




Article

Non-Native Listeners' Use of Information in Parsing Ambiguous Casual Speech

Natasha Warner ^{1,*} , Daniel Brenner ¹, Benjamin V. Tucker ²  and Mirjam Ernestus ³ ¹ Department of Linguistics, University of Arizona, Tucson, AZ 85721, USA; wobaidan@gmail.com² Department of Communication Sciences and Disorders, Northern Arizona University, Flagstaff, AZ 86011, USA; benjamin.tucker@nau.edu³ Centre for Language Studies, Radboud University, 6500 HD Nijmegen, The Netherlands; mirjam.ernestus@ru.nl

* Correspondence: nwarner@arizona.edu; Tel.: +1-520-626-5591

Abstract: During conversation, speakers produce reduced speech, and this can create homophones: 'we were' and 'we're' can both be realized as [ɜ], and 'he was' and 'he's' can be realized as [ɪz]. We investigate the types of information non-native listeners (Dutch L1-English L2) use to perceive the tense of such verbs, making comparisons with previous results from native listeners. The Dutch listeners were almost as successful as natives (average percentage correct for 'is'/'was' in the most accurate condition: 81% for Dutch, 88% for natives). The two groups showed many of the same patterns, indicating that both make strong use of whatever acoustic information is available in the signal, even if it is heavily reduced. The Dutch listeners showed one crucial difference: a minimal amount of context around the target, just enough to signal speech rate, did not help Dutch listeners to recover the longer forms, i.e., was/were, from reduced pronunciations. Only the full utterance context (containing syntactic/semantic information such as 'yesterday' or another tensed verb) helped Dutch listeners to recover from reduction. They were not able to adjust their criteria based on the surrounding speech rate as native listeners were. This study contributes to understanding how L2 learners parse information from spontaneous speech in a World Englishes setting with inputs from multiple dialects.



Academic Editors: Donna Erickson and Adrian P. Simpson

Received: 1 August 2024

Revised: 21 December 2024

Accepted: 31 December 2024

Published: 8 January 2025

Citation: Warner, N., Brenner, D., Tucker, B. V., & Ernestus, M. (2025). Non-Native Listeners' Use of Information in Parsing Ambiguous Casual Speech. *Languages*, 10(1), 8. <https://doi.org/10.3390/languages10010008>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: speech reduction; speech perception; L2 listening; acoustic reduction

1. Introduction

In spontaneous conversation, speakers often produce highly reduced speech, omitting or altering sounds, syllables, and words relative to careful pronunciation of the same words (Greenberg, 1999; Johnson, 2004; Ernestus & Warner, 2011). This paper examines non-native listeners' perception of reduced speech. For example, in recordings of spontaneous speech, we found pronunciations such as [ʃɪzɪ] for 'she wants to be a' or [əʒ lɪk] for 'I was (just) like' (see audio examples at <https://sites.arizona.edu/nwarner/reduced-speech-examples/>, accessed on 31 July 2024). Such reductions can affect content words (e.g., [fɪɛ:] for 'Friday night' in one of our recordings) as well as function words, but are more common in high-frequency function words such as 'I was', 'we were like', etc. Such reductions occur, as far as we know, in every language where someone has looked for them (Ernestus & Warner, 2011, see also audio examples at <https://nascl.rc.nau.edu/resources/reduction-examples/>, accessed on 31 July 2024). Reduced speech usually poses no problem for listeners, as long as they hear the speech in context and it is in their native language.

Listeners do not even detect anything unusual about the speech when listening to a recording of a spontaneous conversation.

This paper examines what information listeners use to comprehend such speech when they are hearing a language that is not their native one. We specifically investigate how listeners comprehend reduced speech when they are rather proficient in their L2 and live in a country where that L2 is used often but in more than one dialect. There are at least three types of information available to listeners to use when comprehending reduced words within the span of an utterance: the acoustics in the word itself, the speech rate of the surrounding speech, and the syntactic and semantic information in the rest of the words and sentence structure of the utterance. For example, when a listener hears an extremely reduced token of 'Friday night' like the one mentioned above (realized as [fɪɛ:]), the acoustics of that stretch of speech itself, such as short duration and nasalization, might lead the listener to think they are hearing the word 'friend'. The speech rate of the surrounding utterance, if it is fast, might lead the listener to think the duration of this stretch is too long to be just the word 'friend' in this fast speech rate, and so something may have been deleted, and it may be some longer word or words. If the rest of the utterance includes words such as 'but Saturday night', this may help the listener to recover the reduced words 'Friday night'. In conversational speech, there are additional sources of information, such as the topic of the discourse and coarticulation with neighboring sounds, but the three discussed here (acoustics within the word(s), speech rate, and syntactic/semantic information) are the ones of primary interest for the current work.

