“Hama”? Reduced pronunciations in non-native natural speech obstruct high-school students’ comprehension at lower processing levels

Wanrooij, K.; Raijmakers, M.E.J.

DOI
10.1016/j.wocn.2021.101082

Publication date
2021

Document Version
Final published version

Published in
Journal of Phonetics

License
CC BY

Journal of Phonetics 88 (2021) 101082

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

"*Hama*"? Reduced pronunciations in non-native natural speech obstruct high-school students' comprehension at lower processing levels

Karin Wanrooij [a,*,1], Maartje E.J. Raijmakers [b]

[a] Educational Sciences, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands
[b] Educational Studies, Free University, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 8 April 2020
Received in revised form 5 June 2021
Accepted 17 June 2021
Available online 28 August 2021

Keywords:
Reduced speech
Non-native listening comprehension
High school
Adolescents
German as a second language
Bottom-up and top-down processing
Acoustic–phonetic cues

ABSTRACT

Native speakers 'reduce' their pronunciations, i.e., they shorten and merge words. For instance, German native speakers may say "hama" for "haben wir" ('have-we'). We examined to what extent such reductions are problematic for adolescent learners of a second language, after four years of high-school training; and whether the problems can be related to inadequate bottom-up and top-down processing. For this, 39 Dutch and 38 German adolescents heard either reduced or unreduced German full phrases and part-phrases (phrase-intelligibility task) and words (lexical decision task). The results show that (1) Learners perceive non-native reduced speech less accurately than unreduced speech and also judge it as less intelligible; (2) This reduced-form disadvantage occurs separately from factors such as speech rate, orthography and voice; (3) The disadvantage for non-native listeners is substantial and larger than that in native listeners. Therefore, it probably reflects a lack of experience with reduced (i.e., real-life) speech; and (4) Non-native reductions induce at least inadequate bottom-up processing in learners, and may make top-down processing less accessible. We interpret the findings as supporting the idea that experience with variants (here: reduced variants) is necessary to strengthen linguistic (word) representations.

1. Introduction

Understanding non-native natural speech is notoriously difficult. One possible but little discussed cause is native speakers' tendency to 'reduce' their pronunciations, i.e., to delete and assimilate speech sounds within and across words. Such reductions occur naturally, in all languages (e.g., German: Kohler, 1998; Dutch: Ernestus, 2000; English: Shockey, 2003, Johnson, 2004; Japanese: Maekawa & Kikuchi, 2005; review in Ernestus & Warner, 2011). Two examples in German, the target language in this article are *hama* [hamɐ] for the word combination *haben wir* [haːbən viːɐ] 'we have' (literally: 'have we') and *haspmmomentsait* [haspmmomɛntsaɪt] for the sentence *Hast du einen Moment Zeit* [hast^{h} duː ʔaɪnən moːmɛnt^{h} tsaɪt^{h}] 'Do you have a moment?' (Kohler, 1998). As visible in these examples, words in reduced speech do not appear in the forms that are familiar from textbooks for learners of a non-native language (henceforth 'L2 learners'; L2 is the common abbreviation for 'second language'; we use it as a reference to any non-native language). In this research, we ask to what extent L2 reductions affect listening comprehension in adolescent L2 learners in high school, and what mechanisms (in terms of bottom-up and top-down processing) may underlie an observed effect. As will be discussed below (Introduction 1.2), reduced-speech research has not paid much attention to L2 learners yet, and, to our knowledge, adolescent L2 learners have not been examined at all, although they represent a prominent learner group in many countries. In our case, the adolescents were Dutch 17-year-old learners of German, who completed four years of high-school training and were about to take their final school exam in listening comprehension. Below, we will first address what is known about reduced-speech perception in native and non-native languages (Introduction 1.1 and 1.2), before elaborating on the contributions of the present study (Introduction 1.3).

* Corresponding author.
E-mail addresses: karin.wanrooij@hotmail.com (K. Wanrooij), m.e.j.raijmakers@vu.nl (M.E.J. Raijmakers).
[1] Present address: Utrecht Institute of Linguistics OTS, Utrecht University, Trans 10, 3512 JK Utrecht, The Netherlands.

https://doi.org/10.1016/j.wocn.2021.101082
0095-4470/

1.1. Native (L1) reduced-speech perception

When listening to our native language (henceforth “L1”, which stands for “first language”), we generally do not notice reductions. We automatically recognize the various reduction patterns and fill in the missing parts. Therefore, it may seem surprising that even though reduced speech is the more natural speaking style, L1 listeners have more problems with reduced than unreduced speech when they are tested in experiments (Ernestus, Baayen, & Schreuder, 2002; Ranbom & Connine, 2007; Tucker & Warner, 2007; Wanrooij & Raijmakers, 2020). However, this can be understood when considering that acoustic–phonetic cues are partly missing and highly variable in reduced speech. For instance, the phrase *haspmmomentsait* [haspmmomɛntsaɪt] for *Hast du einen Moment Zeit* (the example above) can be reduced further to *hasm̰momentsait* [hasm̰momɛntsaɪt] (Kohler, 1998). In the former pronunciation, the cues for the [d] in the word *du* are reduced to the plosive feature of [p]; in the latter pronunciation, only a glottalization (in the [m]) remains. And these are not the only variants possible. In fact, there is a continuum of possibilities, running from ‘not reduced’ to ‘extremely reduced’. In *haben wir* (the other example above) increasingly reduced examples of *haben* include *habn* [haːbn], *habm* [haːbm], *hamm* [haːmm] and *ham* [haːm] (Kohler, 2001). Hence, to perceive a reduced speech fragment, listeners must be able to pick up and use whatever local (acoustic–phonetic) cues remain and combine those with cues present in the speech surrounding the particular fragment. Adult listeners have been shown to use contextual cues of several types, e.g., acoustic–phonetic cues (Janse & Ernestus, 2011), phonological cues (Mitterer, 2011), syntactic cues (Ernestus et al., 2002) and semantic cues (van de Ven, Tucker, & Ernestus, 2011). Of course, when this helping ‘context’ surrounding a reduced speech fragment is limited, it becomes more difficult for listeners to compensate for missing cues. Notice that the more context is removed, the more important the remaining acoustic–phonetic cues and the ability to pick up these cues become. (In *hasm̰momentsait*, for example, it will help listeners if they can pick up the glottalization, and this cue becomes more important when more context is removed, e.g., *hasm̰* [hasm̰]).

In a previous study that is part of the same project as the present study (Wanrooij & Raijmakers, 2020), we examined reduced-speech perception in 16-year-old adolescents and compared their performance to that in slightly older adults (mean age of 24 years). The adolescents were also less competent in perceiving reduced than unreduced L1 speech, just as the adults, but the more remarkable finding was that they were less competent in reduced-speech perception than the adults. This suggests that the skills needed for reduced-speech perception are still developing in adolescence and that experience with reduced variants plays a role in developing these skills. Interestingly, the difference with adults emerged particularly when the reduced fragments were short (mostly confined to single words) and skilled acoustic–phonetic cue use was thus highly beneficial for adequate perception. We interpreted the results as supporting McMurray, Danelz, Rigler, and Seedorff's (2018) hypothesis that what develops with experience is the sensitivity to acoustic–phonetic detail (making speech sound and word representations richer) and a flexibility to handle cue variability better (thereby making the representations more robust). We will get back to what this means for the present study below (Introduction 1.3).

1.2. Non-native (L2) reduced-speech perception

In light of listeners' relative difficulty with L1 reduced as compared to unreduced speech, listeners can be expected to also experience problems with L2 reduced speech. Yet, there is little work on these problems and on how they compare to L1 reduced-speech perception. That reduced speech is difficult for L2 listeners has been reported for a long time (e.g., Bloomfield, 1933; Bowen, 1975; Norris, 1995; Field, 2003; Hulstijn, 2003), but to our knowledge none of the previous studies looked at younger L2 listeners than adults and only few examined the effect of reductions on L2 speech perception through a systematic comparison of reduced- and unreduced-speech perception (as we do in this study), thus leaving open the possibility that other factors account for the results. What do we know from these previous studies based on adult L2 listeners? Henrichsen (1984), the first to study the topic systematically, used a dictation task, in which listeners have to write down what they hear. Listeners' perception was substantially worse for reduced than unreduced L2 speech (in his case English). Similar results were obtained by Ahn (1987; also with English as the L2), Ito (2006; again with English as the L2) and Ernestus, Dikmans, and Giezenaar (2017; Dutch as the L2). Higher-proficiency L2 listeners may score better at perceiving reductions than lower-proficiency L2 listeners (Ahn, 1987; Ito, 2006), but their scores for reduced-speech perception still seem to remain low (Ahn, 1987; Ito, 2006; Ernestus et al., 2017), even when they are motivated to improve listening comprehension (Ahn, 1987). Hence, problems with L2 reduced-speech perception seem tenacious for adult L2 listeners.

Morano, Ernestus, and Ten Bosch (2015) tried to train L2 listeners in the lab. They trained Dutch learners of French with unreduced (with schwa) and reduced (without schwa) versions of new French words and then tested their knowledge in a subsequent auditory lexical decision task (in this task, listeners are presented with words and non-words and must decide as quickly as possible whether an item is a word or not; Holley-Wilcox, 1977; McCusker, Holley-Wilcox & Hillinger, 1979). The learners recognized both versions, but did not generalize the knowledge to well-known, non-trained words. It is not clear to what extent L2 learners can learn to generalize such just-trained reduction patterns. In view of the above, our knowledge of L2 reduced-speech perception is still modest: overall, adult L2 learners experience problems with the perception of L2 reduced speech, and they seem to have limited room for improvement. It is not clear if perhaps younger L2 learners succeed better. (We will get back at a possible advantage for younger learners in Section 1.3.2).

1.3. The present study

The present study contributes to previous work in three ways. First, it removes possible nuisance factors present in previous systematic studies on reduced-speech perception (such as speech rate and orthography; see below), so that

Table 1
Overview of the three main research questions, with the sub-questions, tasks and variables for addressing the questions. PI = phrase-intelligibility task, ALDf= auditory lexical decision task. (Explanation: Introduction 1.3. Outcomes: Table 11).

Questions and sub-questions	Task	Dependent variables
1. Do reductions (separated from factors such as speech rate) hamper perception, and consequently comprehension, in adolescent L2 learners?		
a. Do adolescent L2 learners perceive reduced speech less accurately than unreduced speech?	PI + ALD	Accuracies
b. Do they consciously judge reduced speech as less intelligible than unreduced speech?	PI	Intelligibility judgments
c. Do they process reduced speech slower than unreduced speech?	ALD	Reaction times
If so:		
2. To what extent? Is the 'reduced-form disadvantage' relevant?		
a. What is the effect size?	PI + ALD	See Question 1
b. Is the disadvantage larger than in adolescent L1 listeners?		
3. What processing mechanisms may underlie the disadvantage?		
a. Is bottom-up processing inadequate?		
– Is perception worse for reduced than unreduced speech specifically for part-phrases, i.e., when surrounding phrasal context is removed and listeners are forced to use bottom-up processing most?	PI	Accuracies in part-phrases
– Are non-word substitutions (and other error types) more numerous for reduced than unreduced speech?	PI	Number of Non-word substitutions (and other error types)
b. Is top-down processing inadequate?		
– Does perception improve less for reduced than unreduced speech when phrasal context is provided and possibilities for top-down processing increase?	PI	Accuracies in full phrases after part-phrases

the results reflect true effects of reductions on L2 speech perception better. Second, it examines the effect of reductions in a younger learner group than adults (L2 learners in high school, in our case Dutch learners), and for another L2 language than previously studied (German). Studying high-school students is relevant, because they represent the largest group of L2 learners in the Netherlands and in many other countries and adults are not necessarily good representatives of adolescents when speech perception is concerned, not even in the L1 (Wanrooij & Raijmakers, 2020). German has also not been studied in the literature on L2 reduced-speech perception, where the L2 is almost always English (Introduction 1.2). For the present study, a concomitant advantage of German as the L2 is that Dutch adolescents are not exposed to German as extensively as to English. This allows us to largely avoid the influence on reduced-speech perception of exposure outside the classroom, and hence to study L2 reduced-speech perception as a proper result of high-school training better. Thirdly, the present study dives into the mechanisms behind adolescent L2 reduced-speech perception, specifically into possible obstructions in bottom-up and top-down processing (explained below). Specifically, three main research questions are addressed (Table 1). They are explained below (Introduction 1.3.1, 1.3.2 and 1.3.3 respectively).

1.3.1. Replication for adolescent L2 listeners, while avoiding nuisance factors

First, we simply ask whether reductions hamper perception, and consequently comprehension, in adolescent L2 listeners at all (research question 1). This question may seem superfluous: they most probably do, since a 'reduced-form disadvantage' has been observed in adult L1 and L2 listeners and in adolescent L1 listeners (Introduction 1.1 and 1.2). However, the lack of previous evidence made it necessary to first replicate the disadvantage in this learner group too. Additionally, a replication was necessary to settle that reductions cause a separate disadvantage from possible nuisance factors that could have accounted for the previous results, notably speech rate. The duration of unreduced speech is normally longer than that of the corresponding reduced versions (e.g., Morano et al., 2015; Brand & Ernestus, 2018). Hence, unreduced speech gives listeners not only more cues, but also more time to comprehend the signal. Additionally, processes determining the speech rate are separable from those determining the degree of reduction (Hasegawa, 1979, 2006). To exclude speech rate as a factor, it is therefore necessary to make sure that presented unreduced speech fragments are not longer (or shorter) than reduced speech fragments (Method 2.2.1).

Another possible nuisance factor is orthography. Some previous studies provided orthographic support pertaining to the stimuli or the context of the stimuli (e.g., Ahn, 1987; Ernestus et al., 2017). Phrases in these studies were relatively long so that the support was useful in constraining the burden on memory. However, orthography reflects unreduced speech more closely than reduced speech and is known to affect native and non-native speech perception (e.g., Rastle, McCormick, Bayliss, & Davis, 2011; Escudero & Wanrooij, 2010). To exclude orthography as a factor, it is therefore necessary to avoid orthographic support and to confine the burden on memory by using simple short phrases (Ito, 2006; Method 2.2.1). Still another possible nuisance factor is voice. Some voices are more intelligible than other ones. To exclude voice as a factor, it is necessary to avoid multiple voices (this is also done in previous research, e.g., Ernestus et al., 2017; Method 2.2.1).

