. Kurtosis is traditionally defined as the 'peakedness' or the 'tail thickness' of the distribution (in our case, the distribution of the Herz values in the sound signal per speaker and per language condition). A normal distribution per definition has a Pearson's kurtosis of 3. An adjusted version of Pearson's kurtosis is the 'excess kurtosis', which is Pearson's kurtosis minus 3. In the remainder of this paper, report excess kurtosis values. For a distribution with the same mean and standard deviation, but with 'thinner' inner tails (i.e. fewer outliers that are slightly deviating from the mean) and 'thicker' outer tails (i.e. more outliers that are strongly deviating from the mean) than a normal distribution, and, usually, a higher peak, kurtosis is higher, and excess kurtosis therefore is greater than 0. Accordingly, kurtosis is lower (and excess kurtosis is smaller than 0) for 'flatter' distributions with a 'thicker' inner tail and a 'thinner' outer tail (Westfall, 2014).

Both PVQ and kurtosis are measures of variation (again, in this case of the Hertz values per speaker per condition), but while PVQ is a measure that quantifies mainly how deviant the data points are from the mean *overall* (here: per speaker and condition), kurtosis tells us more about the *distribution* of the deviance.

5. Results

5.1. Speech rates, articulation rates, and reduction ratio in the speech signal

Firstly, we found that the lectures were about 9% shorter on average in L1 Dutch than in L2 English (see Table 2). This pattern could be observed in speakers A (17% shorter) and B (9% shorter), while speaker C talked slightly longer in Dutch than

in English (2% longer). The average difference of 9% is surprisingly comparable to the differences in Hincks's (2010) study, who reported that L1 Swedish presentations were about 10% shorter than L2 English presentations (according to the figures in Table 2 by Hincks, 2010). Thøgersen and Airey (2011), however, reported that L1 Danish lectures were even 17.7% shorter than L2 English lectures in lecture duration (according to the figures in Table 1 by Thøgersen & Airey, 2011).

Given that the content of the two lectures across the conditions was highly comparable in our study, this difference seems to suggest that the speakers' messages were generally transferred faster in their L1 than in their L2. Indeed, it turned out that the speakers had a tendency to speak faster in L1 Dutch than in L2 English, as indicated by a higher canonical speech and articulation rate in Dutch than in English, with L1 Dutch lectures having a 16% higher canonical speech rate than L2 English lectures on average. Again, this is well in line with Hincks' (2010) results, who reported that students' canonical speech rate was about 18.7% higher in L1 Swedish than in L2 English, while the difference of 30.6% reported by Thøgersen and Airey (2011) was substantially higher (according to the figures in their Table 1). The canonical articulation rate in our data showed a consistent increase of 7% to 22% in Dutch compared to English, which renders an average increase of 15%. Thøgersen and Airey (2011) reported a 30.4% difference (according to the figures in their Table 2). Our data thus supports this trend, but we find a less pronounced effect.

The difference in phonetic speech rate confirms the higher rate for L1 Dutch than for L2 English reported above. Interestingly, however, with L1 Dutch being spoken only 7% faster phonetically, this difference is less pronounced than the difference in canonical speech rate. This suggests that the speakers in our sample had a tendency to reduce L1 Dutch to a larger degree than L2 English. This was measured using the syllable reduction ratio. As stated above, the syllable reduction ratio was defined as the number of canonical syllables divided by the number of phonetic syllables. For all three speakers, this ratio is indeed between 1% and 16% higher in L1 Dutch than in L2 English with an average difference of 8%. We have to bear in mind, however, that it is unclear whether these differences are due to typological differences between the two languages (Dutch versus English) or due to differences in language proficiency (L1 versus L2).

Finally, the lexical speech and articulation rates observed in our speakers were on average 7% higher in L1 Dutch than in L2 English. This is a smaller difference than in Vinke's (1995) study, who reported a difference of 17% for L1 Dutch lecturers teaching in L1 Dutch and in L2 English. This could possibly be due to the fact that our speakers were more proficient in English than the speakers in Vinke's (1995) study, but this remains speculative. It also has to be noted that one of the three speakers,

Table 2

Length, number of syllables, speech and articulation rates and syllable reduction ratio for the three speakers in both conditions.