In some cases, reduced pronunciations can become homophonous: 'we were' and 'we're' can both be realized as just [ə], and 'he was' and 'he's' as [ɪz]. Thus, reduction can obscure the tense of the verb. In [Warner et al. \(2022\)](#), we investigated how native English listeners use various sources of information to identify the tense of such verbs (potentially homophonous sequences such as 'he's, he is, and he was', or 'we're, we are, and we were'). Possible sources of information include the acoustic cues in the reduced speech, speech rate, and the meaning of the rest of the utterance. We extracted stimuli such as 'I don't even know what we're gonna do' and 'I called Dad and asked him about the internet and like, he was like' from recordings of telephone conversations between friends or family members who were native speakers of American English. Underlined words constitute the target word/phrase. Following [Ernestus et al. \(2002\)](#), we presented the target phrases (e.g., 'we're' or 'he was', the underlined portion in the examples) to native listeners with various amounts of context ("isolation", "limited", or "full"). For example, for the utterance 'I don't even know what we're gonna do', in the Isolation condition listeners heard only 'we're' (whatever portion of the signal corresponded to 'we're', extracted from the original recording of spontaneous conversation). In the Limited-context condition, they heard the portion corresponding to '-at we're go-' (out to the edges of the surrounding vowels), which should provide information about the speech rate and coarticulation of the word-initial and -final consonants, but the portion of the neighboring words was, in most stimuli, not enough information to recognize those words. In the Full context condition, they heard the entire utterance, which often supplies semantic and/or syntactic information about the target's tense through other words of the utterance and tense of other verbs. For example, if the full utterance contains the word 'yesterday', it is more likely that the target verb is past tense. Listeners identified which tense they heard (e.g., 'we're/we are' vs. 'we were' as the response options). Results showed that the native English listeners prioritized acoustic cues above any syntactic or semantic information in the utterance context ([Warner et al., 2022](#)). They favored the acoustic cues even when those cues misled them because of extreme reduction. In the current study, we examine whether

rather proficient non-native listeners of English (native language is Dutch) use the same sources of information as native listeners.

When listeners (native or non-native) are hearing spontaneous speech, there are many types of information they might be using simultaneously to recognize the words they are hearing in the often fast, continuous, reduced speech stream. The most obvious type of information is the acoustic cues in the words themselves. For example, if a speech signal has low F2, that might provide perceptual information that suggests a /w/, leading the listener to perceive 'she was' rather than 'she's'. If a particular stretch of the signal that could be either 'they're' or 'they were' has relatively long duration, that might lead the listener to perceive 'they were' (the longer of the two forms, at two syllables). The Isolation condition (target word/phrase only, such as 'he's' or 'he was', presented in isolation) provides this type of information to listeners.