The present study features two tasks, a phrase-intelligibility task and an auditory lexical decision task, and both tasks were used to address this first research question (Table 1). In the phrase-intelligibility task, participants were presented with stretches of speech and typed what they heard, without time pressure. The task is also called 'dictation task' in previous work on reduced-speech perception (Henrichsen, 1984; Brown & Hilferty, 1986; Norris, 1995; Wong et al., 2015; ten Bosch, Giezenaar, Boves, & Ernestus, 2016; Ernestus et al., 2017). In an auditory lexical decision task, a series of words and non-words (or pseudo-words) is presented and participants must decide as quickly and as accurately as possible whether an item is a word or not (Holley-Wilcox, 1977; McCusker et al., 1979). Auditory lexical decision tasks have also been used before to study reduced-speech perception (Poelmans, 2003; Connine, Ranbom, & Patterson, 2008; Tucker, 2011; Pitt, Dilley, & Tat, 2011; Morano et al., 2015; Ernestus & Cutler, 2015; Wanrooij & Raijmakers, 2020). We address the first research question more concretely (see Table 1) by asking (a) whether adolescent L2 listeners perceive reduced speech less accurately than unreduced speech (for this, we measured accuracies in the phrase-intelligibility task and the auditory lexical decision task), (b) whether they consciously judge reduced speech as less intelligible than unreduced speech (intelligibility judgments in the phrase-intelligibility task); and (c) whether they process reduced speech slower than unreduced speech (reaction times in the auditory lexical decision task).

1.3.2. Relevance and the role of experience

The second and third research questions were asked after establishing a reduced-form disadvantage in the adolescent L2 learners. The second question is: to what extent do reductions hamper adolescent L2 listeners' perception (Table 1)? This question was asked to determine how relevant the reduced-form disadvantage is, in theory and practice. For this, we calculated the effect sizes of the relevant variables and compared the reduced-form disadvantage to that observed in adolescent L1 listeners. A similar-sized disadvantage as in adolescent L1 listeners might reflect the relative paucity of available cues in reduced speech (recall Introduction 1.1) and hence a *general* (i.e., not specific to adolescent L2 listeners) larger listening challenge compared to unreduced speech. This outcome would imply that adolescent L2 learners are at an advantage in comparison with adult L2 learners (Introduction 1.2) and hence that age of learning, which is known to influence L2 speech acquisition (e.g., Flege, MacKay, & Meador, 1999) has an effect. A *larger* disadvantage than for adolescent L1 listeners, on the other hand, would signal that experience plays a role too, not just experience with the L2 in general, but experience with reduced variants in particular. It is precisely such experience that the L2 listeners in high school are largely lacking now. (As in other countries, L2 learners in the Netherlands are mostly trained with written materials and hence with neatly spelled out unreduced words. According to a questionnaire, which was conducted among 82 teachers of German in high schools across the Netherlands, training materials often lack reduced speech; Wanrooij, 2019). We predict that experience with reduced variants will play a role and hence that the disadvantage will be larger than in L1 listeners. An important reason for this is that the results of the previous study in our project suggest that such experience is still even shaping reduced-speech perception in adolescent L1 listeners (Wanrooij & Raijmakers, 2020; Introduction 1.1).

Theoretically, a larger disadvantage for L2 than L1 listeners would endorse the idea proposed in previous literature on L1 and L2 acquisition, that not only clarity in experienced speech (here: the unreduced variant) supports linguistic category creation, but also variability (here: the reduced variants). For L1 acquisition, the variability of acoustic properties of speech sounds in infant-directed speech has been hypothesized to support the formation of speech sound categories, alongside the exaggerated acoustic properties that make the speech sounds clearer (Kuhl et al., 1997). For L2 acquisition, high-variability phonetic training has been reported to strengthen L2 learners' speech sound categories (Logan, Lively, & Pisoni, 1991; Wanrooij, Escudero, & Raijmakers, 2013). Similarly, experience with reduced word variants may add to the building of speech sound and word representations (McMurray et al., 2018; Wanrooij & Raijmakers, 2020).

Notice that we do not expect that adolescent L2 listeners' perception of *unreduced* speech will be much worse than that of L1 listeners in the experiments in this paper. This is because the speech fragments consist of simple short phrases and words that should be (highly) familiar (including phrases such as *schönen guten Tag*, a greeting; recall that the adolescent L2 listeners were close to taking their final listening comprehension exam after completing four years of high-school training in German). Hence, the L2 listeners' vocabulary knowledge[2] should be sufficient for a relatively adequate perception of the unreduced speech in this paper. Of course, their scores for unreduced-speech perception may still be lower than that of L1 listeners; this would not contradict an experience account (L2 listeners also have less experience with the unreduced pronunciations than L1 listeners). However, the real difference between the language groups should surface in reduced-speech perception (since L2 listeners lack experience with reduced variants in particular).

1.3.3. The effect of L2 reductions on processing

The third main research question is: what mechanisms may underlie a reduced-form disadvantage in the L2 learners? Specifically, can perceptual difficulties with reduced speech be related to obstructions in bottom-up and/or top-down processing (Table 1)? Below, we will first explain bottom-up and top-down processing and why we believe that L2 reductions can impair both, before describing how we address the question in this paper. In linguistic theory, listening comprehension involves several levels of processing from a low-level acoustic analysis of the speech signal to high-level integration of various types of linguistic and non-linguistic information (as in several models of speech comprehension, e.g., in TRACE, McClelland & Elman, 1986 and BiPhon, Boersma, 2011). Two processing directions through these levels are commonly discerned: bottom-up and top-down. The bottom-up route proceeds from lower to higher levels with increasingly larger building blocks, i.e., from acoustic elements (such as frequencies) to phonetic elements (such as formants) and smaller phonological elements (such as speech sounds and syllables) to words forms and meanings and to larger-scale forms and meanings (as in sentences and stretches of sentences). The top-down route proceeds in the reverse direction and refers to the influence of higher-level knowledge on lower-level processing (such as the influence of memorized words on perceived phonological elements). This knowledge makes listeners expect words and other speech patterns. Although models of processing differ in the role they assign to the two processing directions, it is clear that L1 and L2 listeners use both in listening comprehension (Blank, 1979; Studdert-

[2] At least vocabulary breadth (how many words one knows) should be sufficient, as separate from 'vocabulary depth' (how well one knows the words; Read, 2004).

Kennedy, 1974; Tsui & Fullilove, 1998; Hulstijn, 2003; Field, 2004; Vandergrift, 2007).

When reflecting on the processing of reductions for the purpose of the present study, it seemed possible that problems would occur in both routes. Reductions could complicate the bottom-up route, as follows. Many reduction patterns are language-specific, even when comparing closely related languages such as German and Dutch (Darcy, Ramus, Christophe, Kinzler, & Dupoux, 2009; Torreira & Ernestus, 2011; Mitterer & Tuinman, 2012). Consider a Dutch reduction pattern that corresponds to the above-mentioned German pattern for *haben* in the context *haben wir* (Introduction 1.1). When in Dutch *hebben* [ɦɛbən] occurs in the same context (*hebben we*), it does not reduce to *hebn* [ɦɛbn] and *hem* [ɦɛm] (which would result if the pattern was the same as in German), but to *hebbe* [ɦɛbə] and *hew* [ɦɛʊ], at least in Standard Dutch (henceforth 'Dutch'). Hence, listeners have to perceive L2 reduced speech in an L2-specific way. However, there is evidence that they tend to perceive reductions automatically according to the rules of their L1 (Mitterer & Tuinman, 2012). In addition, automatic L1 speech processing is well-known for speech sound perception (Polivanov, 1931/translation 1974) and the segmentation of words from the speech stream (Cutler, 2000), which are both crucial for reduced-speech perception. (As for speech sound perception, recall the perceptual advantage of being able to pick up acoustic–phonetic details such as the glottalization of the [m] in the example *hasm̰mo-mentsait* above. Reductions may aggravate the word segmentation problem, because they can blur word boundaries that seem already absent for L2 listeners in unreduced speech, further; Cutler, 2000).

In this paper, we take advantage of this tendency for low-level L1 speech processing where L2 processing is required, for signalling obstructions in bottom-up processing due to reductions. For this, we focus on a German reduction pattern that is absent in Dutch (Method 2.2.1). If our Dutch listeners to German cannot help but listen in a native way, the pronunciation *hama* for *haben wir* could invoke various 'mistakes' at higher processing levels, namely the failure to recognize any word (the L2 learner cannot map [ham] onto a word), false word segmentations (the L2 learner segments [hamɐ] as a single word), false recognitions of existing words (the L2 learner falsely maps [hamɐ] onto the German word *Hammer* 'hammer') and false 'recognitions' of non-existing words (L2 learners may simply assume that [hamɐ] represents an unknown word). These mistakes can materialize in the answers given in the phrase-intelligibility task as, respectively, deletions or non-word substitutions of words, incorrect word segmentations, word substitutions and non-word substitutions.

Reductions could also complicate the top-down route, as follows. Given that L2 learners are mostly exposed to unreduced (written) materials, they will expect mainly memorized unreduced forms, but they will often not find such forms in reduced speech. Additionally, an earlier false recognition of an unintended word (the false recognition of 'hammer') may make an L2 learner search in vain for related words in the upcoming speech stream (e.g. for 'nail'). Both situations can materialize in the phrase-intelligibility task as deletions of words (the L2 learner gives up). Of course, it is also possible that L2 learners think falsely that they hear an expected word. For instance, when they hear 'nail' and expect 'hammer', they may falsely recognize *Hammer* for *hama*. This mistake can appear in answers in the phrase-intelligibility task as incorrect word segmentations and word substitutions. Notice that all these cases of inadequate top-down processing must involve inadequate bottom-up processing too, since the listener must resolve the L2-specific reduction patterns appropriately, before he or she can use top-down processing effectively. Hence, when exploring different error types in L2 learners' answers in the phrase-intelligibility task, we will use non-word substitutions as a sign of obstructions in bottom-up processing specifically, and deletions, incorrect word segmentations and word substitutions as signs of obstructions in at least bottom-up processing. Error types are thus more suitable for pinpointing inadequate bottom-up than top-down processing. (For non-word substitutions there is less uncertainty as to the processing route, because it is unlikely that listeners anticipate specific non-memorized nonsense sequences). If reductions hamper bottom-up processing, non-word substitutions and the other error types should occur more when L2 learners hear reduced than unreduced speech.

Apart from looking at error types, we also use a second way to explore possible obstructions in bottom-up and top-down processing: we manipulate the 'amount of context' surrounding target words in the phrase-intelligibility task. For this, listeners are presented with part-phrases ('less context') and full phrases ('more context'). We do not exclude either processing route for full or part-phrases (listeners will always have to use acoustic–phonetic cues and hence bottom-up processing, and top-down processing does not require a full phrase: it also occurs when listeners use single-word knowledge to resolve phonemes; Ganong, 1980). However, we make use of the fact that optimal perception of part- versus full phrases requires a shift in the balance between bottom-up and top-down processing. Specifically, the more speech material is removed, the more the remaining cues are local acoustic–phonetic cues and the more perception must be solved in a bottom-up manner. In the phrase-intelligibility task, bottom-up processing can therefore be considered more important in the part- than full phrases, while top-down processing is more possible in the full than part-phrases. Hence, if L2 reductions hamper bottom-up processing in our young L2 listeners, we should find lower perception scores for reduced than unreduced speech at least in the part-phrases, where top-down processing is least possible and bottom-up processing is most required. To appreciate the difficulty of trying to use top-down processing for the part-phrases adequately, consider hearing a reduced part-phrase such as *aini* [aɪnɪ] for *eigentlich* 'actually', without knowing how many words the phrase represents (many part-phrases were single words representing both function words and content words, thus complicating the anticipation of specific words further; Method 2.2.1).

As for a possible influence of reductions on top-down processing, we examine what happens to perception scores, when full phrases are presented after part-phrases, i.e., when more higher-level cues become available and hence the possibilities for top-down processing increase. If our young L2 listeners improve less when listening to reduced than unreduced speech, this will be a sign that reductions limit the possibility to exploit top-down processing to infer the target words in the part-phrase. (Notice that the lower-level cues in the target words remain the same in the full phrases as in the part-phrases, so that the demands for bottom-up processing remain the same).

All in all, we predict to find obstructions in both bottom-up and top-down processing (see the reasoning above in this section). The results of the previous study in our project (with adolescent and adult L1 listeners; Introduction 1.1) make us expect obstructions in at least bottom-up processing. As mentioned above (Introduction 1.1), these results indicated that L1 speech perception still develops in adolescence and this development seemed to reflect a development in the skills to pick up acoustic–phonetic detail and to handle acoustic–phonetic cue variability. If these skills are so difficult to acquire for L1 listeners, L2 learners can be expected to also experience difficulties here. Since the skills represent lower-level analyses in the comprehension of phrases, bottom-up processing is expected to be obstructed.

2. Method

This study belongs to a larger project on the comprehension of German reduced speech by German L1 and Dutch L2 listeners. The method was identical to that used in the previous study in this project, on differences between adult and adolescent L1 perception (Wanrooij & Raijmakers, 2020).

2.1. Participants

39 Dutch and 38 German adolescents participated. The German adolescents were the control group (and the experimental group in Wanrooij & Raijmakers, 2020). All participants were raised monolingually and attended the same school type, namely a high school that prepares its students for university. The German adolescents were recruited in the German state Lower Saxony, the Dutch adolescents in the Dutch province South Holland. All adolescents and their parents signed informed consent forms. The procedure was approved by the Ethical Committee of the Department of Educational Sciences of Leiden University (approval reference number: ECPW-2017/169).

Participants were assigned randomly to either the ‘Reduced’ condition, where they were presented with German reduced speech, or the ‘Unreduced’ condition, where they were presented with German unreduced speech. Hence, ‘Condition’ was a between-subject rather than within-subject variable. It was not an option to present participants with the same stimulus twice – i.e., once reduced and once unreduced – since this would invoke priming. Presenting participants with a part of the stimuli in a reduced version and the other part in an unreduced version was also not adequate, because the available stimuli could not be counterbalanced properly in terms of other variables than the degree of reduction (such as stimulus duration, word familiarity and the number of target words, which may all affect intelligibility too). The available stimuli were confined to examples extracted from a corpus of spontaneous speech (Method 2.2.1 and 2.3.1), because we wanted to be sure that they had occurred in real life and hence contained possible reductions. Table 2 shows the ages and sex distributions per language (L1 = German native listeners, L2 = Dutch non-native listeners) and condition. The German L1 listeners were one high-school year behind the Dutch L2 listeners and on average 9 months younger.

The Dutch adolescents in the two conditions should not differ in listening comprehension capabilities before the experiment. To confirm this *absence of a difference*, we compared their scores on the last German listening comprehension test in school before participating, in a *Bayesian* independent-samples *t*-test (we use a Bayesian test, because we aim to confirm the null hypothesis, which is not possible with frequentist tests). The scores did not differ (Dutch Reduced: mean = M = 5.76; standard deviation = SD = 1.39; 95% confidence interval = CI = 5.11, 6.41. Dutch Unreduced: M = 6.02; SD = 1.00; CI = 5.54, 6.51. Alternative hypothesis: Cauchy distribution; Bayes Factor of null versus alternative hypothesis = BF_{01} = 2.68).

Considering the poor means (maximum possible score = 10), we wanted to ascertain that the adolescents were not ‘abnormally’ poor, i.e., that the high-school scores in each condition did not deviate from the mean national score for this level, which turned out to be 6.1 (mail communication with CITO, the organization that monitors the tests in the Netherlands, on December 6, 2018). This was confirmed (Bayesian one-sample *t*-tests; Reduced: BF_{01} = 2.55; Unreduced: BF_{01} = 3.99). Hence, in both conditions, German listening comprehension proficiency was representative of general Dutch high-school students’ performance.