Measure	Language	Speaker A	Speaker B	Speaker C	Mean
Length (s)	Dutch	496	499	450	482
	English	599	550	440	530
	Dutch: English	0.83	0.91	1.02	0.91
Length excl. silent and filled pauses (s)	Dutch	403	441	371	405
	English	467	496	370	444
	Dutch: English	0.86	0.89	1.00	0.91
Words (N)	Dutch	1154	1537	1151	1281
	English	1196	1597	1140	1311
	Dutch: English	0.96	0.96	1.01	0.98
Canonical syllables (N)	Dutch	1929	2539	1809	2092
	English	1836	2434	1685	1985
	Dutch: English	1.05	1.04	1.07	1.05
Phonetic syllables (N)	Dutch	1249	1467	1176	1297
	English	1375	1422	1178	1325
	Dutch: English	0.91	1.03	1.00	0.98
Canonical speech rate (syll/s)	Dutch	3.89	5.09	4.02	4.33
	English	3.07	4.43	3.83	3.77
	Dutch: English	1.27	1.15	1.05	1.16
Canonical articulation rate (syll/s)	Dutch	5.20	6.01	5.12	5.44
	English	4.24	5.14	4.81	4.73
	Dutch: English	1.23	1.17	1.06	1.15
Phonetic speech rate (syll/s)	Dutch	2.52	2.94	2.61	2.69
	English	2.30	2.59	2.68	2.50
	Dutch: English	1.10	1.14	0.97	1.08
Syllable reduction ($syll_{Can}/syll_{Phon}$)	Dutch	1.54	1.73	1.54	1.61
	English	1.34	1.71	1.43	1.50
	Dutch: English	1.15	1.01	1.08	1.07
Lexical speech rate (wds/s)	Dutch	2.33	3.08	2.56	2.66
	English	2.00	2.90	2.59	2.48
	Dutch: English	1.17	1.06	0.99	1.07
Lexical articulation rate (wds/s)	Dutch	2.86	3.49	3.11	3.16
	English	2.56	3.22	3.08	2.95
	Dutch: English	1.12	1.08	1.01	1.07

speaker C, did not show a substantial difference in lexical speech and articulation rate between Dutch and English. As such, no large claims can be made about the lexical rates in this study.

5.2. Disfluencies

Pausing patterns are indirectly reflected in some of the measures reported in the previous section. In this section, a more detailed report of silent and filled pauses is provided along with data on corrections, repairs, and repetitions per speaker and condition.

When it comes to silent pauses, speaker A paused less often and shorter on average in Dutch than in English, while speaker C showed the exact opposite pattern: he used more and on average longer silent pauses in Dutch than in English. Speaker B displays yet another pattern with fewer, but longer silent pauses in Dutch than in English (see Table 3). In general, therefore, no consistent pattern can be found with regard to silent pauses across conditions, which confirms the ambiguous picture drawn by previous studies. Filled pauses were employed somewhat more often in L1 Dutch than in L2 English by speakers A and B, but for all speakers, filled pauses were substantially shorter in Dutch than in English.

Overall, the speakers used very few disfluencies aside from silent and filled pauses, which made it difficult to detect reliable patterns for the other disfluency categories. On average, the speakers tended to correct themselves less often in Dutch than in English, but the total duration of corrections was slightly longer. Generally, repairs were less frequent and shorter in L1 Dutch than in L2 English, while the opposite was true for repetitions: on average, the speakers tended to repeat themselves more often and use more time for repetitions in Dutch than in English. However, these patterns were not consistent across speakers.

In general, the L1 Dutch lectures contained slightly fewer and shorter disfluencies than the L2 English lectures. This could of course be merely an artefact of the lectures being slightly shorter in Dutch than in English. We therefore calculated the total number and total duration of disfluencies per second, which revealed that the Dutch lectures contained slightly more, but shorter disfluencies than the English lectures per time interval, which resulted in slightly less time spent on disfluencies overall. As there are quite substantial interpersonal differences, however, these data have to be interpreted with caution.

5.3. Lexical diversity

Table 4 shows Guiraud's indices and the *D* statistic for the three speakers in both conditions. As indicated in the table, all three speakers have lower Guiraud's indices for L2 English than for L1 Dutch, indicating slightly less lexical variation in English than in Dutch. Overall, lexical diversity was about 8% higher in L1 Dutch than in L2 English. For the *D* statistic, the difference between Dutch and English lectures is even more pronounced (see Table 4). Here, Dutch lectures showed a 20% higher lexical diversity than English lectures. Importantly, despite some inter-individual differences, both measures for lexical diversity showed consistent patterns in that vocabulary used by our lecturers seems to be more diverse in L1 Dutch than in L2 English.