Another type of information is speech rate of the surrounding speech and any coarticulation present in the neighboring speech. If the speech rate of the surrounding words is slow, that might lead listeners to expect 'he's/he was' to also be produced slowly, in keeping with the surrounding speech rate. This could lead the listener to adjust the duration boundary between 'he's' and 'he was', so that a somewhat longer token can still be perceived as 'he's', because it is expected to be in slow speech. This is similar to listeners' use of speech rate information to adjust the boundary between aspirated and unaspirated stops at the segmental level (Miller & Volaitis, 1989; Volaitis & Miller, 1992). Past research has also shown that listeners use the speech rate of surrounding words to adjust the boundary for perception of shorter vs. longer vowels (Gottfried et al., 1990). Of particular relevance for perception of reduced speech are findings that listeners use speech rate of the surrounding context as a cue to whether the word 'or' is present or not in phrases like 'leisure time/leisure or time' (Dilley & Pitt, 2010). Niebuhr and Kohler (2011) demonstrate a similar case in German. The limited-context condition of our experiment, which allows listeners to hear only out to the edges of the surrounding vowels (e.g., '-at we're go-' from 'I don't even know what we're gonna do') is enough context to supply speech rate information and any coarticulation at the boundaries of the target, but not enough for listeners to recognize the surrounding words in most cases. In the case of the current experiment, since each target begins and ends with the same sounds regardless of verb tense (e.g., 'he's' and 'he was' begin and end with the same segments), coarticulation is unlikely to supply additional cues to the verb tense, but speech rate information could be helpful. In Warner et al. (2022), we found that native listeners are able to use the additional information in the Limited condition, likely speech rate information, to recover a highly reduced 'was' or 'were' from the signal. That is, a given stimulus can sound to native listeners like a token of 'he's' or 'we're' in the Isolation condition, but in the Limited context, fast surrounding speech rate leads the listener to shift the perceptual boundary between 'he was' and 'he's' to a shorter duration. Therefore, the same token is perceived as 'he was' or 'we were' in the Isolation condition. This was reflected in the results as an improvement in accuracy in native listeners' responses to past tense stimuli (was and were stimuli) in the Limited-context condition relative to the Isolation condition. This result is particularly similar to the Dilley and Pitt (2010) result and related subsequent research (e.g., Brown et al., 2012; Heffner et al., 2013), where increased surrounding speech rate shifted listeners' percept from 'leisure or time' to 'leisure time'.

A very different type of information comes from lexical, semantic, and syntactic information in the rest of the utterance. The Full-context condition of our experiment makes this information available to listeners. For example, in 'She was really hyper earlier', the adverb 'earlier' makes a present tense 'she is/she's' unlikely. Individual stimuli differ in how much such information is available. For example, in the stimulus 'But he's still s-(pause) you know', past tense 'he was' is also plausible. This is especially true when the verb

is followed by quotative 'like' as in "And he was like, 'what's wrong?!'", where the speaker could use the historical present to report past speech with no change to the meaning. Two control experiments in our past work (Warner et al., 2022) examined how much information native readers or listeners could get out of only the surrounding utterance information, without hearing the target word/phrase at all, in order to quantify how much syntactic and semantic information was potentially present and compare to conditions where listeners heard the target word/phrase. The results showed that native listeners made surprisingly little use of the syntactic and semantic information in the rest of the utterance, as long as they were able to hear the acoustic cues of the target phrase itself. Utterance context generally did not lead to significant improvements in the accuracy of tense perception, and listeners favored the acoustic cues in the target itself over the syntactic and semantic cues of the rest of the utterance when those two types of information conflicted.

There are additional types of information that listeners could use to understand speech. For example, hearing the entire preceding portion of the conversation provides more information about what topics are being discussed (e.g., this conversation is about wedding planning, or about the speaker's volunteer work with a Girl Scout troop), as well as a greater chance to acclimate to the speaker's voice and dialect. It might also provide more information about whether past events or future plans are being discussed, which could help with the perception of verb tense. The current work, however, as in (Warner et al., 2022), is limited to types of information within the utterance.

Warner et al. (2022) provide a review of the literature on native listeners' use of these various types of information, particularly for perceiving reduced, spontaneous speech. In addition to the work on speech rate as a perceptual cue discussed above, one important finding of the literature is Brouwer et al.'s (2012) finding that listeners seem to adjust their expectations for precision of match the lexicon when hearing spontaneous speech. Brouwer et al. found that listeners allowed more acoustic mismatch to the typical pronunciation of a word if they heard reduced speech before the target word. This suggests that the speech style of the surrounding speech also supplies information by leading the listener to expect other words to also be reduced.

van de Ven and colleagues, in several related works, found that native listeners do extract semantic and syntactic information from reduced speech utterances, but that they extract more information from the acoustics of the signal itself than from the syntactics and semantics (van de Ven et al., 2011, 2012; van de Ven & Ernestus, 2018). They also found that the semantic information in a reduced stimulus only primes recognition of a subsequent word if listeners have more time than usual for processing speech. Podlubny et al. (2018) investigated which specific types of variation in the sound wave provide the strongest acoustic cues in reduced speech (e.g., variously removing duration, pitch, or spectral information), and they compared these to syntactic/semantic contextual cues. Among other findings, this work showed that f_0 (pitch) information in and of itself does not help listeners with perception of reduced speech very much.

