Dutch listeners' responses to Dutch, British and American English accents in three contexts

Warda Nejjari[1], Marinel Gerritsen[1], Roeland van Hout[1] and Brigitte Planken[1]

[1]*Radboud University*

Abstract   The aim of this study was to assess Dutch listeners' responses to native-accented Englishes compared with Dutch-accented English in terms of speech understandability and speaker evaluations in three professional communication contexts. In a matched-guise experiment Dutch listeners (N = 392) responded to a Dutch English, a standard British and American accent in terms of speech understandability (intelligibility, comprehensibility, interpretability) and speaker evaluations (status, affect, dynamism). Dutch listeners evaluated these accents in three communication contexts: Lecture, Audio Tour, Job Pitch. Only context affected speech understandability: comprehensibility and interpretability were higher for the Lecture compared to the Audio Tour and the Job Pitch. Accent only negatively affected status evaluations for Dutch-accented English. Context only evoked more affect in the Audio Tour and the Lecture than in the Job Pitch. Our main conclusion is that Dutch-accented English negatively impacts status, but not understanding, affect and dynamism. Context impacts understanding and affect.

Corresponding author
Warda Nejjari, warda.nejjari@ru.nl

1 Introduction

An increasing range of professional contexts in the Netherlands require that Dutch people communicate in English, for example, in English taught bachelor and master degree programmes and in multinational corporations (Bouma, 2016; Edwards, 2016; Gerritsen et al., 2016; Huygen, 2017; Lizzini et al., 2017; Nickerson, 2005; Van Heest, 2018). Generally, as L2 (non-native) English speakers, the Dutch are considered to possess high English proficiency (cf. the high Dutch ranking on the English Proficiency Index, EF (2018)). Nevertheless, discussions in the Netherlands on the Dutch using English, for

example in a higher education context, are often about the perceived English fluency levels of lecturers and their accents (e.g. Bouma, 2016; Bronkhorst, 2015). Dutch students have indicated that the Dutch-accented English used by professors has hindered their understanding and resulted in less effective lectures (e.g. Huygen, 2017). Such perceptions were partly confirmed in an experiment on self-reported speech understanding of Dutch instructors' accents in English (Hendriks et al., 2016). As self-reported speech perceptions do not necessarily reflect actual speech understanding, our study investigated actual understanding of Dutch-accented English (versus British and American English) by Dutch listeners. Furthermore, it studied the speaker evaluations Dutch-accented English elicits, in various communication contexts. L2 accents can also affect the perceived quality and effectiveness of information transfer (see Nejjari et al., 2012 for an overview). Information transfer involves a continuous evaluation of speech and speakers in interactions in terms of, among other things, the linguistic features of an utterance and the context in which these are used (Lev-Ari & Keysar, 2010). In addition, accents are one of the first signals used to assess someone as similar, intelligent, friendly or even untrustworthy (e.g. Nejjari et al., 2012; Purnell et al., 1999). Research on L2 accents in a lecturing setting shows that L2-accented speech can negatively impact listeners. For instance, lecturers' L2-accented speech can lead to lower perceptions of status, but does not affect a lecturer's likeability (affect) (e.g. Dalton-Puffer et al., 1997; Hendriks et al., 2016). Furthermore, it irritates students as well as negatively impacts perceived lecture understanding (e.g. Bolton & Kuteeva, 2012; Hellekjær, 2010; Hendriks et al., 2016).

Kachru and Smith (2008) offer a fine-grained definition of speech understanding consisting of three components. The first component is intelligibility, which refers to the manner in which utterances are deciphered into words and sentence-level elements. Intelligibility can be measured by requiring listeners to orthographically transcribe individual words or sentences produced by speakers in recorded settings (see also Nejjari et al., 2012; Nelson, 2011; Yorkston et al., 1996). The second component is comprehensibility, which is the manner in which individual words are understood and how words put together in sentences express meaning within a specific context (see Nejjari et al., 2012). Kachru and Smith (2008) indicate that unlike intelligibility, comprehensibility requires that a listener understands the syntax, semantics and even the physical context in which an utterance is heard. The third component is interpretability, which is difficult to distinguish from comprehensibility, because both deal with meaning beyond the recognition of words and phrases. Interpretability refers to whether listeners are able to grasp a speaker's intentions and have the cultural baggage to understand and interpret discourse strategies (see Nejjari et al., 2012).

Nejjari et al. (2012) operationalized Kachru and Smith's model to investigate standard British English speakers' reactions to a standard British English accent and a strong or light Dutch English accent. Both Dutch English accents led to lower *intel-*

ligibility and *comprehensibility*, but did not impact *interpretability*, suggesting that L1 listeners are capable of understanding the intentions of a speaker with an L2 accent without having heard each word correctly or completely comprehending message content. These results demonstrate that measuring speech understandability as a multi-component construct can yield nuanced insights into listeners' understanding of L2 accents.

As discussed above, L2 accents do not only necessarily affect actual speech understanding, but can affect speaker evaluations. In accentedness research, speaker evaluations tend to be measured in terms of status and affect (e.g. Hendriks et al., 2016; Nejjari et al., 2012; Zahn & Hopper, 1985). A number of studies have shown that L1 and L2 speakers of English generally ascribe L2 English accents lower status compared to L1 English accents, even if presented with speech samples in their own L2 accent in English (e.g. Cargile, 1997; Dalton-Puffer et al., 1997; He & Zhang, 2010; Lindemann, 2003; Matsuura et al., 1994; McKenzie, 2008; Nejjari et al., 2012; Ryan & Bulik, 1982). However, L2 accents do not necessarily elicit lower affect (Nejjari et al., 2012; Hendriks et al., 2016).

In addition to status and affect, accentedness research has investigated speaker dynamism, albeit less consistently. According to Zahn and Hopper (1985), dynamism refers to a person's activity level and enthusiasm. It taps into a listener's perception of the self-presentation of a speaker, and this differs from perceptions of a speaker's status or affect, as confirmed by Grondelaers and Van Hout (2015). Their study investigated the perceived prestige of the use of a non-standard object pronoun in Dutch (Dutch: 'hun' (them) in subject position instead of 'zij' (they)). Even though non-standard use of the object pronoun evoked lower status, it did not affect speaker dynamism, suggesting that, while a speaker who uses non-standard features may be perceived as less cultured, their self-presentation will not be perceived differently from that of a standard language user.

Since the aim of speaker evaluation research is to understand listeners' responses to specifically selected accents, it is important to determine whether listeners are able to identify the accents they are asked to evaluate. Research by Yook and Lindemann (2013) has shown that accent recognition triggers speaker evaluations associated with a particular group of people (e.g. nationality, culture, sub-culture), and therefore, speaker evaluations are connected to stereotypical ideas listeners have of certain groups. Consequently, to study evaluations of specific speaker groups (e.g. Dutch-accented English speakers) adequately, we need to determine whether listeners can correctly identify speakers' language background.