2.2. Phrase-intelligibility task

The main task to address research questions 1, 2 and 3 was the phrase-intelligibility task (Table 1; Introduction 1.3). Stimuli consisted of ‘full phrases’ (=phrases comprehensible as stand-alone units, such as *schönen guten Tag* ‘beautiful good day’, a greeting) and part-phrases extracted from full phrases (such as *zusammen* ‘together’, extracted from *zusammen Kaffee trinken* ‘drink coffee together’; stimulus details in Method 2.2.1). Table 3 shows how they were presented. The task was divided into two subtasks, each consisting of 24 trials. In both subtasks, participants typed their answers in each trial twice, i.e., after Presentation A (Answer A) and after Presentation B (answer B). In subtask 1, a full phrase was presented during Presentation A (e.g., *schönen guten Tag*) and this full phrase was repeated twice during Presentation B (e.g., *schönen guten Tag – schönen guten Tag*). This last repetition was added to make subtask 1 similar to subtask 2, where Presentation B also consisted of two parts. In subtask 2, a part-phrase was presented during Presentation A (e.g., *zusammen*), followed in Presentation B by the full phrase plus the part-phrase again (e.g., *zusammen Kaffee trinken – zusammen*). This final presentation of the part-phrase was to remind participants that they only had to write down the part-phrase. Thus, for both Answers A and B, participants wrote down the full phrase in subtask 1 and the part-phrase in subtask 2. Answers A and B yielded the accuracies and the non-word substitutions (and other error types) that we aimed to analyse (As visible in Table 1, the accuracies are relevant for research questions 1 and 2, and in the case of subtask 2 also for question 3; the non-word substitutions are relevant for question 3; Introduction 1.3).

Table 2
Number (N), sex and age of the participants for each Condition in each. Language group (L1 = German native listeners, L2 = Dutch non-native listeners).

Language	Condition	N (sex)	Age in years	
			Mean (SD*)	Range
L1	Reduced	19 (14 F–5 M)	16.19 (0.48)	15.45~17.24
	Unreduced	19 (11 F–8 M)	16.15 (0.29)	15.74~16.69
L2	Reduced	20 (9 F–11 M)	16.99 (0.42)	16.10~17.64
	Unreduced	19 (8 F–11 M)	16.86 (0.38)	16.28~17.44

* : SD = standard deviation.

Table 3
Example stimuli* showing the difference between the subtasks in the phrase-intelligibility task (underlined = what participants were asked to write down, i.e., the full phrase in subtask 1 and the part-phrase in subtask 2; full phrase = 'more context', part-phrase = 'less context').

Subtask	Presentation A (for Answer A)	Presentation B (for Answer B)
1	full phrase *schönen guten Tag*	full phrase – full phrase *schönen guten Tag – schönen guten Tag*
2	part-phrase *zusammen*	full phrase – part-phrase *zusammen Kaffee trinken – zusammen*

* : Participants in the Reduced condition heard a reduced variant of each fragment (e.g., [ʃøːŋgunta:k] for *schönen guten Tag*) and those in the Unreduced condition the unreduced variant (e.g. [ʃøːnəngu:tənta:k]).

The intelligibility judgments (relevant for questions 1 and 2; Table 1) were collected after participants typed Answer A and before they listened to Presentation B. They judged the intelligibility of the phrase (as presented in A) on a 5-point scale, which appeared in German on the laptop screen (1 = highly unintelligible, 2 = unintelligible, 3 = neutral, 4 = intelligible; 5 = highly intelligible). This was done to verify whether the accuracies of the answers corresponded to experienced difficulty.

2.2.1. Stimuli

48 reduced short phrases were extracted from a German corpus with spontaneous dialogues, in which people make appointments (IPDS, 1995, 1996, 1997). To eliminate differences between voices as a potential confound, a male native speaker of German, who had experience in acting and was not a linguist, was asked to copy the phrases. After this, he also pronounced the unreduced versions. The recordings were completed in a sound-proof compartment. To eliminate orthography as a possible confound, there was no orthographic support pertaining to the stimuli or the context of the stimuli (Introduction 1.3). For avoiding that speech rate rather than the reduction patterns would account for the results, reduced and unreduced versions were equated in duration (in the computer program Praat; Boersma & Weenink, 1998–2019; Introduction 1.3).

All phrases are listed in Appendix A, with their IPA (International Phonetic Alphabet) transcriptions. The unreduced versions of most phrases (except seven; see stimulus numbers 25, 35, 37, 41, 43, 44 and 46 in Appendix A) contained unstressed consonant-schwa-nasal consonant sequences, in particular plosive-schwa-/n/ sequences; and the reduced versions featured reductions of these sequences. The German-specific reduction pattern occurring in the plosive-schwa-/n/ strings (Kohler, 2001) is absent in Dutch and is exemplified in Table 4. The table shows increasingly more reduced variants of verbs with such strings: from step 0 with unreduced variants to step 4 with highly reduced variants. The plosive in each example verb has a different place of articulation: /b/ in *haben* is bilabial, /d/ in *werden* 'will' is alveolar and /g/ in *sagen* 'say' is velar. In step 1, the schwa is omitted, resulting in *habn* [haːbn], *werdn* [veːɐdn] and *sagn* [zaːgn]. In step 2, the place of articulation of the final /n/ adapts to that of the preceding plosive, i.e., /n/ is realized as [m] in *haben* and [ŋ] in *sagen* (in *werden* the place of articulation is already the same). In step 3, the manner of articulation of the plosive adapts to that of the subsequent nasal as arisen in the previous step, resulting in nasal geminates. Finally, in step 4, this geminate is reduced to a single nasal. All example variants in Table 4 occur in the above-mentioned corpus. The reduction pattern also surfaces in other word classes than verbs (Appendix A).

Apart from reductions of consonant-schwa-nasal consonant sequences, phrases also included other reduction patterns (as in [flaɪt], a reduced variant of [fɪlaɪçt] *vielleicht* 'perhaps') and further reductions across words (e.g., [hamɐ]). Each phrase contained one to four 'targets', i.e., words that were reduced in the reduced version and unreduced in the unreduced version (the number of targets per phrase is specified in Appendix A). Even though the stimuli contained various degrees of reduction (compare Table 4 and Appendix A), we simply compared adolescents' perception to reduced versus unreduced speech, and hence did not investigate the effect of the specific degree of reduction. This was because we used natural stimuli (rather than synthetic ones), which were mostly highly reduced (rather than exemplifying a solid range of reduction degrees) and which contained complex reductions across words, thereby complicating the scoring of the degree of reduction.

2.2.2. Procedure

Each participant performed the task on a laptop and typed the answers on the laptop keyboard. The stimuli were presented via headphones. The presentation software was E-Prime (Version 2.0, Psychology Software Tools, Inc., 2012). Participants adjusted the sound level to their preferred level before the task.

An experimenter explained the task in German for the German participants and in Dutch for the Dutch participants. Participants practised before each subtask and could ask questions. During the tests, instructions were presented on the laptop screen in German only. If necessary, participants could consult the same instructions on paper, which also included a Dutch translation for the Dutch participants. Participants were requested to type the answers in formal German. They could ignore capital letters (common for nouns in German, but not in Dutch) and umlauts (i.e., the dots on the letters 'ä', 'ö' and 'ü', which are absent in Dutch orthography). The

Table 4
A German reduction pattern in plosive-schwa-/n/ sequences (Kohler, 2001; specific examples: IPDS, 1995; IPDS, 1996; IPDS, 1997), illustrated with increasingly reduced variants, from unreduced (0) to highly reduced (4), for plosives with different places of articulation (bilabial, alveolar, velar). P = plosive; N = nasal; PoA = place of articulation; MoA = manner of articulation.

Step	Reduction (as compared to previous step)	bilabial P	alveolar P	velar P
		haben 'have'	*werden* 'will'	*sagen* 'say'
0	No reduction	[haːbən]	[veːɐdən]	[zaːgən]
1	Pən becomes Pn (schwa drops)	[haːbn]	[veːɐdn]	[zaːgn]
2	Pn becomes PN (PoA adapts)	[haːbm]		[zaːgŋ]
3	PN becomes NN (MoA adapts)	[haːmm]	[veːɐnn]	[zaːŋŋ]
4	NN becomes N (geminate reduces)	[haːm]	[veːɐn]	[zaːŋ]

order of the stimuli in each subtask was randomized for each participant.

2.3. Auditory lexical decision task

The auditory lexical decision task was used for addressing research questions 1 and 2 (Table 1; Introduction 1.3) and particularly for adding a measure of online processing efficiency (reaction times) to the measures collected in the phrase-intelligibility task. The task is performed under time pressure (recall that participants are asked to indicate as quickly as possible whether an auditorily presented item is a word or not). Processing is usually more efficient for words than non-words: listeners can decide immediately once they recognize a word in their mental lexicons, while extra time elapses before they can decide that a presented non-word cannot be found. We will test if this 'item effect' is present as one way to check the validity of the task.

2.3.1. Stimuli

Stimuli were created in the same way as the stimuli for the phrase-intelligibility task (Method 2.2.1). They consisted of 40 German verbs and 40 German pseudo-verbs (total = 80 items), each produced in a reduced and an unreduced version (Appendix B). Each item ended in a consonant-schwa-/n/ sequence (unreduced versions) or a reduced variant of this sequence (reduced versions). Hence, each item was a single target representing the principal German-specific reduction pattern in this paper (Method 2.2.1).

2.3.2. Procedure

Participants performed the task after the phrase-intelligibility task, in the same session. They were seated behind a button box (Chronos Model PST-100430, Psychological Software Tools) and faced a laptop screen. They adjusted the sound level to their preferred level again before the task. Then the task was explained in German for the German participants and in Dutch for the Dutch participants, and participants could ask questions. They also practised (see below). Since the task had to be performed under time and accuracy pressure, there were no additional instructions and also no possibility to ask questions during the task.

Each trial began with a 1000-ms presentation of a star on the laptop screen. Then, participants heard an item and indicated as quickly and accurately as possible whether they considered it a German word or not, by pressing the button of their choice, i.e., by pressing either the leftmost button marked "N" for *nein* 'no' or the rightmost button "J" for *ja* 'yes'. The screen also showed "NEIN" in the top left and "JA" in the top right corner. The next trial started after a button press or after 3 s (post stimulus onset). Participants were informed before the task that all words would be German verbs and all pseudo-words German pseudo-verbs.

Before the task, participants first practised that the 'no'-button was on the left and the 'yes'-button on the right (20 trials). For this, "nein" or "ja" appeared in the middle of the screen (without sound) and participants clicked the appropriate button as quickly as possible. Next trials appeared only after clicking. Then participants did 8 practice trials that proceeded identically to the real-task trials. Practice stimuli did not appear in the real task.

3. Results

The dependent variables and within-subject variables are described below per task. Across tasks, there were two between-subject variables: Language (L1 vs L2, where L1 = German native listeners and L2 = Dutch non-native listeners) and Condition (Reduced vs. Unreduced). The alpha (basis = 5 percent) is corrected for multiple testing, as indicated per analysis.

3.1. Phrase-intelligibility task

Recall from Table 1 that the dependent variables in the phrase-intelligibility task are accuracies (Results 3.1.1), intelligibility judgments (Results 3.1.2) and the number of non-word substitutions and errors of other types (Results 3.1.3). Below, "1″ and "2" refer to subtasks 1 and 2 respectively, "A" and "B" to answers A and B respectively (Table 3; Method 2.2).

3.1.1. Accuracies

A script created in Praat (Boersma & Weenink, 1998–2019) assigned accuracies to each answer of each participant, so that all answers were valued equally. Recall that each stimulus contained one to four target words (which were reduced in the Reduced condition and unreduced in the Unreduced condition; Method 2.2.1) and that participants wrote their answers in response to each stimulus twice (answer A in response to Presentation A; answer B in response to Presentation B; Table 3). For each answer (A, B), the script first determined the accuracy *per target* and then the accuracy for the whole stimulus, i.e., the accuracy *per answer*. The accuracy *per target* was a Boolean value: 0 (incorrect) or 1 (correct). We ignored 'small errors' reflecting that a participant must have recognized the target word, i.e., small spelling mistakes (such as "vieleicht" for *vielleicht* 'maybe'), Dutch spellings for intended German words (such as "Gelegenheid" for *Gelegenheit* 'occasion')

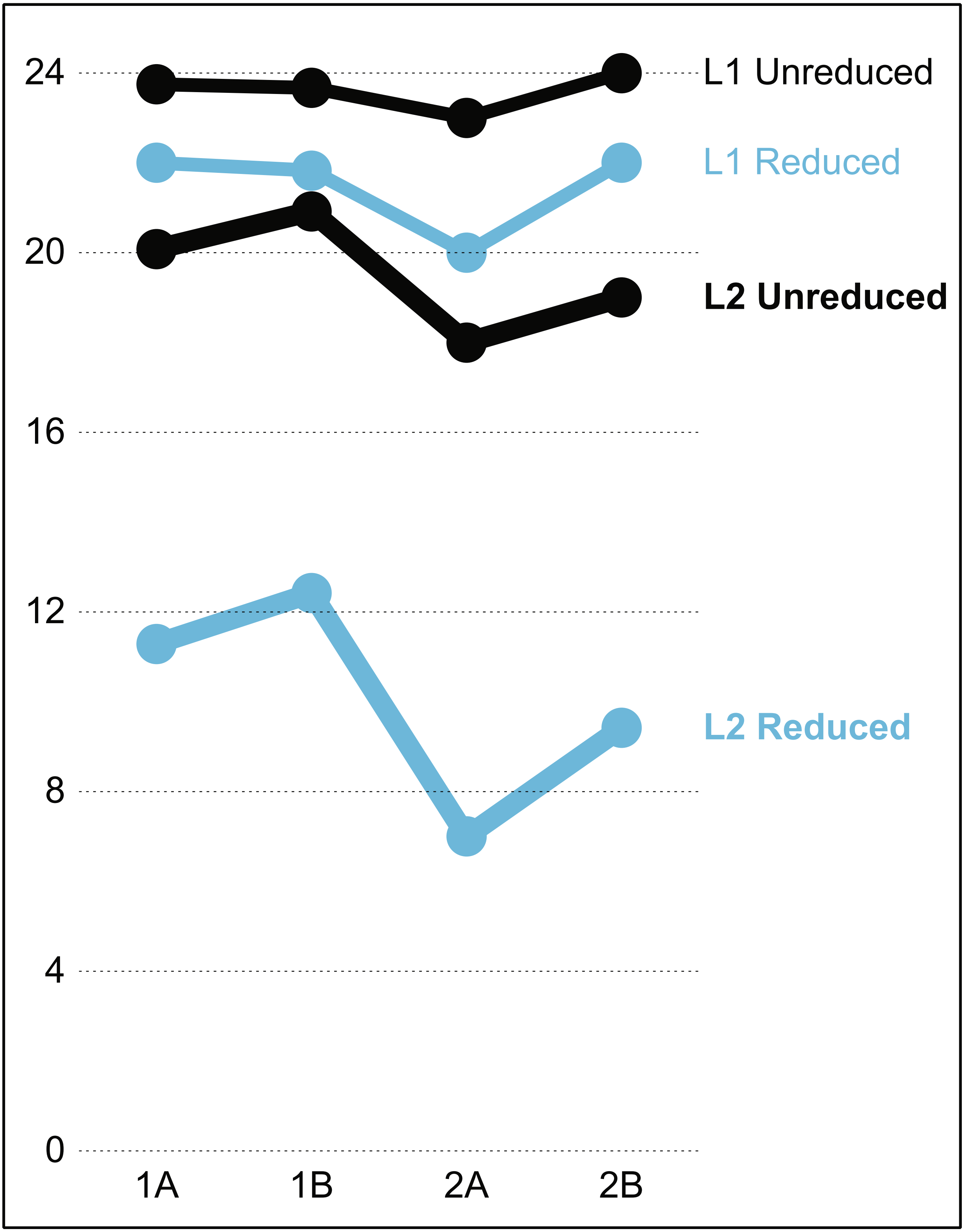

Fig. 1. Median accuracies per Answer in the two Subtasks (1A, 1B, 2A and 2B) in the phrase-intelligibility task, for each Language × Condition (L1 Reduced, L1 Unreduced, L2 Reduced and L2 Unreduced, where L1 = German native listeners and L2 = Dutch non-native listeners).