5.4. Pitch patterns

Overall, speakers had a tendency to use a slightly higher median pitch during the L1 Dutch lectures than during the L2 English lectures (see Table 5). According to Niebuhr and Skarnitzl (2019), this indicates that the speakers might be perceived as slightly more charismatic in Dutch than in English. However, with an average of 2%, this difference was small. A comparable trend could be observed for the mean pitch, but here the effect was even less pronounced as one of the three speakers deviated from this pattern.

In contrast, the PVQ was generally higher for the English than for the Dutch lectures, suggesting more liveliness in L2 English compared to L1 Dutch according to Hincks's (2005) findings. The PVQ values for all six audio files were between 0.15 and 0.21, which means that our group of speakers displays considerably less inter-individual variation than the data analysed by Hincks (2005), in which the PVQ ranges from approximately 0.11 to 0.24 (according to her Figure 5). However, PVQ values in this study clearly represent the middle field of Hincks' (2005) data.

However, we find opposite patterns for charisma if the kurtosis values are interpreted in line with Niebuhr and Skarnitzl's (2019) findings, who reported a negative relation between kurtosis of pitch distribution and perceived charisma of a speaker. Based on the relatively low kurtosis values for Dutch in the current study, then, all our speakers can be assumed to appear more charismatic in Dutch than in English.

6. Discussion

The results in the previous section indicate that speech in L1 Dutch-speaking lecturers varies as a function of language of instruction, even if the speakers are highly proficient in L2 English. Overall, our speakers had a tendency to give their lectures in a shorter amount of time when they were lecturing in L1 Dutch. They managed to do so by consistently speaking faster, which manifested itself in two ways. Firstly, the speakers generally showed a higher degree of reduction in their L1 speech compared to their L2 speech. Secondly, speakers tended to show fewer but longer disfluencies in L2 English than in L1 Dutch. This pattern was inconsistent, however, depending on the type of disfluency as well as on the speaker, so conclusions have to

Table 3
Disfluencies in the lectures by speaker and language.

Disfluency type	Language	Speaker A	Speaker B	Speaker C	Mean
Silent pauses (N)	Dutch	150	103	154	137.67
	English	207	119	137	154.33
	Dutch: English	0.72	0.87	1.12	0.89
Total duration of silent pauses (s)	Dutch	92.64	58.00	79.32	76.65
	English	131.72	54.47	70.17	85.45
	Dutch: English	0.70	1.06	1.13	0.90
Average duration of silent pauses (ms)	Dutch	618	563	515	565.33
	English	636	458	512	535.33
	Dutch: English	0.97	1.23	1.01	1.06
Filled pauses (N)	Dutch	78	67	52	65.67
	English	70	65	53	62.67
	Dutch: English	1.11	1.03	0.98	1.05
Total duration of filled pauses (s)	Dutch	32.25	18.71	17.42	22.79
	English	34.26	22.18	19.85	25.43
	Dutch: English	0.94	0.84	0.88	0.90
Average duration of filled pauses (ms)	Dutch	413	279	335	342.33
	English	489	341	375	401.67
	Dutch: English	0.84	0.82	0.89	0.85
Corrections (N)	Dutch	7	7	4	6.00
	English	9	6	4	6.33
	Dutch: English	0.78	1.17	1.00	0.95
Corrections (s)	Dutch	6.94	2.72	1.71	3.79
	English	5.81	3.22	1.03	3.35
	Dutch: English	1.19	0.84	1.66	1.13
Repairs (N)	Dutch	7	11	9	9.00
	English	18	15	4	12.33
	Dutch: English	0.39	0.73	2.25	0.73
Repairs (s)	Dutch	2.95	4.36	3.94	3.75
	English	12.41	6.91	1.17	6.83
	Dutch: English	0.24	0.63	3.37	0.55
Repetitions (N)	Dutch	7	4	4	5.00
	English	4	6	2	4.00
	Dutch: English	1.75	0.67	2.00	1.25
Repetitions (s)	Dutch	2.23	0.81	0.86	1.30
	English	0.97	1.93	0.46	1.12
	Dutch: English	2.30	0.42	1.87	1.16
Total disfluencies (N)	Dutch	249	192	223	221.33
	English	308	211	200	239.67
	Dutch: English	0.81	0.91	1.12	0.92
Total disfluencies (s)	Dutch	137	85	103	108.33
	English	185	89	93	122.33
	Dutch: English	0.74	0.95	1.11	0.89
Disfluencies per second (N)	Dutch	0.50	0.37	0.50	0.45
	English	0.51	0.37	0.45	0.44
	Dutch: English	0.98	1.00	1.09	1.02
Disfluencies per second (s)	Dutch	0.28	0.16	0.23	0.22
	English	0.31	0.15	0.21	0.22
	Dutch: English	0.90	1.07	1.10	1.00