Most accentedness studies focus on one (professional) communication context, such as higher education (Bolton & Kuteeva, 2012; Hellekjær, 2010; Hendriks et al., 2016) or business sales (Nejjari et al., 2012; Tsalikis et al., 1991). One exception is Cargile (1997) who investigated L1 and L2 listeners' perceptions of the suitability of Mandarin-accented English in a job interview compared with a higher education classroom. He found that listeners viewed the Mandarin-accented English accent as acceptable in a job interview, but not in a higher education classroom. The difference in acceptability could

be connected to the Expectancy theory of Burgoon and Burgoon (2001): people have expectations of, among other things, verbal and non-verbal communications that are 'expected' and/or 'desired' in a certain context. By extension, a particular accent in a certain context may violate what is an expected or desired accent in that context, which may result in negative evaluations, and perhaps even impact understanding (cf. Cargile, 1997; Dalton-Puffer et al., 1997). It is important to note that expectations in terms of context are connected to topic and content (how and what is stated). For example, listeners can have different communication expectations of a lecture on Corona virus in terms of content and topic compared with a lecture on Dutch poetry. Although we acknowledge that a context may be linked to different topics and contents, for practical reasons, only context was included as a variable, with a topic and content typical of each context.

A matched-guise experiment was set up (see 2.1) to investigate listeners' actual understanding of Dutch-accented English compared to standard British and American English accents and listeners' perceptions of the speakers of these accents across three communication contexts. We were also curious whether the three components of speech understandability correlate with the three speaker evaluation dimensions we investigated, as this has not been studied systematically. Three research questions were formulated:

RQ1:   Do accent (standard British English = BrE, standard American English = AmE, Dutch English = DE) and context (Lecture, Audio tour, Job pitch) affect Dutch listeners' speech understandability (intelligibility, comprehensibility, interpretability)?

RQ2:   Do accent (BrE, AmE, DE) and context (Lecture, Audio Tour, Job Pitch) affect Dutch listeners' speaker evaluations (status, affect, dynamism)?

RQ3:   Are speech understandability and speaker evaluations correlated?

2 Method

2.1 Speakers: matched-guises, controls, and filler

A matched-guise technique was used, which meant that one male speaker produced all three tested accents. This technique has the advantage of controlling for the influence of paralinguistic features like voice quality. We discussed different voice evaluation paradigms in Nejjari et al. (2019) leading us to conclude that the matched-guise technique constitutes the most optimal speech evaluation paradigm. The male matched-guise speaker in our study had been assessed in a prior accentedness experiment, which showed that only he could authentically produce the three accents: (1) BrE, (2) AmE, and (3) DE, the typical English accent of L1 speakers of Dutch (see Nejjari et al., 2019 for detailed discussion on the matched-guise technique itself, the reason for employ-

ing the matched-guise technique, and the matched-guise speaker selection process). In the current study, "standard" accents of British and American English refer to the accents generally associated with the national accent norm of the nations in which they originate. Both Englishes are used as models in L2 English education around the world (Robinson, 2019). A typical Dutch English accent contains features that L1 speakers of Dutch and others familiar with Dutch and Dutch English will recognize as such. For example, because Dutch lacks dental consonants [ð] as in *this*, *mother*, *breathe* and [θ] as in *think*, *Martha*, *breath*, they are often mispronounced as stop consonants, [d] and [t] respectively, by Dutch speakers of English. Dutch also lacks voiced fricatives and plosives in the coda, causing the voiced obstruents of English to generally be pronounced as voiceless in Dutch speakers' English (e.g. *live*, *badge*, *bad*, *bag* will be pronounced with [f, tʃ, t, k]) (Gussenhoven & Broeders, 1997). As no standard has been defined for Dutch English, this accent is referred to as 'typical' in the present study (cf. Nejjari et al., 2019).

Listeners only heard one of the nine matched-guise speech samples, preventing listeners from deducing that we used a matched-guise technique. We included (speech samples from) six male control speakers as stimuli: two L1 speakers of BrE, two L1 speakers of AmE, two DE speakers. All but one had been assessed on the representativeness of their accents in the same accentedness experiment as the matched-guise speaker (see Nejjari et al., 2019). The speaker who was not assessed was one of the DE control speakers. He was generally regarded by experienced linguists and Dutch language specialists as a representative speaker of DE. One further male speaker produced a speech sample that was presented to all listeners at the beginning of the experiment (the filler speech sample) to familiarize them with the task. The filler speaker had also been selected from the prior study as an L1 and standard speaker of British English (Nejjari et al., 2019).

2.2 Instrumentation and participants

To ensure that the nine (plus one filler) matched-guise samples could be evaluated in each context for each accent, to avoid repeating the content of each context, and to limit any order effect, 18 listener groups were created, aiming to collect approximately 30 listeners per group (Table 1).

Listeners were highly educated (highest degree attained: 12.3 % A-level; 51.2 % bachelor; 27.6 % master; 2.3 % PhD; 6.3 % other) native speakers of Dutch (mean age 39; 60 % female, 40 % male) with no background in linguistics. They were selected because they represent the part of Dutch society most likely to use English in educational and professional settings (Bouma, 2016; Lizzini et al., 2017). To understand whether self-reported English fluency impacted speech understandability, the listeners were asked to estimate their English language skills (listening, reading, speaking, writing) on a 5-point Likert scale (1: very low; 2: low; 3: average; 4: high; 5: like a

Table 1 Listener groups

N	Version 1		Version 2
1	Filler	10	Filler
n=30	MG BrE Lecture	n=30	MG AmE Lecture
	NS1 AmE Audio Tour		NS2 AmE Audio Tour
	NS1 BrE Job Pitch		NS2 BrE Job Pitch
2	Filler	11	Filler
n=30	NS1 DE Lecture	n=30	NS2 DE Lecture
	MG BrE Audio Tour		MG AmE Audio Tour
	NS1 AmE Job Pitch		NS2 AmE Job Pitch
3	Filler	12	Filler
n=30	NS1 AmE Lecture	n=30	NS2 AmE Lecture
	NS1 DE Audio Tour		NS2 DE Audio Tour
	MG BrE Job Pitch		MG AmE Job Pitch
4	Filler	13	Filler
n=30	MG AmE Lecture	n=30	MG DE Lecture
	NS2 AmE Audio Tour		NS1 AmE Audio Tour
	NS1 DE Job Pitch		NS2 DE Job Pitch
5	Filler	14	Filler
n=30	NS1 BrE Lecture	n=30	NS2 BrE Lecture
	MG AmE Audio Tour		MG DE Audio Tour
	NS2 AmE Job Pitch		NS1 AmE Job Pitch
6	Filler	15	Filler
n=30	NS2 AmE Lecture	n=30	NS1 AmE Lecture
	NS1 BrE Audio Tour		NS2 BrE Audio Tour
	MG AmE Job Pitch		MG DE Job Pitch
7	Filler	16	Filler
n=30	MG DE Lecture	n=30	MG BrE Lecture
	NS2 DE Audio Tour		NS1 DE Audio Tour
	NS2 BrE Job Pitch		NS1 BrE Job Pitch
8	Filler	17	Filler
n=30	NS2 BrE Lecture	n=30	NS1 BrE Lecture
	MG DE Audio Tour		MG BrE Audio Tour
	NS2 DE Job Pitch		NS1 DE Job Pitch
9	Filler	18	Filler
n=30	NS2 DE Lecture	n=30	NS1 DE Lecture
	NS2 BrE Audio Tour		NS1 BrE Audio Tour
	MG DE Job Pitch		MG BrE Job Pitch

N = 540; Filler = filler speaker; MG = matched-guise speaker; NS1 or NS 2= native speaker 1 or 2; BrE = standard British English; AmE = standard American English; DE = Dutch-accented English.

native speaker). Ninety-two percent indicated that their skills in all four areas were average or higher.