There is some past research on non-native listeners' perception of reduced spontaneous speech. Ernestus et al. (2017a) investigated advanced non-native learners of Dutch for their ability to orthographically transcribe stretches of spontaneous Dutch speech that contained clearly pronounced forms or reduced forms (e.g., 'als het goed is', "if it's correct", pronounced as [als ət xut ɪs] (clear) vs. [sxuts] (reduced)). They found that the advanced learners made more errors on reduced pronunciations than clear pronunciations, and that this effect was larger than among native Dutch listeners. They also found that non-native listeners often gave transcriptions of reduced forms that could not be grammatical or make sense in the sentence, suggesting that even advanced non-native listeners made less use of syntactic and semantic cues than of acoustic cues. The experiments by Bradlow and

Alexander (2007) and van de Ven et al. (2010) suggest that non-native listeners simply cannot use the semantic context when it is not clearly produced. The work of van de Ven et al. (2010) specifically shows that non-native listeners who have an Asian language as L1 and English as L2 are less able than natives to use the semantic relatedness of nouns to help them recognize the second noun, at least when the nouns are pronounced in reduced speech. Non-native listeners also cannot always use all cues present in the acoustic signal. Ernestus et al. (2017b) tested native speakers of American English and non-native speakers who had one of several native languages (Mandarin, Dutch, or Spanish) for their perception of sequences such as ‘I can’t imagine’ or ‘I can think’ that were produced by native English or Spanish speakers in spontaneous connected speech. In native American English, ‘can’ is usually produced with a highly reduced vowel or just a syllabic nasal, while ‘can’t’ is sometimes produced with no sign of a /t/, and sometimes produced with some variant of /t/. Mandarin, Dutch, and Spanish differ in whether they allow /nt/ syllable-finally. They found that the Mandarin and Spanish listeners, who lack syllable-final /nt/ in their own languages, showed more difficulty than Dutch or native English listeners in using the subsegmental cues to reduced /t/.

Morano et al. (2019) tested native Dutch speakers’ ability to recognize French words produced with or without devoicing of a vowel (e.g., ‘cycliste’, “cyclist”, produced with the first vowel voiced or devoiced), a reduction process that is not typical of Dutch. They primarily addressed questions of exemplar-based processing. Nijveld et al. (2022) similarly tested priming with match/mismatch of vowel reduction in English words such as ‘cassette’ (first vowel reduced or clear), as perceived by native English or non-native listeners (either Spanish or Dutch natives). Both papers showed that using token-specific details of individual exemplars in recognizing non-native words is possible, but occurs only under very specific, not entirely predictable conditions. Brand and Ernestus (2018) investigated French and Dutch native listeners’ recognition of French words with a schwa present or deleted (e.g., ‘revue’ pronounced as /RƏVY/ or /RVY/), and found that both native and advanced non-native listeners stored more than one variant of such words as part of their lexical representation. Based on the listeners’ own estimates of how frequent the two variants are, which should reflect how often they have heard the reduced and clear pronunciations, this study showed that non-native listeners can behave very similarly to native listeners if they have had enough experience with the reduced forms.

Some researchers examined how non-native listeners perceive especially careful speech, the opposite end of the continuum from reduced speech. Marcoux et al. (2022) tested this for Lombard speech (speech produced in noise). Smiljanić and Bradlow (2011) and Bradlow and Bent (2002) tested native and non-native production and perception of clear speech (asking speakers to pretend they are speaking to someone with a hearing impairment or to a non-native speaker). Both of these groups find that non-native listeners benefit in terms of perception from especially clear speech, just like natives.

The current paper turns to the question of how well non-native listeners can recover ambiguous words from highly reduced spontaneous, conversational speech. It also asks what types of information non-native listeners use to recognize such reduced words as compared to native listeners. The listeners in the current study, as university students in the Netherlands, have relatively high proficiency in their L2 (English). Some university courses are taught in English, and even though they live in a Dutch environment, most young adult Dutch speakers expect that they will need to use English regularly in some situations. As Gerritsen et al. (2016) discuss, people in the Netherlands, especially people who are younger, live in cities, or are involved with higher education, have higher English proficiency than people in other countries that do not historically have English as one of their national languages. However, they argue that English in the Netherlands is still a

foreign language, not a second language, and that the Netherlands is best understood as part of the “expanding circle” of world English usage and does not meet the criteria for an “outer-circle” country (such as India). However, in the 2019 EF English Proficiency Report (E. F. Education First Ltd., 2019), the Netherlands scored the highest of any of the 100 countries tested, outranking many countries with a colonial history of English use (outer-circle countries) such as Singapore, South Africa, Hong Kong, and Malaysia.