be drawn with caution. Interestingly, our findings for three L1 Dutch-speaking lecturers are generally well in line with the results reported by Hincks (2010) for L1 Swedish-speaking presenters, and with Thøgersen & Airey's (2011) findings for an L1 Danish-speaking lecturer, both of which served as starting points for our study.

Importantly, speech rate and reduction ratio have previously been reported to vary not only with different levels of proficiency (e.g. L1 vs. L2), but also across language varieties. Robb et al. (2004) report a higher articulation rate for L1 New Zealand English than for L1 American English. They argue that reduction processes and vowel raising in NZ English increase the articulation rate compared to American English. Hilton, Schüppert, and Gooskens (2011) report that the same message is generally transferred faster, i.e. in a shorter amount of time, in Danish than in its closely related languages Norwegian and Swedish, even when the message entirely consists of cognate words shared by the three varieties. In line with Robb et al.'s (2004) findings, their data suggests that reduction phenomena are the major cause for these differences in speech rate. To our knowledge, a systematic cross-linguistic investigation of speech rate for Dutch and English has not been conducted to date. The differences in speech and articulation rate across L1 Dutch and L2 English therefore need to be interpreted carefully as there are at least two factors involved: proficiency (L1 vs. L2) and language typology (Dutch vs. English). Our study cannot

Table 4

Lexical diversity per speaker per condition as quantified using Guiraud's indices and the *D* statistic.

	Language	Speaker A	Speaker B	Speaker C	Mean
Guiraud's index	Dutch	8.87	9.37	8.00	8.75
	English	8.35	8.04	7.82	8.07
	Dutch: English	1.06	1.17	1.02	1.08
D	Dutch	91.2	74.8	70.0	78.7
	English	72.8	60.0	64.4	65.7
	Dutch: English	1.25	1.25	1.09	1.20

Table 5

Mean and median pitch (Hz), standard deviation, PVQ, and kurtosis values per condition for the three speakers.

Measure	Language	Speaker A	Speaker B	Speaker C	Mean
Mean pitch (Hz)	Dutch	216.21	225.86	138.63	193.57
	English	210.76	221.52	140.07	190.08
	Dutch: English	1.03	1.02	0.99	1.02
Median pitch (Hz)	Dutch	212.08	221.74	133.45	189.09
	English	205.08	217.31	133.15	185.18
	Dutch: English	1.03	1.02	1.00	1.02
Standard Deviation (Hz)	Dutch	35.31	39.13	26.72	33.72
	English	32.01	39.80	30.04	33.95
	Dutch: English	1.10	0.98	0.89	0.99
PVQ	Dutch	0.16	0.17	0.19	0.18
	English	0.15	0.18	0.21	0.18
	Dutch: English	1.08	0.96	0.90	0.97
Kurtosis	Dutch	1.51	0.05	0.98	0.85
	English	1.97	0.31	1.48	1.25
	Dutch: English	0.77	0.16	0.66	0.68

tear these two factors apart. Therefore, future research shedding more light on the similarities and differences between Dutch and English speech overall is highly desirable.

The disfluency pattern that was observed consistently across all three speakers was a higher number and longer duration of filled pauses in the English lectures. Previous research has shown that L1 and L2 speakers generally show comparable pausing behaviour before analysis-of-speech units (AS-units; De Jong, 2016). AS-units are primarily syntactic units that consist of an independent clause (at least a clause including one finite verb). Pauses proceeding these units often reflect language-independent cognitive processes such as structuring or planning what to say next. In contrast, pauses that occur within AS-units are assumed to reflect language-specific processes such as how to formulate a particular message. These pauses have been shown to occur significantly more often and have longer durations in L2 speech than in L1 speech. A more detailed analysis of our data that differentiates between pauses before and within AS-units is beyond the scope of this paper, but the fact that the speakers in our study generally paused more often and longer in L2 English than in L1 Dutch seems to be in line with De Jong's (2016) overall results regarding pausing behaviour and might indicate that L1 Dutch-speaking lecturers run into slightly more problems formulating their messages in English than in Dutch, even if they are highly proficient in English.