Participants were asked to indicate what they believed the speakers' country of origin was. Over 92% correctly indicated that the DE matched-guise speech sample was by a speaker from the Netherlands. For the BrE speech sample 88% correctly indicated that the speech sample was produced by a speaker from Great Britain, and over 81% correctly indicated that the AmE matched-guise speech sample was by a speaker from the United States. Responses of listeners who were not able to correctly identify speakers' country of origin were excluded, as these could reflect interfering associations with other speaker groups, and thus with other accents than those under study.

In order to answer the first two RQs, the responses by listeners who had correctly identified the speaker's country of origin were included in the analyses (N = 392). To answer RQ3, however, correlations were calculated using all responses (N = 545), so as to provide a better assessment of the relationship between the speech understandability dimensions and English language skills, based on the larger range of English skills in the full sample.

2.3 Stimuli

Three texts, (1) an introduction to a marketing lecture, (2) an art gallery audio tour segment and (3) a job pitch for a retail management position, were used as the basis for the speech samples. One filler text on a general topic was used to start each questionnaire and allow listeners to get accustomed to the questions (links and texts: see appendix).

All but the filler text reflect three contexts (independent variable *context*) in which English is an important lingua franca: higher education, tourism, and international business (Gerritsen et al., 2016). Every listener evaluated all three contexts in the same order. To ensure realistic topics and content, the lecture, audio tour, and filler texts were selected from an IELTS Academic English listening test and the job pitch text from a human resources webpage. The matched-guise speaker produced the three accents in all three contexts, resulting in nine speech samples. The six control speakers produced English in their L1 accents (DE, BrE, AmE) in the three contexts, resulting in 18 speech samples, and the additional speaker produced one filler speech sample on a general topic in BrE.

2.3.1 *Speech understandability*
Following Kachru and Smith (2008) and Nejjari et al. (2012), three questions measured speech understandability in terms of ability to (1) literally recognize words (intelligibility), (2) understand the meaning of the words within the context (comprehensibility), and (3) understand the intention of the speaker / purpose of the message (interpretability). To measure intelligibility, listeners were presented with a speech sample consisting of

the first 11–12 words of each of the four stimuli and asked to write down what was literally stated. We counted the number of words correctly transcribed, and an Intraclass Correlation Coefficient for the two raters turned out to be extremely high (.98). To measure comprehensibility, listeners indicated whether one statement per speech sample on the topic of that speech sample was correct or not. Interpretability was measured by having listeners indicate whether one statement per speech sample was correct or not with respect to the speaker's intention.

2.3.2 *Speaker evaluations*

To assess speaker evaluations, the listeners indicated on 5-point Likert scales (1 = strongly disagree; 5 = strongly agree; 3 = neither disagree nor agree) to which extent they believed the speaker possessed 11 personality traits, representing the three dimensions of speaker evaluations this experiment focuses on, namely status (competent, educated, having authority, intelligent and cultured), affect (considerate, pleasant and friendly), and dynamism (energetic, enthusiastic, confident). The traits associated with status and affect are based on Nejjari et al. (2012). Status represents the degree to which a speaker is viewed as intelligent and well-educated, and affect represents the degree to which a speaker is perceived as likeable. Dynamism measured the self-presentation of the speaker and was based on Grondelaers et al. (2015). To confirm the dimensionality of the speaker evaluation items, a principal component analysis, was applied with an Eigenvalue > 1 criterion for factor extraction, and varimax rotation. The personality items showed a resolution into three factors: status, affect, dynamism, as can be seen in Table 2.

2.4 Procedures

An online survey was conducted via a student Facebook page and the online data collection service Qualtrics. This approach enabled the participation of many listeners, but the disadvantage was that many listeners did not complete the questionnaire or produced repetitive or nonsense response patterns. Approximately 41 % of the original data was discarded, resulting in data from 545 listeners. Reasons for excluding data included: incomplete questionnaires (approximately 25 %); nonsense answers or symbols and /or only neutral answers (approximately 15 % of the original data). Listeners who did not speak the required L1 were also excluded (1 %). The median questionnaire duration was a little over 16 minutes for all listeners, including the excluded responses.

2.5 Statistics

Descriptives and frequencies were calculated to establish means and percentages of listener characteristics and responses. We used ANOVAs when the dependent variable

Table 2 Rotated Factor Matrix: factor loadings of the scores on 11 scales with three factors. Only loadings >.550 have been printed.

	Factor 1 Status	Factor 2 Affect	Factor 3 Dynamism
Competent	.690		
Considerate		.805	
Cultured	.855		
Highly educated	.871		
Pleasant		.751	
Energetic			.766
Authoritative	.596		
Friendly		.868	
Enthusiastic			.670
Intelligent	.855		
Confident			.671

was continuous and logistic regression when it was binomial (two values only). The factor analyses we applied to trace underlying dimensions in speaker evaluations were principal component analyses, with varimax rotation.

3 Results

The findings will be reported for the three RQs successively (datafiles are available upon request). All mean and frequency measurements of speech understandability and speaker evaluation with regard to the three accents in three contexts can be found in Table 3 below. Only the matched-guise speaker results are reported here, and not the control speakers, since individual speaker voice quality differences can obscure the impact of accent and context, which is precisely the reason why the matched-guise technique was used.

3.1 Speech understandability, accent, context (RQ1)

In general, speech understandability was high: comprehensibility was 89.2 %, interpretability was 82.2 %. The mean number of correctly transcribed words (intelligibility) was 9.23 (81.0 %), which is fairly high given the maximum correct score of 11 (Lecture, Job Pitch) or 12 (Audio Tour) words.

Table 3 Mean, % speech understandability; means speaker evaluations, per accent (BrE, AmE, DE; 1 = negative; 3 = neutral; 5 = positive; 95 % confidence interval) and context (Lecture, Audio Tour, Job Pitch)