The situation of English in the Netherlands is also interesting because it is not clear what the target dialect is. As Gerritsen et al. (2016) discuss, British English is usually viewed as the goal for English education in the Netherlands. However, since such a large proportion of television, movies, and internet content reach the Netherlands in American English (not dubbed, but subtitled in Dutch), most native Dutch listeners probably have more exposure to American English than to the dialect they are in principle learning. This could be especially important for the perception of reduced spontaneous speech, since speech in media is not always produced as clear speech. Furthermore, as Gerritsen et al. (2016) discuss, there is some increasing tolerance of Dutch English as a regional variety rather than as just imperfect learning of a target native variety. In the current paper, we examine the perception of reduced speech by non-native speakers in an environment of rather high proficiency but mixed dialects in a country where English is still a foreign (rather than second) language.

2. Materials and Methods

2.1. Participants

The participants were 60 students at Radboud University, Nijmegen, in the Netherlands. Human subjects review and approval was provided by the Human Subjects Protection Program of the University of Arizona, project #03-0704-00. All listeners were native speakers of Dutch. Most began taking English classes in school at age 10 or 11 (range of ages at onset of English classes of 8–12, median of 10), and they had had 5–11 years of English courses (median of 8 years). English language television programs in the Netherlands are subtitled in Dutch rather than being dubbed, the general level of English fluency in the Netherlands is high, and university courses in many subjects are taught in English, so Dutch university students are typically very proficient in English. However, all of the participants were living in the Netherlands at the time of the experiment and had grown up entirely or primarily in the Netherlands. The staff who ran the experiment were Dutch and spoke Dutch with the participants. Only two participants had ever lived in an English-speaking country.

2.2. Materials and Procedures

The stimuli were created from 184 utterances taken from spontaneous conversations between 18 native speakers of American English with friends or family members, containing words/phrases like ‘he is’, ‘he’s’, ‘she was’, ‘she’s’, ‘we’re’, and ‘we were’. Examples include “Cuz he already told Steve he was in the wedding” and ‘you know, you were telling me about his roommate’. In each stimulus, the past tense form (e.g., ‘he was’) and the present tense form (‘he’s’ or ‘he is’) had to be reducible to be homophonous (which excludes strings such as ‘I was’, since ‘I’s’ is not possible in this variety of English). Materials were identical to those in Experiment 3 of Warner et al. (2022), which is the native listener version of the current study.

Each of the 184 stimulus utterances was presented with three levels of context: the full-context condition consisted of the entire utterance (e.g., ‘Cuz he already told Steve he was in the wedding’), supplying the listener with whatever syntactic or semantic cues to verb tense are available in the rest of the utterance. However, not all of the utterances

contained enough syntactic or semantic information to disambiguate tense. For example, “Cuz he already told Steve he’s in the wedding” would also be grammatical. The limited-context condition consisted only of the target word/phrase plus a portion of the signal extending out to the outer edge of its surrounding vowels, e.g., /iv hi w_ΛZ I/ (‘-eve he was i-’). This amount of context should be enough to provide the listener with information about speech rate of the utterance, but not enough to provide lexical information about the identity of surrounding words in most cases. (The first author’s judgement is that the only surrounding word that can be readily recognized is ‘like’ when it follows the target.) The Isolation context consisted of just the target word/phrase (e.g., ‘he was’, ‘we’re’, ‘she’s’). In all cases, the stimulus consists of how that portion of speech was actually produced during the spontaneous conversation it was drawn from. This means that the speech is often highly reduced, so that ‘you’re’ or ‘we were’ might consist of a single central vowel, and ‘he’s’ or ‘she was’ might consist of a very short, low-amplitude vowel and brief frication noise. As explained in Warner et al. (2022), the average duration of the Isolation portion (the target word/phrase itself) was just 262 ms, reflecting the reduced nature of spontaneous speech. An example stimulus set appears in Figure 1.