The speakers' L1 Dutch lectures in our study were more varied with regard to vocabulary, as measured by Guiraud's index and the *D* statistics. The lower Guiraud's indices and *D* values in the English lectures indicate that these lecturers used a more limited set of words in L2 English than in L1 Dutch. This is in line with Tilstra and Smakman's (2018) study, who reported lower type-token ratios for English lectures given by L1 Dutch-speaking lecturers than for Dutch lectures given by the same speakers.

Finally, the assumed liveliness of the speech, as measured by PVQ (Hincks, 2010), and the assumed charisma of the speakers were explored by conducting cross-linguistic comparisons of the mean, median and kurtosis of the pitch (Niebuhr & Skarnitzl, 2019). The PVQ seems to suggest that the English lectures were slightly livelier for two of the speakers, while the third lecturer was somewhat livelier in Dutch. The analysis of kurtosis in the pitch data showed a more consistent pattern than the PVQ, with all speakers showing pitch patterns that have been shown to be associated with a higher degree of perceived charisma during their Dutch lecture than during their English lecture. The underlying assumption here is Niebuhr and Skarnitzl's (2019), who concluded that the more speakers are able to make use of his/her entire f0 range, the more charismatic they will be perceived: a high kurtosis signals that more data points are centered close to the mean, fewer data points deviate slightly from the mean, and more data points that deviate strongly from the mean, than is the case for lower kurtosis. Finally, all speakers used a 2% higher median pitch during their Dutch lecture than during their English lecture, which seems to point in a similar direction and indicates that speakers might be perceived as speaking slightly livelier in L1 Dutch than in L2 English (Hincks, 2005). While previous research has shown a correlation between suprasegmental features such as pitch on the one hand, and the speaker's perceived charisma and liveliness on the other, a separate study investigating whether this assumed perception can be supported empirically by our data has yet to be conducted.

7. Conclusion

In this study, L1 Dutch and L2 English lectures by three native speakers of Dutch were compared to determine whether the lecturers showed general speech differences depending on the language of instruction. All in all, the analyses in this exploratory study indicate that medium of instruction is likely to affect a lecturer's speech, but some speech features are more heavily and more consistently affected than others. More specifically, the canonical speech rate and the canonical articulation rate were consistently higher in L1 Dutch than in L2 English, with an average of 16% and 15%, respectively. The total duration of filled pauses was on average 10% lower in Dutch than in English, and for the average duration of filled pauses this difference was even 15%. The lexical diversity was on average 8% (Guiraud's index) and 20% (*D* statistic) higher in Dutch than in English, and the potential perceived charisma of the speakers was higher in the L1 Dutch lecture compared to the English one, as indicated by the lower kurtosis value of their pitch in Dutch than in English. For all remaining measures, our results yielded contradictory results.

While we assume the crossed-design of our data collection to be a strength, our dataset is based on only three speakers and therefore rather limited. Inter-individual differences and/or the setting of the lectures (introductory lecture versus a more research-oriented talk) could have yielded, boostered or deflated differences between L1 Dutch and L2 English lectures. Therefore, a larger dataset of lectures is needed for a robuster picture of lecturers' language use in EMI contexts.

Also, since all three of our lecturers had a background in communication studies and/or linguistics, it is likely that their language abilities or motivation to learn new languages was higher than that of their peers in other academic fields. Therefore, the results reported here might not be entirely representative of the population of Dutch-speaking lecturers in the Netherlands. In order to develop a clearer understanding of the role of EMI in Dutch tertiary education, future studies should include a higher number of lecturers from various disciplines with different levels of language proficiency and teaching experience.