Table 5
Accuracies in the phrase-intelligibility task: medians and ranges for each Language (L1 = German native listeners and L2 = Dutch non-native listeners) × Condition, per Subtask (1 and 2) and Answer (A and B).

Language	Condition	Median	Range
Answer 1A			
L1	Reduced	22.00	18.00 ~ 23.50
	Unreduced	23.75	21.92 ~ 24.00
L2	Reduced	11.29	7.83 ~ 17.75
	Unreduced	20.08	15.92 ~ 22.67
Answer 1B			
L1	Reduced	21.83	18.00 ~ 24.00
	Unreduced	23.67	22.67 ~ 24.00
L2	Reduced	12.42	7.58 ~ 17.33
	Unreduced	20.92	16.17 ~ 23.25
Answer 2A			
L1	Reduced	20.00	16.33 ~ 22.00
	Unreduced	23.00	20.33 ~ 24.00
L2	Reduced	7.00	3.17 ~ 15.00
	Unreduced	18.00	12.67 ~ 24.00
Answer 2B			
L1	Reduced	22.00	16.00 ~ 24.00
	Unreduced	24.00	20.33 ~ 24.00
L2	Reduced	9.42	4.17 ~ 19.00
	Unreduced	19.00	15.33 ~ 24.00

and agreement errors on determiners (including possessive determiners) and adjectives (in German, the form of determiners and adjectives is determined by case, grammatical gender and number). None of these errors induce changes of meaning. The accuracy *per answer* was calculated as the number of correct targets divided by the total number of targets in the answer, so that participants could get a maximum of 1 point per answer and 48 points in the whole phrase-intelligibility task (=24 per subtask).

Fig. 1 visualizes the median accuracies per Subtask and Answer (1A, 1B, 2A and 2B) for each Language × Condition. Table 5 lists the precise medians and ranges. The German L1 listeners, in particular in the Unreduced condition, scored (near) ceiling, which supports the validity of the test.

Below, we use non-parametric statistical tests, because the accuracies were not normally distributed and could not be transformed meaningfully, due to L1 listeners' ceiling scores.[3] Accuracies were consistently lower for adolescent L2 than L1 listeners, which supports the validity of the task (eight Mann-Whitney tests, one per Condition per Subtask per Answer, alpha = 0.00625; Reduced: all *U*s $\geq$ 0 and $\leq$1.0; all *p*s < 0.001; all *z*s ≤ -5.3; all effect sizes $r \leq -0.85$; Unreduced: all *U*s $\geq$ 3.0 and $\leq$ 32.5; all *p*s < 0.001; all *z*s ≤ -4.4; all *r*s ≤ -0.71; here and below $r = z/\sqrt{N}$; Rosenthal, 1991). More importantly, accuracies were also consistently lower for the Reduced than the Unreduced Condition, for both L1 and L2 listeners (eight Mann-Whitney tests, one per Language per Subtask per Answer, alpha = 0.00625; L1: all *U*s $\geq$ 6.5 and $\leq$ 79.5; all *p*s $\leq$ 0.0018; all *z*s ≤ -3.0; all *r*s ≤ -0.49; L2: all *U*s $\geq$ 1.0 and $\leq$ 9.0; all *p*s < 0.001; all *z*s ≤ -5.1; all *r*s ≤ -0.81). This answers research question 1 in the affirmative: L2 reductions (corrected for the possible confounding factors speech rate, orthography and voice) indeed hamper 17-year-old high-school students' perception, in that adolescents' perception of L2 reduced speech is less accurate than that of unreduced speech (question 1a in Table 1). The effect sizes show that the disadvantage is substantial (question 2a in Table 1) which is visible in Fig. 1 too. That the reductions also hamper adolescent L1 listeners' perception (albeit to a much lesser extent; see below) supports the validity of the task (Introduction 1.1).

Judging from the medians (Fig. 1) and the effect sizes (the just-reported *r*s), the reduced-form disadvantage looks substantially larger for adolescent L2 than L1 listeners (research question 2; Table 1). Since non-parametric tests do not reveal possible interactions, we assessed this difference between L1 and L2 listeners (=the interaction Language × Condition) in another way. First, we calculated the accuracy for each *stimulus* (rather than for each participant) as the percentage[4] of participants in each Language × Condition group (=L1 Reduced, L1 Unreduced, L2 Reduced and L2 Unreduced) with correct answers in response to that stimulus. Then we calculated the reduced-form disadvantage (=accuracy in Unreduced condition – accuracy in Reduced condition) per stimulus, for each Language. Four Wilcoxon signed-ranks tests (pairing the L1 and L2 reduced-form disadvantages per stimulus; one test per Subtask per Answer, alpha = 0.0125) confirmed a larger disadvantage for adolescent L2 than L1 listeners, for both answers in both subtasks, i.e., for 1A (L1 median = 2.63, range = -1.32 to 50.00 percent; L2 median = 34.67, range = -5.26 to 94.74 percent, T = 18.00, $z = -3.772$, $p < 0.001$, $r = -0.54$), 1B (L1 median = 0, range = -5.26 to 47.37 percent, L2 median = 28.22, range = -2.89 to 94.74 percent, T = 7.00, $z = -3.984$, $p < 0.001$, $r = -0.58$), 2A (L1 median = 5.26, range = -15.79 to 94.74 percent, L2 median = 39.21, range = -2.37 to 100 percent, T = 41.00, $z = -3.115$, $p = 0.001$, $r = -0.45$), and 2B (L1 median = 5.26, range = -10.53 to 36.84 percent, L2 median = 41.25, range = -6.84 to 95.00 percent, T = 5.00, $z = -4.143$, $p < 0.001$, $r = -0.60$). In sum, the reduced-form disadvantage is larger for adolescent L2 than L1 listeners (question 2b in Table 1), thus suggesting that experience with reduced variants is lacking in these young listeners and that this experience in particular is necessary for learning to perceive L2 speech in real life (Introduction 1.3.2).

We had also predicted that the L2 learners would perform rather well on the unreduced stimuli (Introduction 1.3.2). This is indeed visible in Fig. 1 (but see Results 3.1.3). It also follows from the just reported analyses for accuracies per stimulus that the difference between the Language groups (i.e., accuracies for L1 – accuracies for L2) is smaller for unreduced than reduced speech.

One way to explore the effect of reductions on bottom-up and top-down processing was to manipulate the amount of context surrounding target words in Subtask 2 (research question 3 in Table 1; Introduction 1.3.3). Recall that in this subtask, participants were first presented with part-phrases and then with the corresponding full phrases (Table 3). In the

[3] Notice that the accuracies are not simply binomially distributed: the number of possible outcomes depends on the stimulus: it is two (one target: 0 or 1 point for the answer), three (two targets: 0 or 1/2 or 1 point), four (three targets: 0 or 1/3 or 2/3 or 1 point) or five (four targets: 1 or 1/4 or 1/2 or 3/4 or 1 point). Changing the way of scoring into a simple "accurate" versus "inaccurate" for all answers would minimize the sensitivity of the task. Therefore, a logistic regression or a linear mixed-effects analysis with a binomial dependent variable is not possible as an alternative for the non-parametric tests used here.

[4] The percentage is used rather than the number of participants to compensate for the fact that the German group contained one participant less than the Dutch group in the Reduced condition. However, taking the number of participants yields the same results.

part-phrases, the possibilities for using top-down processing were limited, and participants had to rely predominantly on bottom-up processing. Hence, a hampering influence of reductions on bottom-up processing would have to surface as lower accuracies for reduced than unreduced speech at least here (i.e., in the part-phrases = answer 2A). As indicated above, these scores were indeed lower for the Reduced than Unreduced condition. In fact, they were the lowest accuracies in the task (significance values: see above for the difference between the Reduced and Unreduced condition and below for the difference between part- and full phrases; see also Fig. 1).

To detect signs of inadequate top-down processing, we examined L2 listeners' capability to exploit higher-level cues in the full phrases presented after the part-phrases. If listeners were able to exploit those cues, their scores were expected to be higher in the full phrases (answer 2B) than the part-phrases (answer 2A). A hampering effect of reductions on top-down processing would have to surface as a smaller accuracy gain (2B – 2A) in the Reduced than Unreduced condition. Accuracies turned out to be significantly lower in the part-phrases than full phrases, for both conditions, showing that adolescent L2 learners, irrespective of whether they hear reduced or unreduced speech, can exploit the availability of higher-level cues when more context is provided in this task (two Wilcoxon signed-ranks tests; alpha = 0.025; L2 Reduced: median = 1.50, range = −0.67 to 6.00, T = 6.00, z = −3.463, p < 0.001, r = −0.77; L2 Unreduced: median = 1.00, range = −1.67 to 4.00, T = 12.50, z = −2.877, p = 0.002, r = −0.66). L2 listeners' accuracy gain with more context (2B – 2A) was not significantly larger for the Unreduced than Reduced condition (U = 166.50, z = −0.662, p = 0.526). Hence, it is not clear whether our young L2 listeners benefit more from the availability of higher-level cues in unreduced than reduced speech, when after a part-phrase the corresponding full phrase is presented. Further non-planned exploratory comparisons demonstrated that when disregarding the L1 listeners in the Unreduced condition (who scored ceiling already for the part-phrases and hence had no room for improvement), the accuracy gain was not significantly different across Language × Condition groups (L1 Reduced: median = 2.50, range = −2.17 to 5.00, T = 9.00, z = −3.469, p < 0.001, r = −0.80; L1 Unreduced: median = 0, range = −1.00 to 2.00, T = 14.00, z = −1.029, p = 0.371; L2: see above; three pairwise Mann-Whitney comparisons, alpha = 0.017; ps ≥ 0.029). Hence, adolescent listeners irrespective of whether they listen to their L1 or L2 and regardless of whether they listen to reduced or unreduced speech, seem well able to exploit top-down processing when more context is provided in a task like the phrase-intelligibility task. All in all, the manipulation of the amount of context in the phrase-intelligibility task (research question 3; Table 1) suggests that L2 reductions hamper at least bottom-up processing, while for top-down processing the hampering effect is unclear.

Finally, notice that simply repeating a full phrase (1A versus 1B) was also significantly beneficial for L2 listeners, in both conditions (two Wilcoxon signed-ranks tests, alpha = 0.025; L2 Reduced: median = 1.00, range = −1.75 to 2.50, T = 38.00, z = −2.299, p = 0.020, r = −0.51; L2 Unreduced: median = 0.92, range = −1.17 to 2.33, T = 21.00, z = −2.812, p = 0. 003; r = −0.65). The accuracy gain (1B – 1A) was not significantly different between the two conditions (U = 189.50, z = −0.014, p = 0.994). Hence, it is unclear if the repetition of full phrases helps adolescent L2 listeners more when the phrases are reduced or unreduced.

Table 6
Intelligibility judgments in the phrase-intelligibility task: medians and ranges for each Language (L1 = German native listeners, L2 = Dutch non-native listeners) × Condition, per Subtask (1 and 2).

Language	Condition	Median	Range
Subtask 1			
L1	Reduced	3.5	3–5
	Unreduced*	5	4–5
L2	Reduced	3	1–4
	Unreduced	4	3–5
Subtask 2			
L1	Reduced	3	1–4
	Unreduced*	4	3–5
L2	Reduced	2	2–4
	Unreduced	4	2–5

* = excluding two participants (see text).

3.1.2. Intelligibility judgments

Two L1 listeners in the Unreduced condition probably judged the items reversely (i.e., they judged the stimuli for which they had the lowest scores as the 'easiest' and stimuli for which they had the highest scores as the 'hardest') and were removed from the dataset. Medians and ranges of intelligibility judgments are listed in Table 6. Recall that the judgments represent a five-point scale and that they were given after answers A only (Method 2.2).

Overall, the effects observed in the accuracies were also effects in the intelligibility judgments, thus supporting that the accuracies reflect consciously experienced difficulty. Specifically, adolescent L2 listeners judged the stimuli to be less intelligible than L1 listeners in Subtask 1 and in the Reduced condition also in Subtask 2 (four Mann-Whitney tests, alpha = 0.0125; Reduced 1: U = 73.00, z = −3.548, p < 0.001, r = −0.57; Reduced 2: U = 84.50, z = −3.120, p = 0.002, r = −0.50; Unreduced 1: U = 86.50, z = −2.631, p = 0.009, r = −0.43; Unreduced 2: U = 137.00, z = −0.821, p = 0.412). More importantly, adolescents in the Unreduced condition judged the stimuli to be more intelligible than adolescents in the Reduced condition, irrespective of whether adolescents listened to their L1 or to the L2 (four Mann-Whitney tests, alpha = 0.0125; L1 Subtask 1: U = 60.50, z = −3.363, p < 0.001, r = −0.55; L1 Subtask 2: U = 51.00, z = −3.644, p < 0.001, r = −0.59; L2 Subtask 1: U = 37.00, z = −4.519, p < 0.001, r = −0.72; L2 Subtask 2: U = 25.00, z = −4.786, p < 0.001, r = −0.77). Hence, the intelligibility judgments further support the affirmative answer to research question 1 (i.e., that reductions hamper perception in adolescent L2 listeners; Table 1). The effect sizes for the L2 learners are also substantial, again mirroring the effects observed in the accuracies (research question 2; Table 1).

3.1.3. Adolescent L2 listeners' errors of different types

For further exploring how reductions may hamper processing (research question 3; Table 1), Dutch L2 listeners' first answers in the phrase-intelligibility task (answers A) were coded for the presence of several error types: false segmentations of targets, deletions of targets, and substitutions of tar-

Table 7
Examples of substitutions (in bold) of target words by non-words, German words and Dutch words in the first answers (answers A) given by adolescent Dutch listeners to German *reduced* phrases in the phrase-intelligibility task.