Accent, Context	*Speech understandability*			*Speaker evaluations*		
	Intelligibility Mean (SD)	Comprehensibility % correct	Interpretability % correct	Status Mean (SD)	Affect Mean (SD)	Dynamism Mean (SD)
Dutch English Lecture*	9.02(2.69) n = 63	95.8% n = 68	88.6% n = 62	2.84(.70) n = 63	3.28(.75) n = 62	3.43(.76) n = 63
Dutch English Audio Tour*	9.41(2.64) n = 34	82.9% n = 29	82.9% n = 29	2.69(.70) n = 39	3.10(.67) n = 32	3.36(.81) n = 34
Dutch English Job Pitch**	9.26(2.79) n = 62	78.8% n = 52	71.2% n = 47	2.76(.78) n = 62	2.80(.75) n = 61	2.72(.76) n = 62
British English Lecture	9.49(2.16) n = 55	100% n = 69	94.2% n = 65	3.99(.57) n = 55	3.34(.78) n = 55	3.30(.60) n = 55
British English Audio Tour	7.97(3.24) n = 29	82.5% n = 52	84.1% n = 53	3.85(.64) n = 29	3.40(.74) n = 29	3.49(.71) n = 29
British English Job Pitch	8.84(2.79) n = 43	86.8% n = 59	67.6% n = 46	3.74(.58) n = 43	2.68(.80) n = 43	3.14(.74) n = 43
American English Lecture	9.91(1.63) n = 47	100% n = 61	93.4% n = 57	3.74(.55) n = 46	2.94(.68) n = 47	2.96(.72) n = 47
American English Audio Tour	10.56(2.18) n = 18	80.3% n = 53	86.4% n = 57	3.53(.36) n = 18	3.33(.54) n = 18	3.39(.57) n = 18
American English Job Pitch	8.95(2.61) n = 42	82.1% n = 55	70.1% n = 47	3.71(.65) n = 42	2.67(.84) n = 42	3.17(.76) n = 42

* = max. 11 words; ** = max. 12 words intelligible. N = 392; n = number of listeners per accent and context

Results for intelligibility (proportion correct words) in relation to context and accent are shown in Figure 1, with 95 % confidence intervals. They show substantial overlap in the confidence intervals between bars, indicating that there are no strong differences between the accents and the contexts. An analysis of variance was applied to investigate the effects of context and accent and their interaction. There was an interaction effect ($F(4,383) = 2.84$, $p = .02$, $\eta_p^2 = .03$). There was a main effect for accent ($F(2,383) = 4.30$, $p = .01$, $\eta_p^2 = .02$), but not for context ($F(2,383) = 1.27$, $p = .28$, $\eta_p^2 = .01$). However, the post-hoc test (HSD) for accent did not show significant differences.

Results for comprehensibility in relation to context and accent are shown in Figure 2, with 95 % confidence intervals. There are no error bars for the Lecture in British and American English because the comprehension scores were 100 %. The remaining error
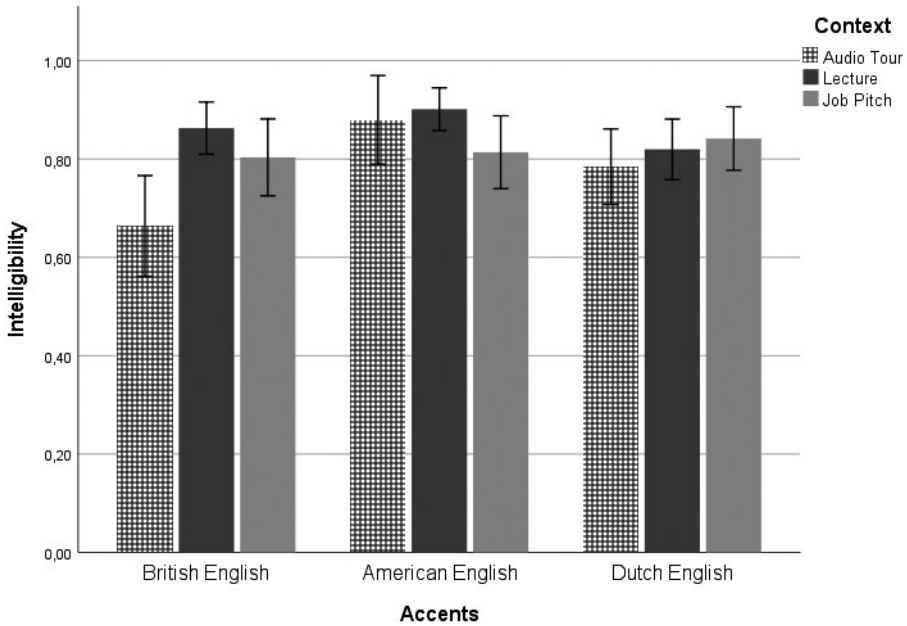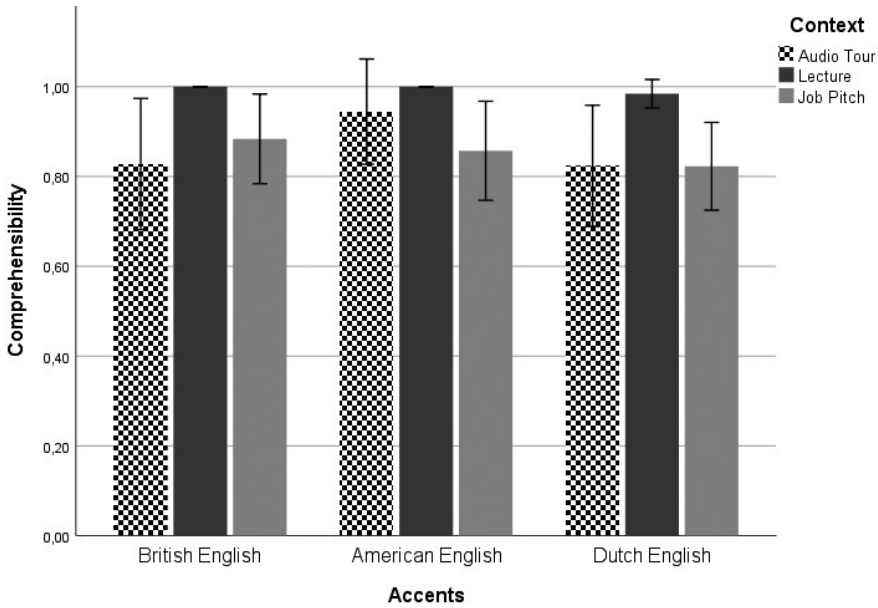
Figure 1 Mean proportions of correct intelligibility for accent (BrE, AmE, DE) and context (Audio Tour (12 words), Lecture (11 words), Job Pitch (11 words)). Error bars 95%

bars show substantial overlap in the confidence intervals between bars, indicating that there are no strong differences between the three accents and the three contexts. A logistic regression (comprehensibility is a yes/no variable) was applied to test the effects of context and accent as well as their interaction. The interaction effect could be removed (Deviance score = 5.58, df = 4, p = .23). The same applied to accent in the next step (Deviance score = 1.80, df = 2, p = .41), but not for the remaining factor, context (Wald = 17.77, df = 2, p < .01). Post-hoc analysis (Bonferroni) revealed that the Lecture was comprehended better than the Audio Tour and Job Pitch.

Results for interpretability in relation to context and accent are shown in Figure 3, with 95% confidence intervals. They demonstrate substantial overlap in the confidence intervals between bars, indicating that there are no strong differences between the three accents and three contexts. A logistic regression (interpretability is a yes/no variable) was applied to test the effects of context and accent as well as their interaction. The interaction effect could be removed (Deviance score between the model with the interaction and the model without interaction was 3.00, df = 4, p = .56). The same applied to accent in the next step (Deviance score = .42, df = 2, p > .05), but not for the remaining factor, context (Wald = 33.85, df = 2, p < .01). Post-hoc analysis (Bonferroni) revealed that interpretability was higher for the Lecture than the Audio Tour and Job Pitch.