Importantly, it is still an open question to what extent the reported linguistic differences affect students' experiences in class and their academic achievements overall. For instance, clear speech has been shown to manifest itself, among others, through a slower speaking tempo, more and longer pauses, and an increase in pitch range (Picheny, Durlach, & Braida, 1986; Picheny, Durlach, & Braida, 1989; Smiljanic & Bradlow, 2009). In addition, slow speech is generally more intelligible than fast speech, irrespective of whether the lower speech rate is accompanied by a clearer pronunciation or not (Schüppert, Hilton, & Gooskens, 2016). This suggests that the mere reduction of speech rate generally enhances speech comprehension, which in turn is likely to affect students' learning and their academic achievements in EMI settings in a positive way. Similarly, it is unclear whether the lecturers had a smaller vocabulary at their disposal in English than in Dutch, which would potentially result in a slightly poorer input for the students attending EMI lectures compared to DMI lectures, or whether the lecturers in our study aligned their presentation consciously or subconsciously to an anticipated audience of non-native English speakers. This alignment might manifest itself by keeping to the same terminology rather than rephrasing content using a more varied vocabulary. In the latter case, the seemingly reduced vocabulary used in the English lectures would not reflect the lecturers' less diverse vocabulary in English, but might in fact be a manifestation of the lecturers' awareness of students' linguistic needs in their L2. This assumption would be in line with the claims made by Giles, Coupland, and Coupland (1991) in their communication accommodation theory (CAT). The CAT postulates that speakers generally adapt their speech to one another in order to establish common ground, to express courtesy toward the interlocutor, and to increase mutual understanding. However, the theory is based on native-speaker dialogues in the first place and as Costa, Pickering, and Sorace (2008) point out, empirical studies investigating the cognitive processes underlying alignment in L1 speakers talking to L2 speakers are scarce. A recent study by Drljača Margić (2017) analysed self-reported data by 377 native speakers of various English varieties. Her study revealed that native speakers claim to be aligning to the anticipated language proficiency of non-native listeners by articulating more clearly, using fewer idioms and speaking more slowly than when they speak to native speakers. Therefore, although our findings are in line with findings on L1-L2 speech differences reported in earlier studies, the slower speech and articulation rates, the clearer pronunciation as well as the smaller vocabulary that our speakers showed in L2 lectures could also considered to be manifestations of conscious or subconscious alignment of the highly proficient L2 English-speaking lecturers to an anticipated audience of non-native English-speaking students. Further research into how lecturers align their speech to the student audience in the EMI classroom is needed to shed more light on this issue.

Likewise, investigating how engaging a lecture actually is and how accessible and/or competent a lecturer is perceived to be is a crucial question for the future of EMI research. Given that enthusiasm is one of the main driving forces behind students' intrinsic motivation (Patrick, Hisley, & Kempler, 2000), a closer investigation of the links between language of instruction and the perception of lecturers appears fundamental within EMI research. Finally, experimental studies on the link between typical characteristics of DMI and EMI lectures on the one hand and students' academic achievements on the other hand would yield data that HE institutions, policymakers and other stakeholders urgently need to formulate and safeguard a sustainable language policy for Dutch higher education.

Author statement

Anja Schüppert: Conceptualization, Methodology, Formal analysis, Writing - Original Draft, Review & Editing, Supervision, Project administration.

Penny Heisterkamp: Investigation, Data Curation, Writing - Review & Editing.

Acknowledgements

References

Archibald, J. (1998). *second language phonology*. Amsterdam: John Benjamins.

Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer [Computer program]. Version 6.0.46.* http://www.praat.org/.

Bulté, B., Housen, A., Pierrard, M., & Van Daele, S. (2008). *Investigating lexical proficiency development over time - the case of Dut*.

Chambers, F. (1997). What do we mean by fluency? *System, 25*(4), 535–544. https://doi.org/10.1016/S0346-251X(97)00046-8

Costa, A., Pickering, M. J., & Sorace, A. (2008). Alignment in second language dialogue. *Language & Cognitive Processes, 23*(4), 528–556. https://doi.org/10.1080/01690960801920545

De Jong, N. (2016). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching, 54*(2), 113–132. https://doi.org/10.1515/iral-2016-9993

De Jong, N., Groenhout, R., Schoonen, R., & Hulstijn, J. (2015). second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied PsychoLinguistics, 36*(2), 223–243. https://doi.org/10.1017/S0142716413000210

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2012). The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers. In A. Housen, I. Vedder, & F. Kuiken (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 121–142). Amsterdam/Philadelphia: John Benjamins.