Substitution by:	Subtask 1	Subtask 2
	"einen gemütlichen abend verbringen"	"das können wir"
non-words	**eigemütlichen anverbring**	das **könwa**
	ein gemütlichen **anfutling**	das **kömmer**
German words	eine gemütlichen abend **von mir** ('from me')	das **Kummer** ('sorrow')
	ein gemütlichen **arm** verbringen ('arm')	das **kann mal** ('can' 'once')
	"wegen der Uhrzeit"	"dann sind wir uns"
non-words	**wingerwutszeit**	**sommers**
	winder uhr seit	bis **ziemals**
German words	**wenn** uhrzeit ('when')	**das Zimmer** ('the room')
	wenige uhrzeit ('few')	dann **sie muss** ('she must')
	"ganz hervorragend passen"	"allerdings"
Dutch words	ganzer **voorrang** fassen ('priority')	**alleen** ('alone')
	"an einem Wochenende"	"im Grunde genommen"
	an **einde** wochenende ('end')	im **groningen** kann **men** (place name; 'one')

gets by non-intended German words, non-words and Dutch words (somewhat unexpectedly, listeners also replaced German words with Dutch words). As discussed in the Introduction (1.3.3), non-word substitutions are taken as signs of failing bottom-up processing, and deletions, incorrect word segmentations and word substitutions as signs of at least unsuccessful bottom-up processing.

Each answer was assigned the value 0 (error absent) or 1 (error present). Since there were 48 answers per adolescent, the maximum possible number of errors per error type was 48 per participant. Table 7 gives examples of substitutions in the Reduced condition. The examples also illustrate false segmentations. For instance, the answer "eigemütlichen anverbring" (two non-words) for *einen gemütlichen Abend verbringen* 'to spend a cosy evening' contains two word mergers: it does not correctly insert a boundary between *einen* and *gemütlichen*, and also not between *Abend* and *verbringen*. The answer "eine gemütlichen abend von mir" in response to the same stimulus contains a word split: *verbringen* is perceived as two words *von* 'from' and *mir* 'me'. Both answers are assigned the value 1 for false segmentations. The value 1 is also assigned to the first answer for non-word substitutions and to the second answer for German word substitutions.

Participants occasionally replaced target words (partly) by Dutch words (Table 7). These Dutch substitutions included Dutch spellings for intended German words that were counted as correct (Results 3.1.1). To ensure transparency, we coded ambiguous substitutions for all possibilities. For instance, the substitution "arm" for *Abend* (in the answer "ein gemütlichen arm verbringen"; Table 7) can be both a German and a Dutch word. Therefore, the answer received the value 1 for German word substitutions and Dutch word substitutions.

Answers were coded for deletions only when they lacked any trace of the targets. Thus, the answer "eigemütlichen anverbring" (Table 7) was assigned the value 0 for deletions: the "ei" in "eigemütlichen" could be considered a trace of *einen*, the rest of this non-word reflected *gemütlichen*, the "an" in "anverbringen" could be a trace of *Abend* and the remainder of this non-word contained *verbringen*.

Fig. 2 illustrates the number of answers with false segmentations, deletions and substitutions of target words by non-words, German words and Dutch words, per Dutch participant in each condition (medians and ranges in Appendix C). Adolescents in the Reduced condition made more errors of every type than those in the Unreduced condition, i.e., more answers with false segmentations (Mann-Whitney $U = 0.00$, $z = -5.373$, $p < 0.001$, $r = -0.86$), deletions ($U = 26.00$, $z = -4.654$, $p < 0.001$; $r = -0.75$), non-word substitutions ($U = 34.00$, $z = -4.393$, $p < 0.001$; $r = -0.70$), German word substitutions ($U = 0.00$, $z = -5.353$, $p < 0.001$; $r = -0.86$) and Dutch word substitutions ($U = 94.00$, $z = -2.800$, $p = 0.004$; $r = -0.45$). Hence, Dutch adolescents miss words and misunderstand words substantially more often when German speech is reduced than when it is not reduced. The larger number of errors in the Reduced condition (particularly the larger number of non-word substitutions) are indicative of inadequate bottom-up processing (Introduction 1.3.3).

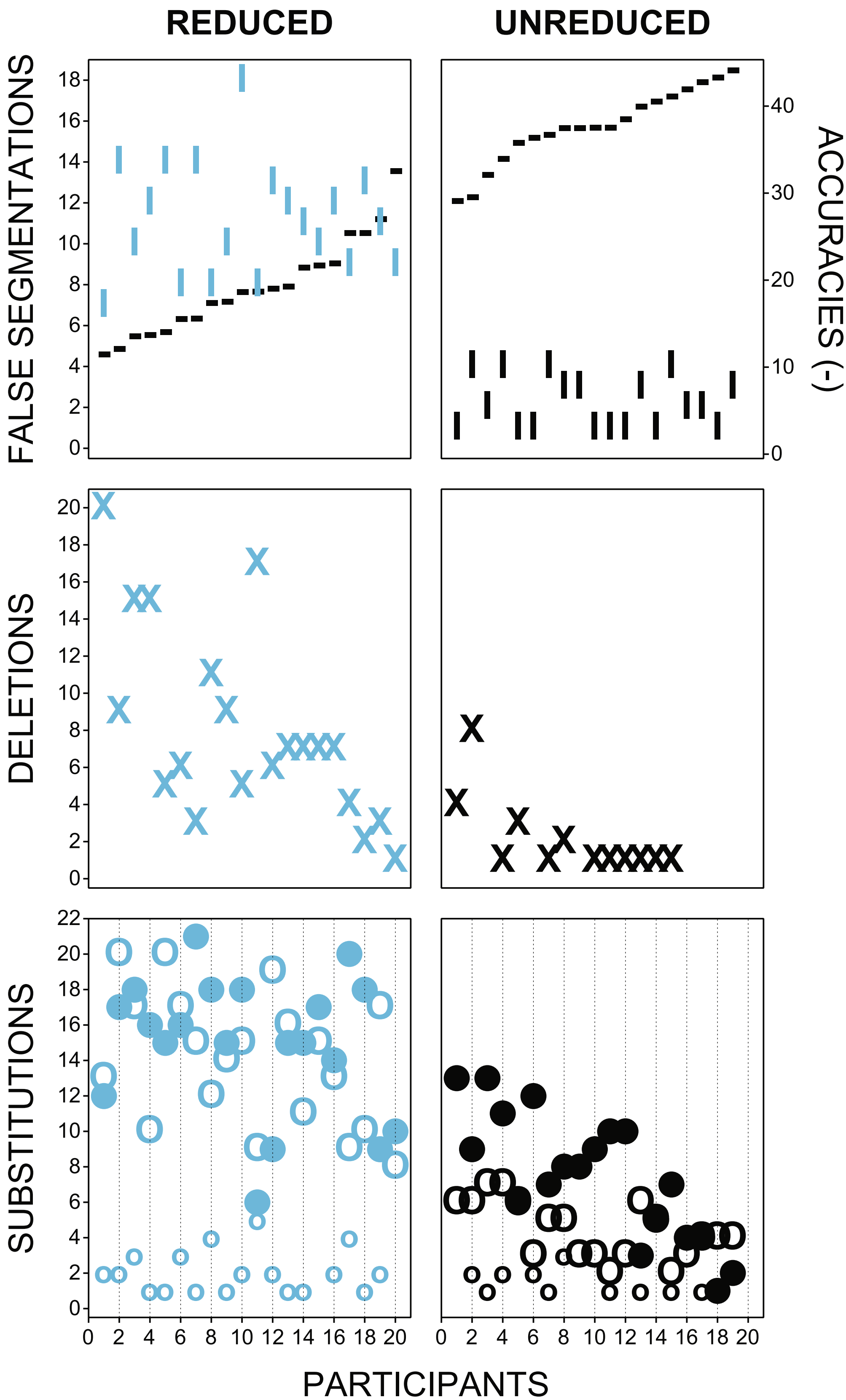

Fig. 2. Dutch adolescents' errors in answers A in the German phrase-intelligibility task. The number of answers with false segmentations (top), deletions (middle) and substitutions (bottom), for each Dutch adolescent in the Reduced (left, blue) and Unreduced (right, black) Condition. Adolescents are arranged from the lowest (left) to highest (right) accuracies (dashes, top). I = false segmentations; X = deletions; ● = non-word substitutions; O = German word substitutions; o = Dutch word substitutions.

Table 8
The number of answers per participant included in the analyses of the accuracies and reaction times in the auditory lexical decision task, for the Words and Pseudo-words separately: mean (range between brackets) in each Language (L1 = German native listeners and L2 = Dutch non-native listeners) × Condition. The maximum number per participant was 40, both for Words and Pseudo-words.

Language	Condition	Words	Pseudo-words
L1	Reduced	38 (35–40)	37 (31–40)
	Unreduced	38 (33–40)	37 (32–39)
L2	Reduced	37 (35–40)	38 (35–40)
	Unreduced	38 (35–40)	38 (34–40)

As visible in Fig. 2, all adolescents used word *and* non-word substitutions. Almost all adolescents (except one in the Reduced condition and 8 in the Unreduced condition) also did so at least once within a single answer (e.g., Table 7: *bis ziemals* = a German word 'until' + a non-word). For assessing the relative use of non-word and German word substitutions in a conservative way, these 'mixed' answers were disregarded. Specifically, we subtracted the number of answers with non-word substitutions only from that with German word substitutions only, for each participant. A one-sample *t*-test comparing these scores to zero showed that adolescents' answers with non-word substitutions outnumber those with German word substitutions (mean difference German – non-word substitutions = −2.05, *CI* = −3.50, −0.60, *t*[38] = −2.857, *p* = 0.007). The difference score did not differ significantly between the conditions (*t*[37] = −1.375, *p* = 0.177). This suggests that deficiencies in adolescents' L2 speech perception in high school, irrespective of whether they hear reduced or unreduced speech, are related to at least inadequate bottom-up processing.

3.2. Auditory lexical decision task

The analysis was confined to right-handed participants: 18 L1 Reduced (13F, 5 M; Age: *M* = 16.23, *SD* = 0.46 years), 16 L1 Unreduced (9F, 7 M; Age: *M* = 16.13, *SD* = 0.29 years), 17 L2 Reduced (7F, 10 M; Age: *M* = 16.93, *SD* = 0.53 years), 19 L2 Unreduced (8F, 11 M; Age: *M* = 16.86, *SD* = 0.38 years). Reaction times were measured from stimulus onset and log-transformed (base 10) in order to make the distributions approximately normal. Undefined reaction times (when participants had not given their answers within 3 s) and log-transformed reaction times beyond 2 standard deviations from the mean of each Group's answers to each Item Type (Word vs. Pseudo-word) were deleted from the dataset (where Groups = Language × Condition, i.e., L1 Reduced, L1 Unreduced, L2 Reduced and L2 Unreduced). The remaining answers (numbers in Table 8) were included in the analyses. Table 8 shows that in each Group, 37 to 38 answers per participant on average (out of 40 answers) were included in the analyses, for both words and pseudo-words.

Because the number of participants per group was unbalanced (see the removal of left-handers above), we applied mixed-effect models to analyse the answers (Baayen, Davidson, & Bates, 2008; Jaeger, 2008; Winter, 2013; Bates, Maechler, Bolker, & Walker, 2015). For the models, we used the package *lme4* (Bates et al., 2015) in the program R (R Core Team, 2020). The accuracies and log-transformed reaction times (dependent variables) were analysed separately, the accuracies with a generalized linear mixed-effects model (using the *lme4*-function *glmer*) and the log-transformed reaction times with linear mixed-effect models (*lmer*). Participant and Item were included in the models as random effects; Item Type (Words vs. Pseudo-words), Language (L1 vs. L2) and Condition (Reduced vs. Unreduced) as fixed effects. The random effects were only modelled as random intercepts, not as random slopes (the models were over-parametrized with random slopes). The precise models are reported in Appendix E. The alpha for testing the significance of each main and interaction effect was set at 0.00714 (=0.05/7 tests; one test per main or interaction effect).

3.2.1. Accuracies

Fig. 3 (top; precise numbers in Appendix D) presents the accuracies obtained by the adolescent L1 and L2 listeners in each Condition, for Words and Pseudo-words separately. The generalized mixed-effects model yielded the outcomes in Table 9. As for the fixed effects, participants scored significantly above chance (*p*-value of intercept in Table 9), thus confirming the validity of the test. Additionally, there were significant fixed effects of Item Type, Language, Condition, Item Type × Language and Item Type × Condition; the other interactions, i.e., Language × Condition and Item Type × Language × Condition, were not significant (*p*-values in Table 9; Notice that the conclusions remain the same when the insignificant effects are removed from the model).

We interpret the significant effects by looking at the odds. Specifically, the odds of the main effects show that adolescents have a 2.63 (*CI* = 1.50, 4.60) times higher chance of a correct answer for a Word than a Pseudo-word (odds for Item Type in Table 9); adolescent L1 listeners have a 2.61 (*CI* = 1.77, 3.86) times higher chance to get an answer correct than adolescent L2 listeners (odds for Language) and adolescents listening to Unreduced items have a 2.13 (*CI* = 1.46, 3.11) times higher chance to answer correctly than adolescents listening to Reduced items (odds for Condition). The last effect further supports the finding reported for the phrase-intelligibility task above, that reductions complicate perception (research question 1). Given the non-significance of Language × Condition, accuracies in the auditory lexical decision task do not clarify whether this reduced-form disadvantage is larger for L2 than L1 listeners (research question 2).

The odds of Item Type × Language indicate that the difference in accuracies between Words and Pseudo-words is larger for adolescent L1 than L2 listeners. Specifically, when the odds are defined as *P(correct Word) / P(correct Pseudo-word)*, i.e., as the chance of a correct answer for a Word versus the chance of a correct answer for a Pseudo-word, then the odds for L1 are 4.60 times larger than for L2. This reflects that adolescent L1 listeners are better at distinguishing words from pseudo-words in this task than adolescent L2 listeners.