Interpretability showed a positive, but weak, correlation with comprehensibility (r(545) = .26, p < .01) and intelligibility (r(545) = .20, p < .01). Comprehensibility showed a

Figure 2 Mean proportions of positive comprehensibility for accent (BrE, AmE, DE) and context (Audio Tour, Lecture, Job Pitch). Error bars 95%
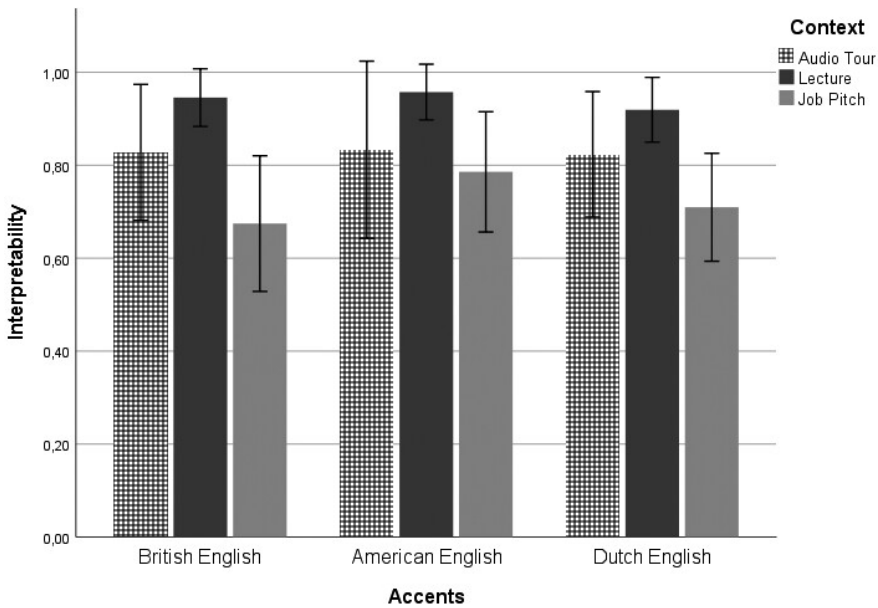
Figure 3 Mean proportions of correct interpretability for accent (BrE, AmE, DE) and context (Audio Tour, Lecture, Job Pitch). Error bars 95%
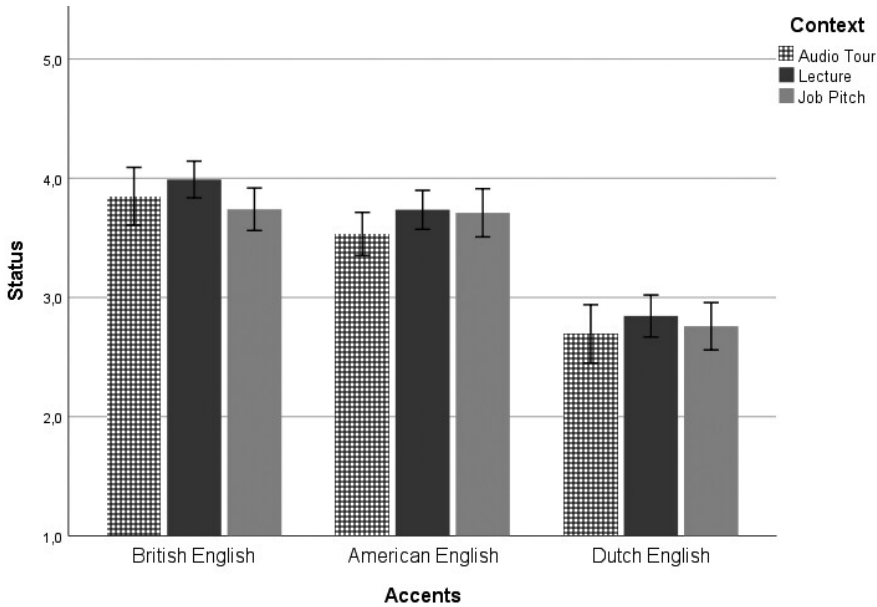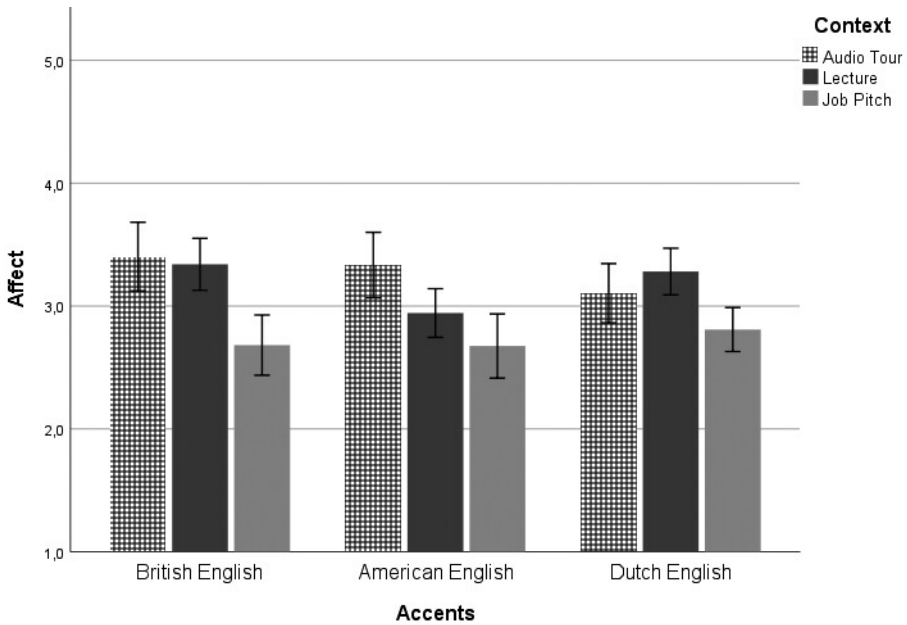
Figure 4 Mean scores of status (1 = negative; 3 = neutral; 5 = positive) for accent (BrE, AmE, DE) and context (Audio Tour, Lecture, Job Pitch). Error bars 95 %

weak, positive correlation with intelligibility ($r(545)$ = .13, p < .01). To check whether their self-reported language skills had an effect on speech understandability, correlations were computed with the three understandability variables. None were significant: interpretability: $r(544)$ = .07, p = .11; comprehensibility: $r(544)$ = −.02, p = .70; intelligibility: $r(545)$ = .15, p = .06.

3.2 Speaker evaluations, accent, context (RQ2)

Results for status in relation to context and accent are shown in Figure 4, with 95 % confidence intervals. An analysis of variance was applied to investigate the effects of context and accent and their interaction. There was no interaction effect ($F(4,383)$ = .53, p = .72, η_p^2 = .01). There was a main effect for accent ($F(2,383)$ = 106.86, p < .01, η_p^2 = .36), but not for context ($F(2,383)$ = 2.15, p = .12, η_p^2 = .01). Post-hoc comparisons (HSD) showed higher status for British English (M = 3.87, SD = .59) and American English (M = 3.69, SD = .57) compared to Dutch English (M = 2.77, SD = .73, p < .001).