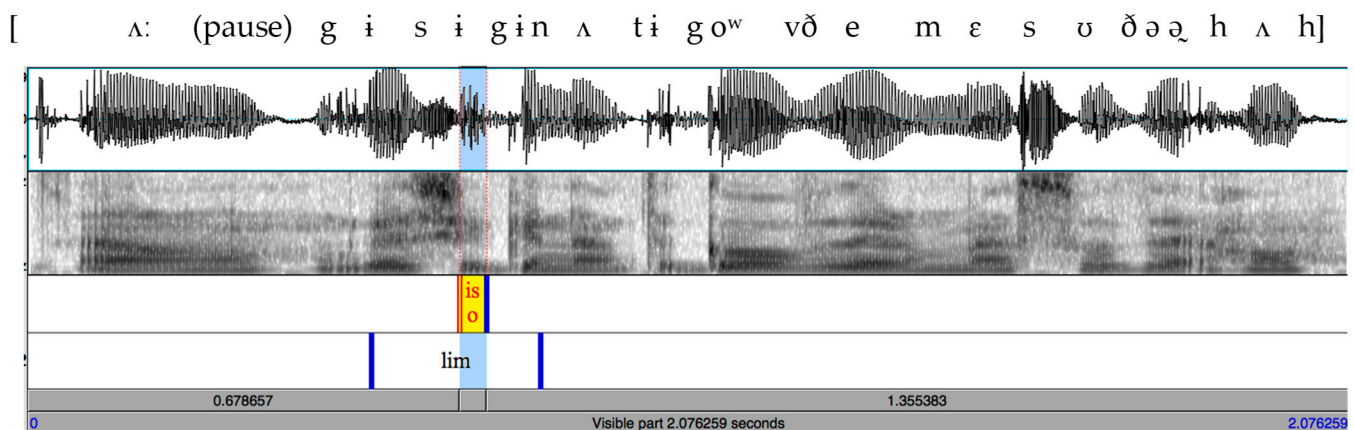


Figure 1. Waveform and spectrogram of the stimulus ‘Oh, guess you’re gonna hafta go over there and mess with it, huh?’ (referring to repairing a computer), containing highly reduced speech, with the target ‘you’re’ realized as a single central vowel. The portion marked “iso” is the portion corresponding to the target ‘you’re’ and constitutes the isolation condition stimulus. The portion labeled “lim” constitutes the limited-context stimulus, and the entire figure constitutes the full-context stimulus.

Details of acoustic criteria for setting boundaries of these stimuli are presented in Experiments 2 and 3 of Warner et al. (2022) and follow typical phonetic segmentation methods. These were applied based on the sounds that were actually produced in a given token, not on the sounds expected in a careful pronunciation. For example, in Figure 1, the boundaries of the word ‘you’re’ were located based on criteria for a vowel after an [s] and before a voiced stop, not based on criteria for the boundaries of a careful production [jɔːr].

The experimental procedures were identical to those of Experiment 3 in Warner et al. (2022), except that the experiment was run at Radboud University, Nijmegen, the Netherlands, and the listeners responded to additional language background questions and participated in an additional short lexical decision task (~5 min) in order to measure their English proficiency (Lemhöfer & Broersma, 2012). The additional language background survey contained questions about listeners’ acquisition of English, such as how many years of English courses they had taken and how often they watch English language television. They were also asked to self-rate their English proficiency on a 10-point scale. The participants’ median LexTale score was 73.81, approximately comparable to the Dutch

student population tested by [Lemhöfer and Broersma \(2012\)](#). Their self-rating of their English proficiency had a median of 7 on a scale of 1–10. Tests of correlation with these measures are included below.

The listeners participated in the experiment while seated in a sound-protected booth and heard the stimuli over headphones (comparable to the native listener experiment in [Warner et al., 2022](#)). The E-Prime software was used to administer the program and collect responses. For each stimulus (regardless of amount of context), listeners were asked to press a button to indicate which of two words/phrases displayed on the screen they heard in the stimulus (e.g., ‘he’s/he is’ or ‘he was’ as the two response options). Listeners were not asked to distinguish between contracted and uncontracted present tense forms (‘he’s’ vs. ‘he is’); they were only asked to select either the present or past option.