The odds of Item Type × Condition signal that the difference in accuracies between Words and Pseudo-words is *smaller* for adolescents listening to Unreduced items than to those listening to Reduced items. Specifically, the odds (as just defined) for adolescents listening to Unreduced items are 0.57 times the odds for adolescents listening to Reduced items. This out-

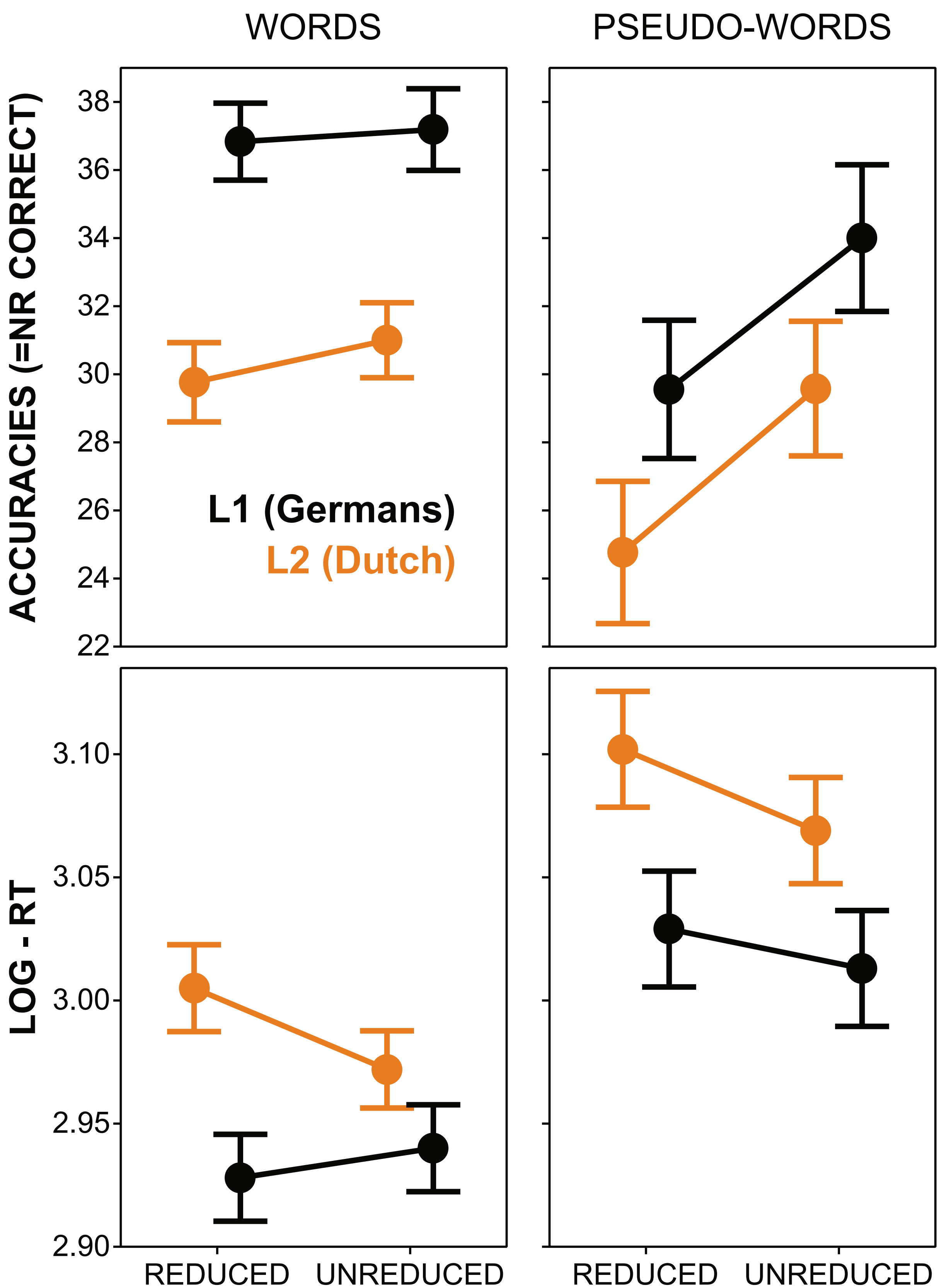

Fig. 3. Accuracies (top) and log-transformed reaction times (bottom, in ms) in the auditory lexical decision task, in response to Words (left) and Pseudo-words (right), per Language (Dutch L2 listeners = orange; German L1 listeners = Black) x Condition (titles at the horizontal axis). Group means (circles) and 95% confidence intervals (vertical lines). Precise numbers in Appendix D.

Table 9
Generalized linear mixed-effects model for the accuracies in the auditory lexical decision task: Fixed effects (top*), with, for each effect: beta estimate, its standard error (SE), the test statistic (*z*-value), the significance (*p*-value) and, for ease of interpretation, the odds plus confidence interval (*CI*). **Random effects (bottom): the standard deviation (*SD*) of the random intercepts for Participant and Item. (Model details in Appendix E).

Fixed effects	Beta	*SE*	*z*	*p*	Odds	*CI*
Intercept	0.763	0.220	3.469	0.000522	2.14	1.39 ~ 3.30
Item Type (Pseudo-word to Word)	0.966	0.285	3.383	0.000717	2.63	1.50 ~ 4.60
Language (L2 to L1)	0.961	0.199	4.821	1.43e−06	2.61	1.77 ~ 3.86
Condition (Reduced to Unreduced)	0.758	0.193	3.937	8.24e−05	2.13	1.46 ~ 3.11
Item Type x Language	1.526	0.292	5.231	1.69e−07	4.60	2.60 ~ 8.15
Item Type x Condition	−0.565	0.201	−2.810	0.004948	0.57	0.38 ~ 0.84
Language x Condition	0.352	0.305	1.156	0.25: n.s.		
Item Type x Language x Condition	−0.634	0.427	−1.486	0.14: n.s.		
Random effects (intercepts)		*SD*				
Participant		0.413				
Item		1.090				

* : For each main fixed variable, the reference category and the other category are indicated between parentheses. The reference category is mentioned first.
** : $p = \Pr(>|z|)$. The *p*-value reflects whether the estimate is significantly different from zero. Odds = e^{beta}; $CI = e^{(beta - 1.96\ SE)} \sim e^{(beta + 1.96\ SE)}$.

Table 10
Linear mixed-effects model for the log-transformed reaction times in the auditory lexical decision task: Fixed effects (top*), with, for each effect: beta estimate, its standard error (*SE*), the test statistic (χ^2) and the significance (*p*-value) of the comparison between a model with and a model without the fixed effect (details: Appendix E). Random effects (bottom): the standard deviation (*SD*) of the random intercepts for Participant and Item.**

Fixed effects	Beta	*SE*	χ^2(df)	*p*
Intercept	3.0993	0.0115		
Item Type (Pseudo-word to Word)	−0.0764	0.0102	57.08 (1)	4.18e−14
Language (L2 to L1)	−0.0715	0.0136	36.14 (1)	1.84e−09
Condition (Reduced to Unreduced)	−0.0299	0.0134	2.94 (1)	0.09: n.s.
Item Type x Language	−0.0221	0.0076	2.69 (1)	0.10: n.s.
Item Type x Condition	0.0033	0.0075	3.55 (1)	0.06: n.s.
Language x Condition	0.0172	0.0193	2.78 (1)	0.10: n.s.
Item Type x Language x Condition	0.0277	0.0111	6.62 (1)	0.01: n.s.
Random effects (intercepts)		*SD*		
Participant		0.03573		
Item		0.04001		
Residual		0.09409		

* : For each main fixed variable, the reference category and the other category are indicated between parentheses. The reference category is mentioned first.
** : A *t*-value can be calculated by: Beta/SE. Alpha = 0.00714 (=0.05/7 tests; see text).

come reflects adolescents' larger readiness to recognize a word in a reduced than unreduced pseudo-word.

3.2.2. Reaction times

Fig. 3 (bottom; numbers in Appendix D) shows the log-transformed reaction times in response to Words and Pseudo-words, for each Language × Condition. Table 10 presents the outcomes of the linear effects modelling (precise models in Appendix E).

There were significant main effects of Item Type and Language (*p*-values in Table 10): the log-transformed reaction times were 0.0764 (*CI* = 0.0564, 0.0964) milliseconds (ms) faster for Words than Pseudo-words and 0.0715 ms (*CI* = 0.0448, 0.0982) faster for adolescent L1 than L2 listeners (beta estimates and *SE*s in Table 10; *CI* = beta ± (1.96 * SE)). For ease of interpretation, we calculated the mean non-transformed reaction times for Words and Pseudo-words separately and for L1 and L2 listeners separately, based on the mean reaction times per participant: yielding 916 (*CI* = 897, 933) ms for Words versus 1130 (*CI* = 1099, 1161) ms for Pseudo-words, and 951 (*CI* = 912, 977) ms for L1 listeners versus 1089 (*CI* = 1057, 1122) ms for L2 listeners. This is in line with the expectations (Method 2.3). Given the non-significant effect of Condition, it is uncertain whether isolated reduced items are processed slower than isolated unreduced items (research question 1c). Interaction effects with Condition were also not significant (*p*-values in Table 10). Hence, we did not find clear evidence that the reduced-form disadvantage in reaction times is larger for adolescent L2 than L1 listeners in this task (question 2). Table 11 summarizes all outcomes in the phrase-intelligibility task and the auditory lexical decision task.

4. Discussion

4.1. Non-native reductions hamper adolescents' perception

This study is the first to systematically examine the effect of reduced pronunciations on L2 speech perception, and accordingly listening comprehension, in adolescents. It shows how 17-year-old Dutch second-language (L2) learners of German in high school struggle with understanding even simple German speech when this speech is 'reduced' as compared to fully pronounced. Here, 'reduced' indicates that words are shortened and merged (e.g., *hama* [hamɐ] for *haben wir* [haːbən viːɐ] 'have we') as is normal in natural speech. Crucially, the present study settles that the perceptual struggles are due to reduction patterns separately from possible other factors not controlled for in previous work with adult listeners

Table 11
Overview of the research questions and the outcomes. PI = phrase-intelligibility task, ALD = auditory lexical decision task.

1.	**Do reductions (separated from factors such as speech rate) hamper perception, and consequently comprehension, in adolescent L2 learners? (both tasks)** Yes, although it remains unclear whether L2 reduced speech also impacts the speed of processing (ALD task), L2 listeners perceive reduced speech less accurately than unreduced speech (PI task, across subtasks, and ALD task), and also judge it as less intelligible (PI task). Hence, the 'reduced-form disadvantage' observed earlier in adult listeners and adolescent L1 listeners is now confirmed for adolescent L2 listeners. Additionally, the results settle that reductions cause a separate disadvantage from speech rate, orthography and voice.
2.	**To what extent? Is the reduced-form disadvantage relevant? (both tasks).** The effect sizes of the disadvantage for adolescent L2 learners are large, both for accuracies (all rs $\leq$ –0.81; Fig. 1) and intelligibility judgments (rs $\leq$ –0.72) in the PI task, and for accuracies in the ALD task (adolescents have a more than two times higher chance to answer correctly in response to unreduced than reduced speech; Fig. 3). Also, the disadvantage is larger than in adolescent L1 listeners (PI task: rs $\leq$ –0.45; Fig. 1). Thus, adolescent L2 learners are not prepared for perceiving and understanding real-life speech (i.e., with reductions) at the end of high school. The problems with reduced speech do not seem to reflect just a general (as in L1 listeners) larger listening challenge of reduced than unreduced speech, but also a lack of experience with specifically reduced variants.
3.	**What processing mechanisms may underlie the disadvantage? (PI task).** Reductions hamper at least bottom-up processing: when top-down processing is made difficult and L2 listeners must rely on bottom-up processing most (in the part-phrases), perception is less accurate and intelligibility judgments are lower for reduced than unreduced speech. Also, non-word substitutions (and all other error types) are more numerous for reduced than unreduced speech. Whether reductions hamper top-down processing too, is less clear: when possibilities for top-down processing increase (in full phrases after part-phrases), perception does not improve less (or more) for reduced than unreduced speech.

(notably speech rate, which also varies in natural speech). The adolescents had studied German in high school for four years and were representative of the most advanced high-school students in the Netherlands.

The impairing effect of reductions on perception (a 'reduced-form disadvantage') was robust, as it appeared across the two tasks, a phrase-intelligibility task and an auditory lexical decision task. Adolescents presented with reduced speech ('Reduced condition') had lower accuracies and, in the phrase-intelligibility task, also lower intelligibility judgments and more errors of several types (deletions, false segmentations and non-word and word substitutions) than those presented with unreduced speech ('Unreduced condition'). Only one measure (the reaction times in the auditory lexical decision task) did not provide clear evidence of the disadvantage. Hence, it remains unclear whether processing is slower (or faster) for reduced than unreduced isolated (pseudo-)verbs. It is not uncommon that reaction times in this task yield less straightforward results than accuracies and this may be due to subtle factors influencing processing time, such as the relative frequencies of variants (Connine et al., 2008; Tucker, 2011; Brand & Ernestus, 2018).

4.2. The problems signal a lack of experience with reduced variants

In line with previous work (Ernestus et al., 2002; Ranbom & Connine, 2007; Tucker & Warner, 2007; Wanrooij & Raijmakers, 2020), a reduced-form disadvantage also appeared in the control group of German L1 listeners (after all, listeners have fewer cues to rely on; Introduction 1.1), but it was larger for the L2 listeners (Fig. 1). Therefore, the disadvantage in the L2 listeners must be based on more than only a general (i.e., not specific to L2 listeners) larger listening challenge for reduced than unreduced speech. We believe that an important factor is a lack of experience with reduced variants. This interpretation is in line with the other study in our project, in which the same tasks were used (Wanrooij & Raijmakers, 2020). In this previous study, adult L1 listeners ('more' experience) outperform adolescent L1 listeners ('less' experience) in L1 (German) reduced-speech perception; in the present study, the adolescent L1 listeners outperform the adolescent L2 listeners (even less experience with German reduced variants). Notice that the results for unreduced speech perception in this paper also match the experience account: scores were better for L1 than L2 listeners (L2 listeners have less experience with the unreduced variants too), but the difference is smaller than for reduced speech (L2 listeners have even less experience with reduced variants).

In view of the above, our results support the idea that experience with 'variants' is necessary for robust linguistic category formation (Introduction 1.3.2). This idea has been proposed earlier for speech sound categories in the L1 (Kuhl et al., 1997; McMurray et al., 2018) and L2 (Logan et al., 1991; Wanrooij et al., 2013) and also for word categories in the L1 (McMurray et al., 2018). Variability in these studies was based on differences in voice or was defined broadly. The previous and present study in our project suggest that experience with variability in reduction patterns specifically also adds to more robust linguistic categories.

4.3. At least bottom-up processing is inadequate

This study also explored the mechanisms behind the reduced-form disadvantage in adolescent L2 learners, in the phrase-intelligibility task. We reasoned that reductions could cause inadequate bottom-up and top-down processing (Introduction 1.3.3) but found evidence particularly for inadequate bottom-up processing. As mentioned, the accuracies and intelligibility judgments for reduced speech were lower than for unreduced speech across the task. Relevant here is that they were lower in the part-phrases, where possibilities for top-down processing were minimized and the need for skilled bottom-up processing was maximized. This held particularly for the *reduced* part-phrases, where the number of available cues was the smallest in the task. Indeed, the accuracies for reduced part-phrases were the lowest. Additionally, non-word substitutions (which represent incorrect sound to word mappings and hence could be considered signs of inadequate bottom-up processing) were more numerous for listeners to reduced than unreduced speech across the task. The same held for all other error types (which we reasoned to represent at least inadequate bottom-up processing; Introduction 1.3.3). We interpret these signals of inadequate bottom-up processing in line with the results of the related previous study in our project (just mentioned above in Discussion 4.2; Wanrooij & Raijmakers, 2020), in which the continuing development of adolescents' native speech perception was related to an increasing sensitivity to acoustic–phonetic details of speech sounds and, concomitantly, "richer" speech sound representations, thereby following McMurray et al. (2018). McMurray et al. observed an increasing sensitivity to within-phoneme-category differences and hypothesized that this increasing sensitivity allows listeners to handle variability in the speech signal better. Since sensitivity to acoustic–phonetic detail and the ability to handle variability in the speech signal are key to adequate reduced-speech perception (Introduction 1.1), and since native listeners apparently need a long time to develop these skills in full (the previous study), we hypothesize that problems in non-native bottom-up processing of reduced speech must arise at least at this level of analysis (analysis of acoustic–phonetic information).