Results for affect in relation to context and accent are shown in Figure 5, with 95 % confidence intervals. An analysis of variance was applied to investigate the effects of context and accent and their interaction. There was no interaction effect ($F(4, 381)$ = 1.90, p = 11, η_p^2 = .02). There was no main effect for accent ($F(2, 381)$ = 1.07, p = .31, η_p^2 = .01), but there was for context ($F(2, 381)$ = 20.29, p < .01, η_p^2 = .10). Post-
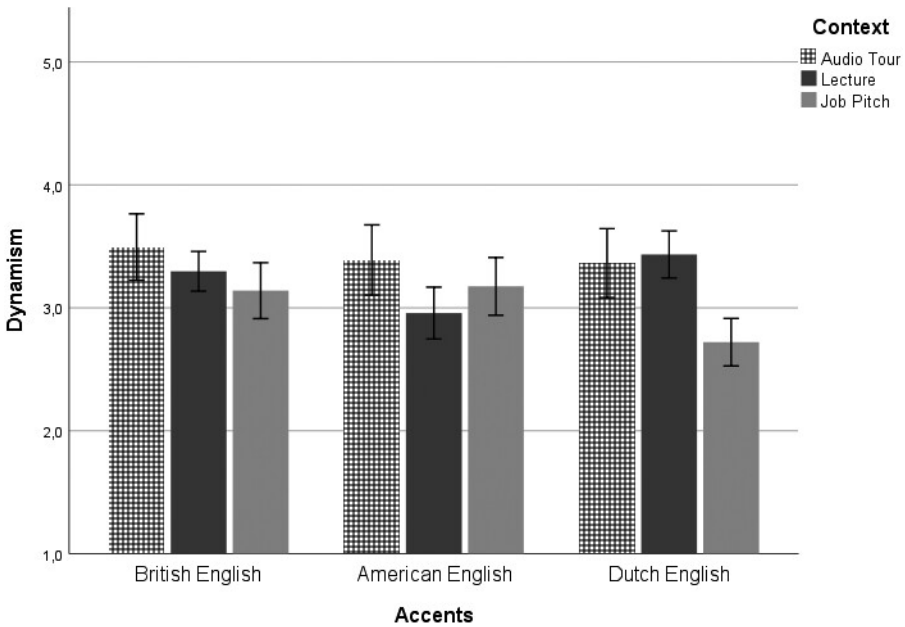
Figure 5 Mean scores of affect (1 = negative; 3 = neutral; 5 = positive) for accent (BrE, AmE, DE) and context (Audio Tour, Lecture, Job Pitch). Error bars 95%

hoc comparisons (HSD) showed that the Audio Tour ($M = 3.27$, $SD = .67$) and Lecture ($M = 3.22$, $SD = .74$) aroused higher affect than the Job Pitch ($M = 2.73$, $SD = .77$, $p < .001$).

The results for dynamism are shown in Figure 6, with 95% confidence intervals. An analysis of variance was applied to investigate the effects of context and accent and their interaction. There was an interaction effect ($F(4,384) = 75.59$, $p < .01$, $\eta_p{}^2 = .06$). There was no main effect for accent ($F(2,384) = 1.42$, $p = .24$, $\eta_p{}^2 = .01$), but there was for context ($F(2,384) = 8.19$, $p < .01$, $\eta_p{}^2 = .04$). Post-hoc comparisons (HSD) showed that Job Pitch ($M = 2.97$, $SD = .78$) evoked lower dynamism than the Audio Tour ($M = 3.42$, $SD = .72$) and Lecture ($M = 3.26$, $SD = .71$, $p < .01$), but the interaction clarified that this was due to the negative evaluation of the Dutch English Job Pitch.

3.3 Correlation speech understandability and speaker evaluations (RQ3)

Table 4 gives the correlations between speech understandability and speaker evaluations. There was only one negative correlation between comprehensibility and affect, but with a low value ($r = -.11$, $p = .03$).

Figure 6 Mean scores of dynamism (1 = negative, 3 = neutral, 5 = positive) for accent (BrE, AmE, DE) and context (Audio Tour, Lecture, Job Pitch). Error bars 95 %

Table 4 Correlations between speech understandability and speaker evaluations

	Interpretability	Comprehensibility	Intelligibility
Status	.08 (p=.10)	.01 (p=.99)	.03 (p=.59)
Affect	.05 (p=.33)	−.11 (p=.03)*	.04 (p=.44)
Dynamism	.02 (p=.68)	.07 (p=.17)	.09 (p=.09)

*significant < .05

4 Conclusion and discussion

The aim of our experiment was to study whether accent (standard BrE, standard AmE, Dutch-accented English) and context (Lecture, Audio tour, Job pitch) affect Dutch listeners' speech understandability and speaker evaluations, and whether speech understandability and speaker evaluations are correlated.

4.1 Speech understandability, accent, context (RQ1)

The effects of accent and context on speech understandability (RQ1) are only significant for context. Accent did not influence speech understandability. This indicates that Dutch-accented English is as comprehensible as standard British and American English.

Context had an effect on interpretability and comprehensibility. The Lecture was more interpretable and comprehensible than the Audio Tour and the Job Pitch regardless of accent. Most listeners were highly educated, and might have had more frequent experience with academic educational settings. This concurs with Cargile (1997), who found that if listeners were confronted with an accent in a unexpected context, their responses to that speaker were negatively affected.

Speech understandability was high for all accents and contexts. The experiment was conducted with highly educated Dutch listeners who are generally regarded as having good English language skills (EF, 2018). This could mean that our listeners had high enough English fluency to easily deal with the content in our stimuli.

The high speech understandability caused a ceiling effect, which could have obscured the observed interpretability effects, and also contributed to weak, positive correlations between the three dimensions. This means that, when listeners could decipher words and phrases (intelligibility), this positively influenced their ability to understand them (comprehensibility) and to interpret speaker's intention (interpretability). Comprehensibility and interpretability were measured by asking listeners to confirm or deny the validity of a statement on the content (comprehensibility) and communicative purpose (interpretability) of each sample. This method might have positively biased results; however, the positive correlation between the three components suggests that higher intelligibility helps comprehensibility and interpretability.

4.2 Speaker evaluations, accent, context (RQ2)

With regard to the effects of accent and context on speaker evaluations (RQ2), it can be concluded that having a Dutch English accent has a negative effect on a speaker's status compared with standard British and American English. Context has no effect on the perceived status of a speaker. This indicates that accent matters more than context in perceptions of speaker status.

In terms of perceptions of affect and dynamism, context appears to be more important than accent, because context had an effect on affect and dynamism, with the Job Pitch context evoking lower affect and dynamism compared to the Lecture and Audio Tour contexts. Yet, even though the means for all three contexts were mostly neutral, and despite the fact that for dynamism and context the statistically significant results were caused by the Dutch English job pitch, it is striking that the Job Pitch context evoked such a negative response.