De Jong, N., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods, 41*(2), 385–390. https://doi.org/10.3758/BRM.41.2.385

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition, 31*(4), 533–557. https://doi.org/10.1017/S0272263109990015

Dewaele, J.-M., & Pavlenko, A. (2003). Productivity and lexical diversity in native and non-native speech: A study of cross-cultural effects. In V. Cook (Ed.), *Effects of the second language on the first* (pp. 120–141). Clevedon/Buffalo/Toronto/Sydney: Multilingual Matters.

Drljača Margić, B. (2017). Communication courtesy or condescension? Linguistic accommodation of native to non-native speakers of English. *Journal of English As a Lingua Franca, 6*(1), 29–55. https://doi.org/10.1515/jelf-2017-0006

Foster, P., & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language Learning, 59*(4), 866–896. https://doi.org/10.1111/j.1467-9922.2009.00528.x

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge: University Press.

Haan, J., & Van Heuven, V. J. (1999). Male vs. female pitch range in Dutch questions. In , *14. Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1581–1584).

Hilton, Nanna H., Schüppert, Anja, & Gooskens, Charlotte (2011). Syllable reduction and articulation rates in Danish, Norwegian and Swedish. *Nordic Journal of Linguistics, 34*, 215–237. https://doi.org/10.1017/S0332586511000175

Hincks, R. (2005). Measures and perceptions of liveliness in student oral presentation speech: A proposal for an automatic feedback mechanism. *System, 33*(4), 575–591. https://doi.org/10.1016/j.system.2005.04.002

Hincks, R. (2010). Speaking rate and information content in English lingua franca oral presentations. *English for Specific Purposes, 29*(1), 4–18. https://doi.org/10.1016/j.esp.2009.05.004

Hyltenstam, K., Bartning, I., & Fant, L. (2018). *High-level language proficiency in second language and multilingual contexts*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316809686

Lennes, M. (2017). *SpeCT - speech Corpus Toolkit for Praat (v1.0.0). First release on GitHub. Zenodo.* https://doi.org/10.5281/zenodo.375923

Maiworm, F., & Wächter, B. (2002). *English-Language-Taught degree programmes in European higher education — trends and success factors*. Bonn: Lemmens.

Malvern, D., Richards, B., Chipere, N., & Duran, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Basingstoke: Palgrave Macmillan.

Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics, 19*(2), 313–330.

Meara, P. M., & Miralpeix, I. (2007). *D_Tools (Version 2.0; _lognostics: Tools for vocabulary researchers: Free software from _lognostics) [Computer software]*. University of Wales Swansea. http://www.lognostics.co.uk/tools/index.htm.

Niebuhr, O., & Skarnitzl, R. (2019). Measuring a speaker's acoustic correlates of pitch - but which? A contrastive analysis based on perceived speaker charisma. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th international congress of phonetic sciences, Melbourne, Australia 2019* (pp. 1174–1177). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Patrick, B. C., Hisley, J., & Kempler, T. (2000). "What's everybody so excited about?": The effects of teacher enthusiasm on student intrinsic motivation and vitality. *The Journal of Experimental Education, 68*(3), 217–236. https://doi.org/10.1080/00220970009600093

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. *Journal of Speech & Hearing Research, 29*(4), 434–446. https://doi.org/10.1044/jshr.2904.434

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1989). Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *Journal of Speech & Hearing Research, 32*(3), 600–603. https://doi.org/10.1044/jshr.3203.600

Pickering, L. (2004). The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse. *English for Specific Purposes, 23*(1), 19–43.

Raupach, M. (1980). Temporal variables in first and second language speech production. In H. W. Dechert, & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (pp. 263–270). The Hague/Paris/New York: Mouton.

Riazantseva, A. (2001). Second language proficiency and pausing. A Study of Russian Speakers of English. *Studies in Second Language Acquisition, 23*(4), 497–526. https://doi.org/10.1017/S027226310100403X

Roach, P., Hartman, J., Setter, J., & Jones, D. (2006). *Cambridge English pronouncing dictionary*. Cambridge: University Press.

Robb, M., Maclagan, M. A., & Chen, Y. (2004). Speaking rates of American and New Zealand varieties of English. *Clinical Linguistics and Phonetics, 18*(1), 1–15. https://doi.org/10.1080/0269920031000105336

Schmid, H. (1994). *Probabilistic part-of-speech tagging using decision trees. Proceedings of international conference on new methods in language processing. Manchester, UK*.