The evidence that reduced speech also hampers top-down processing was less clear. L2 learners in both conditions (i.e., irrespective of whether they heard reduced or unreduced speech) improved their accuracies when after a part-phrase they heard the corresponding full phrase. The extent of this improvement was not significantly different between the conditions, and not even between the L1 and L2 listeners. This shows that adolescent listeners can make use of top-down processing (in our case: the use of cues in the phrasal context provided in the full phrase, to infer the target words in the part-phrase) irrespective of the speaking style (reduced or unreduced) and the language (native of non-native). Interestingly, the literature on adults' perception of speech in noise also shows that L2 listeners are well capable of using top-down processing, in this case of using semantic cues to predict an upcoming word (Bradlow & Alexander, 2007). However, in Bradlow and Alexander, the listeners needed a clearer speech signal, with less noise, for this than adult L1 listeners and the present study does not yield comparable evidence (i.e., that L2 listeners need a clearer speech signal, without reductions, for using top-down processing than L1 listeners). All in all, the comprehension problems that adolescent L2 learners in high school face with reduced speech seem to be related to at least inadequate bottom-up processing. Of course, inadequate bottom-up processing also reduces the chances of accurate top-down processing. It is possible, therefore, that we missed obstructions in top-down processing because they were more difficult to pinpoint.

4.4. Consequences for the classroom

Although L2 learners showed more non-word substitutions and errors of other types in response to reduced than unreduced speech in the phrase-intelligibility task, these error types also occurred when listening to the latter speaking style. In fact, *every* Dutch adolescent used non-word substitutions and non-word substitutions outnumbered word substitutions in both conditions. In addition, every adolescent used word substitutions, and virtually every adolescent mixed substitution types within answers, sometimes even mixing non-word, German word and Dutch word substitutions. Hence, adolescents in both conditions resorted to every means possible to make sense of the signal before giving up (as indicated by deletions). Overall, the errors match the profile of beginning, low-proficiency L2 learners, namely a profile characterized by substantial breakdowns in bottom-up processing and attempts to compensate for this via top-down processing (Voss, 1984; Field, 2004). The profile shows that Dutch high-school training in German is not sufficient for making students understand even simple natural speech. It should be remembered (Method 2.1) that the L2 learners in the present study were representative of high-school students that attend the 'highest' level of high-school education in the Netherlands (namely the level preparing students for university) and had had the longest training (four years).

In view of the above, the next question is how to improve high-school students' L2 listening comprehension. There are numerous studies on teaching L2 listening comprehension and training *in general* (i.e., not specifically of reduced speech). Since the 1990s, the focus has been increasingly on teaching learners to use 'top-down' strategies (Goh, 2008), such as preparing for the word 'nail' when perceiving 'hammer'. Although such strategies can definitely be effective (Goh, 2008), they do not necessarily solve the problems L2 learners have with reductions (for instance, it is useless to prepare for 'nail' when misperceiving *hama* in the meaning *haben wir* as *Hammer*, 'hammer'; see also Field, 2003; Hulstijn, 2003). The results in the present study endorse this objection and suggest that for learning to understand real-life speech, bottom-up processing should be improved in the first place (Discussion 4.3). Although future work is needed to find out what training would be best for this, the results in this study suggest that building up experience with reduced variants is a prerequisite (Discussion 4.2).

5. Conclusion

This study shows that adolescents who have learned a second language in high school for four years at the most advanced level can still not understand simple phrases in the second language when these phrases are pronounced as in natural conversations, that is, with reduced pronunciations. Problems arise in any case in bottom-up processing. Extensive experience with naturally reduced speech variants seems needed to strengthen linguistic (word) representations.

Declaration of interest

This work was supported by the Netherlands Initiative for Education Research (Dutch acronym NRO; grant number 405-16-382 awarded to the first author). The NRO was not involved in any aspect of the study.

Acknowledgements

This work was supported by the Netherlands Initiative for Education Research (Dutch acronym NRO; grant number 405-16-382 awarded to the first author). We thank Jos Pacilly, Malte Steinhoff and Mané van Veldhuizen for their indispensable assistance with technical matters and participant recruitment, and Angela Bak, Manon Willems and other student assistants for their help with participant recruitment and testing. We are also grateful to professor Marc Brysbaert and the anonymous reviewers for their valuable comments, which helped improve the paper. Finally, we thank all high-school students who participated in the study, as well as their parents, teachers and school administrators for their time, efforts, interest and enthusiasm.

Appendix A. Stimuli in the phrase-intelligibility task

Table A.1

Table A.1
Stimuli in the phrase-intelligibility task.

Nr	Stimulus	Phonetic transcription of targets ([. . .] = non-targets)		Number of targets
		Reduced	Unreduced	
	Subtask 1			
1	damit einverstanden	[. . .] a̰ɪnfɐʃtann(n)	[. . .] ʔaɪnfɐʃtandən	1
2	gegen 14, 15 Uhr	geːŋ [. . .]	geːgən [. . .]	1
3	bei dieser Gelegenheit	[. . .] gəleːgŋhaɪt	[. . .] gəleːgənhaɪt	1
4	wenn Ihnen das nichts ausmacht	[. . .] iːn [. . .]	[. . .] ʔiːnən [. . .]	1
5	zu Ihrem Büro	[. . .] ḭːɐm [. . .]	[. . .] ʔiːrəm [. . .]	1
6	gleich morgens	[. . .] mɔɐgŋs	[. . .] mɔɐgəns	1
7	im selben Monat	[. . .] zɛlm(m) [. . .]	[. . .] zɛlbən [. . .]	1
8	Wiederhören	viːdɐhøːɐn	viːdɐhøːrən	1
9	kleinen Augenblick	klaɪn a̰ʊŋblɪk	klaɪnən ʔaʊgənblɪk	2
10	nach dem Essen	[. . .] m ɛ̰sn	[. . .] deːm ʔɛsən	2
11	einen halben Tag	n halm(m) [. . .]	ʔaɪnən halbən [. . .]	2
12	ganz hervorragend passen	[. . .] hɛɐfoːɐraːŋ pasn	[. . .] hɛɐfoːɐraːgənt pasən	2
13	in meinem Kalender	[. . .] maɪm khalɛn(n)ɐ	[. . .] maɪnəm kalɛndɐ	2
14	die letzten beiden Tage	[. . .] lɛtstn baɪn(n) [. . .]	[. . .] lɛtstən baɪdən [. . .]	2
15	nehmen wir das	neːm ʋiːɐ [. . .]	neːmən viːɐ [. . .]	2
16	offenbar ganz was Spannendes	ɔfmaːɐ [. . .] ʃpan(n)dəs	ʔɔfənbaːɐ [. . .] ʃpanəndəs	2
17	schönen guten Tag	ʃøːŋ gun [. . .]	ʃøːnən guːtən [. . .]	2
18	wegen der Uhrzeit	veːŋ dɐ [. . .]	veːgən deːɐ [. . .]	2
19	an einem Wochenende	[. . .] a̰ɪm vɔxŋɛ̰n(n)ə	[. . .] ʔaɪnəm vɔxənʔɛndə	2
20	mit anderen Worten	[. . .] a̰nɐn vɔɐ̰n	[. . .] ʔandɐrən vɔɐtən	2
21	noch mal folgendes überlegen	nɔ̃ maː fɔlŋəs yːbɐleːgŋ	nɔx maːl fɔlgəndəs ʔyːbɐleːgən	3*
22	mal eben sehen	maː ḛːbm zeːn	maːl ʔeːbən zeːən	3
23	einen gemütlichen Abend verbringen	a̰ɪŋ gəmyːtlɪçn a̰ːmp fɐbrɪŋŋ	ʔaɪnən gəmyːtlɪçən ʔaːbənt fɐbrɪŋən	4
24	die wichtigen noch offenen Fragen klären	[. . .] vɪçtɪgŋ [. . .] ɔ̰f(ə)nən fraːgŋ khlɛːɐn	[. . .] vɪçtɪgən [. . .] ʔɔfənən fraːgən klɛːrən	4
	Subtask 2			
25	(donnerstags sieht es) allerdings (schlecht aus)	a̰lɐɪŋs	ʔalɐdɪŋs	1
26	(immer) ausserordentlich (gut)	a̰ʊsɐɔ̰ɐəntlɪç	ʔaʊsɐʔɔɐdəntlɪç	1
27	(ich bin) eigentlich (gar nicht da)	aɪnɪ	ʔaɪgəntlɪç	1
28	entschuldigen (Sie bitte)	ʔɛnʃʊlɪŋ	ʔɛntʃʊldɪgən	1
29	(zuviel Arbeit zu) erledigen	ɛ̰ɐleːdɪŋ	ʔɛɐleːdɪgən	1
30	(ich könnte) höchstens (nach den Feiertagen)	høːstns	høːçstəns	1
31	(Anfang November) irgendwann (oder am Montag)	ʔɪ̰ɐgŋvan	ʔɪɐgəntvan	1
32	auf jeden Fall (haben wir ja die Einladung)	[. . .] jeːn [. . .]	[. . .] jeːdən [. . .]	1
33	(noch genauer) verabreden	fɐa̰preːdn	fɐʔapreːdən	1
34	(okay) vereinbaren (wir das)	fɐa̰ɪnmaːɐn	fɐʔaɪnbaːrən	1
35	vielleicht (wieder 8 Uhr)	flaɪt	fɪlaɪçt	1
36	ab dem sechzehnten (bis Weihnachten)	[. . .] zɛçtseːʔn	[. . .] zɛçtseːntən	1
37	(ich fände es) wirklich (toll)	vɪɐkɪç	vɪɐklɪç	1
38	zusammen (Kaffee trinken)	tsʊzam(m)	tsʊzamən	1
39	(das ist) im Grunde genommen (keine schlechte Idee)	[. . .] grʊn(n)ə gənɔmm	[. . .] grʊndə gənɔmən	2
40	dann haben wir (das auch gleich mit eingetragen)	haːmɐ	haːbən viːɐ	2
41	(da haben Sie) natürlich nicht (ganz unrecht)	natyːɐ(lɪ)ç nɪç	natyːɐlɪç nɪçt	2
42	(ich denke dass das) mit sieben Tagen (auch nicht hinkommen wird)	[. . .] ziːm taːŋ	[. . .] ziːbən taːgən	2
43	zum Beispiel (vom 3ten bis zum 11ten Februar)	tsʊ̃ maɪʃl	tsʊm baɪʃpiːl	2
44	da habe ich (nicht dran gedacht)	daːb ɪ̰h	daː haːbə ʔɪç	3
45	das können wir (also machen)	s kœm miːɐ	das kœnən viːɐ	3
46	dann sind wir uns (da doch noch einig geworden)	[. . .] zɪm ʋ ʊns	[. . .] zɪnt viːɐ ʔʊns	3
47	(das wird dann) wahrscheinlich ein bißchen (schwierig)	vaʃaɪn(l) m bɪsçn	vaːɐʃaɪnlɪç ʔaɪn bɪsçən	3
48	zwei Stunden werden wir (brauchen)	[. . .] ʃtʊnʔn ʋeːɐm ʋɐ	[. . .] ʃtʊndən veːɐdən viːɐ	3

* : The word combination *noch mal* in the fragment *noch mal folgendes überlegen* was viewed as one target word (Wanrooij & Raijmakers, 2020).

Appendix B. Stimuli in the auditory lexical decision task

Table B.1

Table B.1
Stimuli in the auditory lexical decision task. Each stimulus existed in a reduced version and an unreduced version.

Nr	Verbs	Pseudo-verbs
	2 syllables	
1	bleiben	bleffen
2	brauchen	brufen
3	fassen	fagen
4	fehlen	falchen
5	finden	fauben
6	fragen	frocken
7	geben	gachen
8	gehen	gelchen
9	halten	hehen
10	helfen	hinsen
11	klappen	kaffen
12	kucken	klanen
13	legen	lalen
14	machen	muchen
15	melden	mählen
16	planen	plappen
17	rufen	relen
18	sagen	saffen
19	schaffen	scheigen
20	treffen	trehen
21	wollen	welben
22	wählen	weten
	3 syllables	
23	anbieten	ankongen
24	anfangen	anschreichen
25	aufschreiben	aufreigen
26	ausreichen	ausfaggen
27	ausschlafen	aussuden
28	beeilen	bekrielen
29	beenden	benüten
30	bekommen	beschlaren
31	besprechen	bezichten
32	eintragen	einfagen
33	genügen	gesteden
34	hinkriegen	hinschlaben
35	losfahren	losschiegen
36	verschieben	verbieben
37	verstehen	vereigen
38	versuchen	verspreffen
39	verzichten	vertramen
40	vorschlagen	vorenden

Appendix C. The number of errors of each type for Dutch adolescents in the phrase-intelligibility task

Table C.1

Table C.1
The number of Dutch high-school students' answers with deletions, substitutions and false segmentations in the phrase-intelligibility task: medians and ranges per Condition (Reduced and Unreduced). The answers considered are answers A (see text).

Error type		Reduced		Unreduced	
		Median	Range	Median	Range
Deletions		7	1~20	1	0~8
Substitutions by	- non-words	15.5	6~21	8	1~13
	- German words	14.5	8~20	4	2~7
	- Dutch words	2	0~5	1	0~3
False segmentations		11	7~18	2	1~4

Appendix D. Accuracies and reaction times in the auditory lexical decision task

Table D.1

Table D.1
Accuracies (top) and reaction times (bottom) in the auditory lexical decision task.* In response to Words (left) and Pseudo-words (right), per Language x Condition: means and 95% confidence intervals (*CI*s). L1 = German native listeners, L2 = Dutch non-native listeners. Reaction times (in milliseconds) are log-transformed, with non-transformed equivalents between brackets. Maximum accuracy is 40 for Words and 40 for Pseudo-words.

Language	Condition	Words		Pseudo-words	
		Mean	*CI*	Mean	*CI*
Accuracies					
L1	Reduced	36.83	35.70 ~ 37.96	29.56	27.53 ~ 31.59
	Unreduced	37.19	35.99 ~ 38.39	34.00	31.85 ~ 36.15
L2	Reduced	29.77	28.60 ~ 30.93	24.77	22.68 ~ 26.85
	Unreduced	31.00	29.90 ~ 32.10	29.58	27.60 ~ 31.56
Reaction Times					
L1	Reduced	2.93 (847.23)	2.91 ~ 2.95	3.03 (1069.05)	3.01 ~ 3.05
	Unreduced	2.94 (870.96)	2.92 ~ 2.96	3.01 (1030.39)	2.99 ~ 3.04
L2	Reduced	3.00 (1011.58)	2.99 ~ 3.02	3.10 (1264.74)	3.08 ~ 3.13
	Unreduced	2.97 (937.56)	2.96 ~ 2.99	3.07 (1172.20)	3.05 ~ 3.09

* : Calculations are based on mean accuracies per participant.