The above results concur with previous research that found that L1 and L2 speakers of English tend to assign L2 English accents lower status compared to L1 English accents, including their own L2 accent in English (e.g. Nejjari et al., 2012; Cargile, 1997; Dalton-Puffer et al., 1997; He & Zhang, 2010; Hendriks et al., 2016; Lindemann, 2003; Matsuura et al., 1994; McKenzie, 2008; Ryan & Bulik, 1982). In line with other research, our results showed that an L2 accent does not necessarily evoke lower affect (Nejjari et al., 2012; Hendriks et al., 2016) or dynamism (Nejjari et al., 2020), and that speaker evaluations can be impacted by context (Cargile, 1997).

4.3 Correlations speech understandability and speaker evaluations (RQ3)

The last research question focused on correlations between speech understandability and speaker evaluations (RQ3). Status and dynamism were not correlated with speech understandability. This means that a listener's evaluation of the status and dynamism of a speaker is not related to their ability to comprehend a speaker. For affect there was a weak, negative correlation with comprehensibility. These results contradict Nejjari et al. (2012), who found a positive correlation between status and intelligibility and comprehensibility as well as between affect and comprehensibility. The explanation may be that their 2012 study involved British listeners, half of whom were not familiar with Dutch accented-English, who reacted to standard British English accent and two degrees of Dutch-accented English. These listeners probably needed to put more effort into comprehending accents and content than the L1 speakers of Dutch in the current study, who were all familiar with all three accents. They achieved high levels of *speech understandability*, resulting in the ceiling effect discussed above.

4.4 Limitations and future research

This study has a number of limitations. First, we only had one, male matched-guise speaker, because it was extremely challenging to find a speaker who could produce representative native as well as non-native English accents. However, future matched-guise studies could benefit from more diverse matched guise speakers. Second, the contexts for the questionnaire versions were presented in one order (filler, Lecture, Audio Tour, Job Pitch) without counterbalancing. We did not apply counterbalancing as this would have resulted in more questionnaire versions which would have required more respondents. We recommend counterbalancing for future studies. Third, the listeners were assumed to represent people who would likely be familiar with the three selected contexts. With hindsight, selecting professionals who regularly interact in these contexts, such as HR managers in the job interview context, would have been better.

The results for context provide nuance to earlier L1 and L2 accentedness studies. Future research should investigate other contexts and L2 and L1 accents to help language learners understand the effects of their accents and provide insights into which

accents are deemed desirable for which contexts. The assessment of speech understand-ability in terms of Kachru and Smith's (2008) three components yielded results that provide useful insights into the levels at which listeners are able to understand speech. However, the specific method employed to measure comprehensibility and interpretabil-ity resulted in high scores, ceiling effects, and weak correlations between the speech understandability dimensions. This suggests that the questions might have positively impacted the results. Therefore, this method should be further validated with listener groups with various English language levels and language backgrounds. While the use of the matched-guise technique contributed to the validity of our results, since they cannot be attributed to voice characteristics of individual speakers (as might have been the case with verbal guises), future studies should aim to use more than one matched-guise speaker.

Although our methodology yielded insightful results in terms of the understanding and evaluations of speakers, future research should assess how L2 and L1 English accents affect L1 and L2 English speakers' behavior in specific contexts (see e.g. Purnell et al., 1999). With the increased globalization of academia and the job market, it is relevant to investigate how people view one another and behave towards each other on the basis of accents in these particular communication contexts.

4.5 Implications

Accent training aimed at becoming as L1 as possible for advanced Dutch learners of English need not be emphasized in language teaching when the aim is to be understand-able and to evoke higher affect and dynamism. However, accent training to sound like an L1 speaker can be beneficial if the aim of language teaching is to evoke perceptions of high status. As context was shown to affect attitudes towards speakers, creating awareness of such potential effects can help learners understand the impression they make in English in different contexts.

References

Bolton, K., & Kuteeva, M. (2012). English as an academic language at a Swedish university: Parallel language use and the 'threat' of English. *Journal of Multilingual and Multicultural Development*, 33(5), 429–447. DOI: https://doi.org/10.1080/01434632.2012.670241

Bouma, K. (2016, 26 August). Meer dan de helft van de studies volledig in het Engels. *De Volkskrant*. https://www.volkskrant.nl/binnenland/meer-dan-de-helft-van-de-studies-volledig-in-het-engels~a4364526/

Bronkhorst, X. (2015). *DUB-panel: Engels als onderwijstaal? Meer voor- dan nadelen*. University Utrecht, 6 May 2015. https://www.dub.uu.nl/nl/achtergrond/dub-panel-engels-als-onderwijstaal-meer-voor-dan-nadelen

Burgoon, J.K., & Burgoon, M. (2001). Expectancy theories. In W.P. Robinson, & H. Giles (Eds.), *The new handbook of language and social psychology* (2nd ed., pp. 79–102). Wiley.

Cargile, A. (1997). Attitudes toward Chinese-accented speech: An investigation in two contexts. *Journal of Language and Social Psychology*, 16, 434–443. DOI: https://doi.org/10.1177/0261927X970164004

Dalton-Puffer, C., Kaltenböck, G., & Smit, U. (1997). Learner attitudes and L2 pronunciation in Austria. *World Englishes*, 16(1), 115–128. DOI: https://doi.org/10.1111/1467-971X.00052

Edwards, A. (2016). *English in the Netherlands: Functions, forms and attitudes*. John Benjamins. DOI: https://doi.org/10.1075/veaw.g56

EF (2018). EF English Proficiency Index. *EF*, 2018. https://www.ef.com/wwen/epi/

Gerritsen, M., Van Meurs, F., Planken, B., & Korzilius, H. (2016). A reconsideration of the status of English in the Netherlands within the Kachruvin Three Circles model. *World Englishes*, 35(3), 457–475. DOI: https://doi.org/10.1111/weng.12206

Gussenhoven, C., & Broeders, T. (1997). *English pronunciation for student teachers*. Wolters-Noordhof.

Grondelaers, S., & Van Hout, R. (2015). How (in)coherent can standard languages be? A perceptual perspective on co-variation. *Lingua*, 172, 62–71. DOI: https://doi.org/10.1016/j.lingua.2015.10.012

He, D., & Zhang, Q. (2010). Native speaker norms and China English: From the perspective of learners and teachers in China. *TESOL Quarterly*, 44(4), 769–789. DOI: https://doi.org/10.5054/tq.2010.235995

Hellekjær, G.O. (2010). Lecture comprehension in English-medium higher education. *Hermes-Journal of Language and Communication Studies*, 23(45), 11–34. DOI: https://doi.org/10.7146/hjlcb.v23i45.97343

Hendriks, B., Van Meurs, F., & Hogervorst, N. (2016). Effects of degree of accentedness in lecturers' Dutch-English pronunciation on Dutch students' attitudes and perceptions of comprehensibility. *Dutch Journal of Applied Linguistics*, 5(1), 1–17. DOI: https://doi.org/10.1075/dujal.5.1.01hen

Huygen, M. (2017, 21 October). Uhm ... how do you say that in English? *NRC Weekend*. https://www.nrc.nl/nieuws/2017/10/21/uhm-how-do-you-say-that-in-english-13593132-a1578131

Kachru, Y., & Smith, L. (2008). *Cultures, contexts, and world Englishes*. Routledge. DOI: 10.1111/j.1467–971X.2008.01574_3.x

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, 46(6), 1093–1096. DOI: https://doi.org/10.1016/j.jesp.2010.05.025

Lindemann, S. (2003). Koreans, Chinese, or Indians? Attitudes and ideologies about nonnative English speakers in the United States. *Journal of Sociolinguistics*, 7(3), 348–364. DOI: https://doi.org/10.1111/1467-9481.00228

Lizzini, O., Martijn, M., Munk, R., & De Regt, H. (2017, 3 July). Academisch onderwijs kan niet zonder het Engels. *De Volkskrant*. https://www.volkskrant.nl/opinie/academisch-onderwijs-kan-niet-zonder-het-engels~a4503932/

Matsuura, H., Chiba, R., & Yamamoto, A. (1994). Japanese college students attitudes towards non-

native varieties of English. In D. Graddol, & J. Swann (Eds.), *Evaluating Language* (pp. 52–61). BAAL/Multilingual Matters.