Schmid, H. (1995). *Improvements in part-of-speech tagging with an application to German. Proceedings of the ACL SIGDAT-workshop*. Dublin: Ireland.

Schüppert, Anja, Hilton, Nanna H., & Gooskens, Charlotte (2016). Why is Danish so difficult to understand for fellow Scandinavians? *Speech Communication, 79*, 47–60. https://doi.org/10.1016/j.specom.2016.02.001

Smiljanic, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass, 3*(1), 236–264. https://doi.org/10.1111/j.1749-818X.2008.00112.x

Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal, 65*(1), 71–79. https://doi.org/10.1093/elt/ccq020

Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning, 58*(2), 439–473. https://doi.org/10.1111/j.1467-9922.2008.00446.x

Thøgersen, J., & Airey, J. (2011). Lecturing undergraduate science in Danish and in English: A comparison of speaking rate and rhetorical style. *English for Specific Purposes, 30*(3), 209–221. https://doi.org/10.1016/j.esp.2011.01.002

Tilstra, K., & Smakman, D. (2018). The spoken academic English of Dutch university lecturers. *English Studies, 99*(5), 566–579. https://doi.org/10.1080/0013838X.2018.1483620

Traunmüller, H., & Eriksson, A. (1995). The perceptual evaluation of F0 excursions in speech as evidenced in liveliness estimations. *Journal of the Acoustical Society of America, 97*(3), 1905–1915. https://doi.org/10.1121/1.412942

Trouvain, J., Fauth, C., & Möbius, B. (2016, May 31 – June 3). *Breath and non-breath pauses in fluent and disfluent phases of German and French L1 and L2 read speech [Poster session]. Speech Prosody, Boston, United States.* https://doi.org/10.21437/SpeechProsody.2016-7

Trouvain, J., & Möbius, B. (2014, May 20–23). *Sources of variation of articulation rate in native and non-native speech: Comparisons of French and German [Poster session]. Speech Prosody, Dublin, Ireland.* https://doi.org/10.21437/SpeechProsody.2014-42

Tsao, Y.-C., Weismer, G., & Iqbal, K. (2006). Interspeaker variation in habitual speaking rate: Additional evidence. *Journal of Speech, Language, and Hearing Research, 49*(5), 1156–1164. https://doi.org/10.1044/1092-4388(2006/083)

Van Hout, R., & Vermeer, A. (2007). Comparing measures of lexical richness. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 93–115). Cambridge: Cambridge UP.

Vinke, A. A. (1995). *English as the medium of instruction in Dutch Engineering Education.* Doctoral thesis. Delft: Department of Communication and Education, Delft University of Technology.

Vinke, A. A., Snippe, J., & Jochems, W. (1998). English-medium content courses in non-English higher education: A study of lecturer experiences and teaching behaviours. *Teaching in Higher Education, 3*(3), 383–394. https://doi.org/10.1080/1356215980030307

Wächter, B., & Maiworm, F. (2008). *English-taught programmes in European higher education. Bonn: Lemmens.*

Wächter, B., & Maiworm, F. (2014). *English-taught programmes in European higher education. The state of play in 2014.* Bonn: Lemmens.

Westfall, P. H. (2014). Kurtosis as peakedness, 1905–2014. R.I.P. *The American Statistician, 68*(3), 191–195. https://doi.org/10.1080/00031305.2014.917055

Wilkinson, R. (2013). English-Medium instruction at a Dutch university: Challenges and pitfalls. In A. Doiz, D. Lasagabaster, & J. M. Sierra (Eds.), *English-medium instruction at Universities: Global challenges* (pp. 3–24). Bristol: Multilingual Matters.

Zimmerer, F., Jügler, J., Andreeva, B., Möbius, B., & Trouvain, J. (2014). Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and German speakers. In N. Campbell, D. Gibbon, & D. Hirst (Eds.), *Social and linguistic speech prosody* (pp. 1037–1041). Dublin http://www.coli.uni-saarland.de/~trouvain/Zimmerer_et_al_2014.pdf.

Anja Schüppert is Assistant Professor of European Linguistics at the University of Groningen. Her research is in the field of multilingualism, with a particular focus on speech perception, language attitudes, and the use of English as medium of instruction.

Penny Heisterkamp is studying for a Master's degree in Language and Cognition at the University of Groningen. Her research interests include second language development, applied linguistics and sociolinguistics.