Appendix E. Mixed-effects models for the analysis of the data in the auditory lexical decision task

The models were defined in the program R (R Core Team, 2020) with the package *lme4* (Bates et al., 2015). The accuracies and log-transformed reaction times were analysed separately.

1. Model for accuracies

The accuracies (dependent variable) were analysed with the following generalized linear mixed-effects model (using the *lme4*-function *glmer*):

E.1.
```
model_full = glmer(accuracy ~ ItemType * Language * Condition
                  + (1|participant) + (1|item),
                  data = data_accuracies, family = binomial,
                  control = glmerControl(optimizer = "bobyqa",
                  optCtrl = list(maxfun = 2e5))).
```

Hence, the model included three main fixed effects: Item Type (Word versus Pseudo-word), Language (L1 = German native listeners versus L2 = Dutch non-native listeners) and Condition (Reduced versus Unreduced). Additionally, the model included all interaction effects: Item Type × Language, Item Type × Condition, Language × Condition and Item Type × Language × Condition. Participant and Item were modelled as random intercepts.

2. Models for reaction times

The log-transformed reaction times were analysed with linear mixed-effect models (using the *lme4*-function *lmer*). Fixed and random effects were the same as specified for the accuracies-model above. The *lmer*-function does not provide the *p*-values for the factors entered in a model (Bates et al., 2015). Significance values can be obtained by comparing the likelihood of two models, one with and the other without a certain factor (Bates et al., 2015; Winter, 2013; Baayen et al., 2008). For significance testing, we therefore used a sequence of models, starting with the null model and ending with the full model (both given below). Factors were added in the order in which they are listed in the full model. This order reflects the estimated order of importance (therefore, Item Type was added first: this factor should be significant for the task to be valid). We tested the significance of each factor by comparing the model with the factor (*modelY*) to the previous model without the factor (*modelX*). For this, we used the anova-function, i.e., *anova(modelX, modelY)*. The beta estimates and the standard errors reported in Table 10 are those of the full model.

E.2.
```
model_null = lmer(rtlog ~ 1 + (1|participant) + (1|item),
                 data = data_reactiontimes, REML = FALSE)
```

E.3.
```
model_full = lmer(rtlog ~ ItemType + Language + Condition
                 + ItemType:Language + ItemType:Condition
                 + Language:Condition + ItemType:Language:Condition
                 + (1|participant) + (1|item),
                 data = data_reactiontimes, REML = FALSE)
```

Appendix F. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wocn.2021.101082.

References

Ahn, S.-W. (1987). *Sandhi-variation and affective factors as input filters to comprehension of spoken English among Korean learners* Doctoral dissertation. Austin, Texas (USA): The University of Texas at Austin.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Blank, M. A. (1979). *Dual-mode processing of phonemes in fluent speech* Doctoral dissertation. Austin, Texas (USA): The University of Texas at Austin.

Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart & Winston.

Boersma, P. (2011). A programme for bidirectional phonology and phonetics and their acquisition and evolution. In A. Benz & J. Mattausch (Eds.), *Bidirectional optimality theory* (pp. 33–72). Amsterdam: John Benjamins.

Boersma, P., & Weenink, D. (1998–2019). Praat: Doing Phonetics by Computer [Computer program]. Version 6.0.23. Retrieved January 19, 2017 from http://www.praat.org/.

Bowen, J. D. (1975). *Patterns of English Pronunciation*. Rowley, Mass.: Newbury House.

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America, 121*(4), 2339–2349.

Brand, S., & Ernestus, M. (2018). Listeners' processing of a given reduced word pronunciation variant directly reflects their exposure to this variant: Evidence from native listeners and learners of French. *Quarterly Journal of Experimental Psychology, 71*(5), 1240–1259. https://doi.org/10.1080/17470218.2017.1313282.

Brown, J. D., & Hilferty, A. (1986). The effectiveness of teaching reduced forms of listening comprehension. *RELC Journal, 17*(2), 59–70.

Connine, C. M., Ranbom, L. J., & Patterson, D. J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics, 70*(3), 403–411.

Cutler, A. (2000). Listening to a second language through the ears of a first. *Interpreting, 5*(1), 1–23. https://doi.org/10.1075/intp10.1075/intp.5.110.1075/intp.5.1.02cut.

Darcy, I., Ramus, F., Christophe, A., Kinzler, K., & Dupoux, E. (2009). Phonological knowledge in compensation for native and non-native assimilation. In F. Kügler, C. Féry, & R. van de Vijver (Eds.), *Variation and gradience in phonetics and phonology* (pp. 265–309). Berlin, Germany: Mouton De Gruyter.

Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch. A corpus-based study of the phonology-phonetics interface* Doctoral dissertation. Utrecht: LOT.

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language, 81*(1–3), 162–173. https://doi.org/10.1006/brln.2001.2514.

Ernestus, M., & Cutler, A. (2015). BALDEY: A database of auditory lexical decisions. *The Quarterly Journal of Experimental Psychology, 68*(8), 1469–1488. https://doi.org/10.1080/17470218.2014.984730.

Ernestus, M., Dikmans, M. E., & Giezenaar, G. (2017). Advanced second language learners experience difficulties processing reduced word pronunciation variants. *Dutch Journal of Applied Linguistics, 6*(1), 1–20. https://doi.org/10.1075/dujal.6.1.01ern.

Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics, 39*, 253–260. https://doi.org/10.1016/S0095-4470(11)00055-6.

Escudero, P., & Wanrooij, K. (2010). The effect of L1 orthography on non-native vowel perception. *Language and Speech, 53*(3), 343–365. https://doi.org/10.1177/0023830910371447.

Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT (English Language Teaching) Journal, 57*(4), 325–334. https://doi.org/10.1093/elt/57.4.325.

Field, J. (2004). An insight into listeners' problems: Too much bottom-up or too much top-down? *System, 32*(3), 363–377. https://doi.org/10.1016/j.system.2004.05.002.

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America, 106*(5), 2973–2987.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*(1), 110–125 https://doi-org.proxy.library.uu.nl/10.1037//0096-1523.6.1.110.

Goh, C. C. M. (2008). Metacognitive instruction for second language listening development theory, practice and research implications. *RELC Journal, 39*(2), 188–213. https://doi.org/10.1177/0033688208092184.

Hasegawa, N. (1979). Casual speech vs. fast speech. *CLS (Chicago Linguistics Society), 15*, 126–137.

Hasegawa, N. (2006). On casual speech: how it differs from fast speech. In: Brown, James Dean & Kimi Kondo-Brown (Eds.), Perspectives on teaching connected speech to second language speakers. Honolulu, HI: University of Hawai'i at Manoa, National Foreign Language Resource Center.

Henrichsen, L. E. (1984). Sandhi variation: A filter of input for learners of ESL. *Language Learning, 34*(3), 103–126. https://doi.org/10.1111/j.1467-1770.1984.tb00343.x.

Holley-Wilcox, P. (1977). The effect of homophony with auditory presentation of stimuli. Midwestern Psychological Association, Chicago, United States of America, May (cited in McCusker et al., 1981).

Hulstijn, J. H. (2003). Connectionist Models of Language Processing and the Training of Listening Skills With the Aid of Multimedia Software. Computer Assisted Language Learning, 16(5), 413–425. https://doi.org/10.1076/call.16.5.413.29488.

IPDS (1996). *The Kiel Corpus of spontaneous speech* (Vol. 2, CD-ROM#3) Kie: IPDS.

IPDS (1997). *The Kiel Corpus of spontaneous speech* (Vol. 3, CD-ROM#4) Kiel: IPDS.

IPDS (Institut für Phonetik und digitale Sprachverarbeitung der Christian-Albrechts-Universität zu Kiel). 1995. The Kiel Corpus of spontaneous speech,Vol.1, CD-ROM#2. Kiel: IPDS.

Ito, Y. (2006). The comprehension of English reduced forms by second language learners and its effect on input-intake process. In: James Dean Brown & Kimi Kondo-Brown (Eds.), Perspectives on teaching connected speech to second language speakers (pp.67–81). Honolulu, HI (USA): University of Hawai'i, National Foreign Language Resource Center.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446.

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

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.), Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium, 29–54. Tokyo, Japan: The National International Institute for Japanese Language.

Kohler, K. J. (1998). The disappearance of words in connected speech. *ZAS Working Papers Linguistics, 11*, 21–34.

Kohler, K. J. (2001). Articulatory dynamics of vowels and consonants in speech communication. *Journal of the International Phonetic Association, 31*(1), 1–16. https://doi.org/10.1017/S0025100301001013.

Kuhl, P., Andruski, J., Chistovich, I., Chistovich, L., Kozhevnikova, E., Ryskina, V., et al. (2017). Cross-language analysis of phonetic units in language addressed to infants. *Science, 227*(5326), 684–686. https://doi.org/10.1126/science.277.5326.684.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify /r/ and /l/: A first report. *The Journal of the Acoustical Society of America, 89*(2), 874–886. https://doi.org/10.1121/1.1894649.

Maekawa, K. & Kikuchi, H. (2005). Corpus-based analysis of vowel devoicing in spontaneous Japanese. An interim report. In J. van de Weijer, K. Nanjo & T. Nishihara (Eds.), Voicing in Japanese, 205–228. Berlin: Mouton de Gruyter.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0.

McCusker, L. X., Holley-Wilcox, P. & Hillinger, M. L. (1979). Frequency effects in auditory and visual word recognition. Southwestern Psychological Association, San Antonio, Texas, United States of America, April (cited in McCusker et al., 1981).

McMurray, B., Danelz, A., Rigler, H., & Seedorff, M. (2018). Speech categorization develops slowly through adolescence. *Developmental Psychology, 54*(8), 1472–1491. https://doi.org/10.1037/dev0000542.

Mitterer, H. (2011). Recognizing reduced forms: Different processing mechanisms for similar reductions. *Journal of Phonetics, 39*(3), 298–303.

Mitterer, H., & Tuinman, A. (2012). The role of native-language knowledge in the perception of casual speech in a second language. *Frontiers in Psychology, 3*, 1–13. Article 249.

Morano, L., Ernestus, M., & ten Bosch, L. (2015). Schwa reduction in low-proficiency L2 speakers: learning and generalization. In M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, & J. Scobbie (Eds.), Proceedings of the 18th international congress of phonetic sciences (ICPhS 2015). Glasgow: University of Glasgow.

Norris, R. W. (1995). Teaching reduced forms: Putting the horse before the cart. *English Teaching Forum, 33*, 47–50.

Pitt, M. A., Dilley, L., & Tat, M. (2011). Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics, 39*(3), 304–311.

Poelmans, P. (2003). *Developing second-language listening comprehension: Effects of training lower-order skills versus higher-order strategy* Doctoral dissertation. Utrecht, the Netherlands: LOT.

Polivanov, E.D. (1931). La perception des sons d' une langue étrangère. Travaux du Cercle Linguistique de Prague, 4, 79–96. English Translation: The subjective nature of the perceptions of language sounds. In: E.D. Polivanov (1974). Selected works: articles on general linguistics. The Hague: Mouton, 223–237.

R Core Team (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ranbom, L, & Connine, C (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language, 57*(2), 273–298.

Rastle, K., McCormick, S. F., Bayliss, L., & Davis, C. J. (2011). Orthography influences the perception and production of speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(6), 1588–1594. https://doi.org/10.1037/a0024833.

Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 209–227). Amsterdam: Benjamins.

Rosenthal, R. (1991). *Meta-analytic procedures for social research (revised)*. Newbury Park, CA: Sage Publications.

Shockey, L. (2003). *Sound patterns of Spoken English*. Cambridge: Blackwell.

Studdert-Kennedy, M. (1974). The perception of speech. In T.A. Sebeok (Ed.). Current trends in linguistics (pp. 2349–2385). Research Center for the Language Sciences, Indiana University. The Hague: Mouton & CO.

ten Bosch, L., Giezenaar, G., Boves, L. & Ernestus, M. (2016). Modeling language-learners' errors in understanding casual speech. In G. Adda, V. Barbu Mititelu, J. Mariani, D. Tufiş & I. Vasilescu (eds), Errors by humans and machines in multimedia, multimodal, multilingual data processing. Proceedings of Errare 2015, 107–121. Bucharest: Editura Academiei Române.

Torreira, F., & Ernestus, M. (2011). Realization of voiceless stops and vowels in conversational French and Spanish. *Laboratory Phonology, 2*, 331–353. https://doi.org/10.1515/labphon.2011.012.

Tsui, A. B. M., & Fullilove, J. (1998). Bottom-up or top-down processing as a discriminator of L2 listening performance. *Applied Linguistics, 19*(4), 432–451. https://doi.org/10.1093/applin/19.4.432.

Tucker, B. V. (2011). The effect of reduction on the processing of flaps and /g/ in isolated words. *Journal of Phonetics, 39*(3), 312–318.

Tucker, B. V., & Warner, N. (2007). Inhibition of processing due to reduction of the American English flap. In *Inhibition of processing due to reduction of the American English flap* (pp. 1949–1952). Saarbruecken.

van de Ven, M., Tucker, B. V., & Ernestus, M. (2011). Semantic context effects in the comprehension of reduced pronunciation variants. *Memory and Cognition, 39*(7), 1301–1316.

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching, 40*(3), 191–210. https://doi.org/10.1017/S0261444807004338.

Voss, B. (1984). *Slips of the ear*. Tübingen: Gunter Narr Verlag.

Wanrooij, K., Escudero, P., & Raijmakers, M. E. J. (2013). What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *Journal of Phonetics, 41*(5), 307–319. https://doi.org/10.1016/j.wocn.2013.03.005.

Wanrooij, K., & Raijmakers, M. E. J. (2020). Evidence for immature perception in adolescents: Adults process reduced speech better and faster than 16-year olds. *Language Acquisition, 27*(4), 434–459. https://doi.org/10.1080/10489223.2020.1769627.

Wanrooij, K. (2019). Luistervaardigheidstraining Duits op Nederlandse middelbare scholen. Uitkomsten van een vragenlijst ingevuld door docenten Duits in 2018. (The practice of teaching German in Dutch high schools. Results obtained in a questionnaire filled in by teachers of German throughout the Netherlands). https://www.karinwanrooij.nl/wp-content/uploads/2019/06/VerslagVragenlijst_VaardigVerstaan.pdf. 51 pages (in Dutch)

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499. [http://arxiv.org/pdf/1308.5499.pdf].

Wong, S. W. L., Mok, P. P. K., Chung, K. K.-H., Leung, V. W. H., Bishop, D. V. M., & Chow, B. W.-Y. (2015). Perception of native English reduced forms in Chinese learners: Its role in listening comprehension and its phonological correlates. *TESOL Quarterly, 51*(1), 7–31.