McKenzie, R.M. (2008). The role of variety recognition in Japanese university students' attitudes towards English speech varieties. *Journal of Multilingual and Multicultural Development*, 29(2), 139–153. DOI: 10.2167/jmmd565.0

Nejjari, W., Gerritsen, M., Van der Haagen. M., & Korzilius, H. (2012). Responses to Dutch-accented English. *World Englishes*, 31(2), 248–268. DOI: https://doi.org/10.1111/j.1467-971X.2012.01754.x

Nejjari, W., Gerritsen, M., Van Hout, R., & Planken, B. (2019). Refinement of the matched-guise technique for the study of the effect of non-native accents compared to native accents. *Lingua*, 219, 90–105. DOI: https://doi.org/10.1016/j.lingua.2018.12.001

Nejjari, W., Gerritsen, M., Van Hout, R., Planken, B. (2020). Where does a 'foreign' accent matter? German, Spanish and Singaporean listeners' reactions to Dutch-accented English, and standard British and American English accents. *PLoS ONE* 15(4): e0231089. DOI: https://doi.org/10.1371/journal.pone.0231089

Nelson, C.L. (2011). *Intelligibility in world Englishes: Theory and application*. Routledge. DOI: https://doi.org/10.4324/9780203832578

Nickerson, C. (2005). English as a lingua franca in international business contexts. *English for Specific Purposes*, 24(4), 367–380. DOI: https://doi.org/10.1016/j.esp.2005.02.001

Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology*, 18(1), 10–30. DOI: 10.1177/0261927X99018001002

Robinson, J. (2019). Minority Ethnic English. *British library*, 24 Apr 2019. http://www.bl.uk/learning/langlit/sounds/case-studies/minority-ethnic/

Ryan, E.B., & Bulik, C.M. (1982). Evaluations of middle class and lower class speakers of standard American and German-accented English. *Journal of Language and Social Psychology*, 1(1), 51–61. DOI: 10.1177/0261927X8200100104

Tsalikis, J., deShields Jr, O.W., & La Tour, M.S. (1991). The role of accent on the credibility and effectiveness of the salesperson. *Journal of Personal Selling and Sales Management*, 11(19), 31–41. DOI: 10.1080/08853134.1991.10753857

Van Heest, H. (2018). Minister overweegt de wet te wijzigen over verengelsing. *ScienceGuide*, 4 April 2018. https://www.scienceguide.nl/2018/04/minister-overweegt-de-wet-te-wijzigen-over-verengelsing/

Yook, C., & Lindemann, S. (2013). The role of speaker identification in Korean university students' attitudes towards five varieties of English. *Journal of Multilingual and Multicultural Development* 34 (3), 279–296. DOI: 10.1080/01434632.2012.734509

Yorkston, K., Strand, E., & Kennedy, M. (1996). Comprehensibility of dysarthric speech: Implications for assessment and treatment planning. *American Journal of Speech-Language Pathology*, 5(1), 55–66. DOI: 10.1044/1058–0360.0501.55

Zahn, C.J., & Hopper, R. (1985). Measuring language attitudes: The speech evaluation instrument. *Journal of Language and Social Psychology*, 4(2), 113–123. DOI: 10.1177/0261927X8500400203

Appendix: Speech sample links: matched guises, filler, controls

The speech sample links below redirect to one speech sample per accent (standard British, standard American English, Dutch English) and communication context (lecture, audio tour, job pitch), per speaker (matched-guise speaker; filler speaker; control speakers). For the controls, three speech sample links are available as examples (for each accent one).

Matched-guise speaker Lecture
standard British English   https://cls.ru.nl/webexp-media/HV_BE_L.html
standard American English   https://cls.ru.nl/webexp-media/HV_AE_L.html
Dutch English   https://cls.ru.nl/webexp-media/HV_DE_L.html

Lecture speech sample text:
Last week we looked at some general principles associated with marketing and today I'd like to look at some of those points in a little more detail. So what is marketing? Or to put it another way, what does the term marketing mean? Many people think of it simply as the process of selling and advertising. And this is hardly surprising when every day we are bombarded with telephone sales. But selling and advertising are only two functions of marketing. In fact, marketing, more than any other business function, deals with customers. So perhaps the simplest definition is this: marketing is the delivery of customer value and satisfaction at a profit. In other words, finding customers, keeping those customers happy and making money out of the process.

Matched-guise speaker Audio tour
standard British English   https://cls.ru.nl/webexp-media/HV_BE_A.html
standard American English   https://cls.ru.nl/webexp-media/HV_AE_A.html
Dutch English   https://cls.ru.nl/webexp-media/HV_DE_A.html

Audio tour speech sample text:
Welcome to the City Art Gallery and to our gallery audio tour. The gallery was first established in 1875 and is now one of the city's most popular attractions. The building has undergone a number of changes over the years, the most recent addition being the extension to the ground floor, which was opened in 1988. The Gallery houses some of the finest works of art in Australia, including Aboriginal, European and Asian paintings. There is a comprehensive Australian collection, which includes works from the early colonial period to the present day. In addition to the paintings on display, we have an excellent collection of photography, with photographs dating from the nineteenth century. Now let's begin our tour in the nineteenth century Australian room.

Matched-guise speaker Job pitch
standard British English   https://cls.ru.nl/webexp-media/HV_BE_J.html
standard American English   https://cls.ru.nl/webexp-media/HV_AE_J.html
Dutch English   https://cls.ru.nl/webexp-media/HV_DE_J.html

Job Pitch speech sample text:
I'm a seasoned Retail Manager strong in developing training programs and loss prevention techniques that have resulted in revenue savings of over 2.3 Million during the past 11 years. I would say that my greatest strength is my ability to follow through. I have found that I am most successful when I pay attention to every piece of the product cycle, from the first contact, to the thank you at the completion of the project. My greatest weakness is my tendency to overthink a situation. I sometimes take too much time to strategize on a sale, and find in the end that my initial plan was the one that was the best.

Filler speaker
standard British English   https://cls.ru.nl/webexp-media/CG_BE_F.html

Examples of controls or native speakers for each accent (in the lecture context only)
standard British English   https://cls.ru.nl/webexp-media/JC_BE_L.html
standard American English   https://cls.ru.nl/webexp-media/PG_AE_L.html
Dutch English   https://cls.ru.nl/webexp-media/CJ_DE_L.html