Alternatively, we could have given only ‘he is’ (instead of ‘he’s/he is’) vs. ‘he was’ as response options in order to avoid contractions, as participants may be biased to the contracted ‘he’s’ option when they hear reduced pronunciations. However, many stimuli clearly sounded like contracted forms (e.g., ‘he’s’, ‘we’re’, etc.), so if one ran the experiment this way, one would have to instruct listeners that if they hear ‘he’s’ that counts as ‘he is’, so it is not clear that this alternative method would provide any additional information. Furthermore, there are also many other possible sources of bias. For instance, the lexical item ‘he’s’ is shorter than ‘he was’ in number of phonemes and typically in duration. Therefore, if a stimulus has a short duration, perhaps because of spontaneous speech reduction or a fast speech rate, listeners may be biased toward the present tense response because of stimulus duration, apart from the issue of spelled contractions as response options. The previous work with native listeners ([Warner et al., 2022](#)) did indeed find bias toward either present or past responses in the native listener results, and argued for more than one cause of bias. In the current work, the interpretation of bias is discussed in Sections 4.1 and 4.3 below.

The response options on the screen were always appropriate to the stimulus (e.g., ‘we’re/we are’ and ‘we were’ for tokens containing the present or past of those targets, etc.). The use of ‘he’s/he is’ as the present tense option, representing both contracted and uncontracted forms, could introduce bias into the experiment if participants are more inclined to choose the present because its orthographic representation, being contracted, better matches the speech register they hear. However, other potential response options would introduce other biases, and this option keeps the results comparable to those obtained for the native listeners in ([Warner et al., 2022](#)). Bias from this and other sources is analyzed in Section 3.3 below.

The three levels of context were presented in separate blocks, with Full context as the first block, Limited context next, and Isolation as the last block, in order to maintain comparability with the native listener study. If listeners failed to respond within 9 s from the onset of the stimuli, the E-Prime system presented the next stimulus. Within each context block, listeners first heard five practice items made from similar utterances. Within each context block, stimuli were additionally blocked by speakers to allow listeners to adapt to specific speakers’ voices, and one acclimation item by that speaker was included at the start of the speaker’s block, with no data collected from that item. Further details of the procedures are included in ([Warner et al., 2022](#)). Participants failed to respond within the 9 s time-out period in just 103 trials (0.3%). The median reaction time to all other trials was slightly more than 1100 ms from the onset of the stimulus. The total duration of the experiment was approximately 50 min.

2.3. Statistical Analysis

The data were analyzed with the regression-based method of generalized linear mixed effects models with a binomial link function, using the lme4 package (Bates et al., 2015, version 1.1-35.5) of R (using glmer), with the correctness of the response as the dependent variable. Model selection was performed using Anova comparison. Random intercepts for Subject (participant) and Item (sentence), as well as appropriate random slopes by Subject, were included if the model converged and did not give singular fit warnings. Random intercepts for speaker (who produced the stimulus) were also tested, but generally either caused a failure to converge or singular fit warnings or failed to improve the model significantly. Simple effects were tested by splitting a factor only when motivated by a significant interaction. Although the model that tested the interaction sometimes gave a warning message because of its additional complexity, these interactions did provide a reason for following up with tests of simple effects. The exact model used for each test reported here appears in the footnotes.

3. Results

3.1. Proportion Correct

The results, presented as proportion correct, appear in Figure 2. During the analysis of the data in (Warner et al., 2022), it became clear that native listeners' behavior differs substantially in stimuli with vs. without a quotative 'like' after the target word/phrase (as in the stimulus "And he was like, 'What's wrong?!'" as opposed to examples above without 'like'). This pattern also held for the non-native listeners' data. Therefore, the presence/absence of 'like' immediately after the target word/phrase was included as a predictor variable in the statistical analysis. However, because this variable was not planned during experimental design, there were too few 'are like' and 'were like' items for statistical analysis. As such, the few items with 'are/were like' were excluded, and the 'like' factor was only analyzed for singular verbs ('is/was like').

For this study, the factor of primary interest is Context (with Limited as the reference level), since the main question concerns the types of information non-native listeners use to perceive ambiguous reduced speech. Models of all of the singular verb data (is/was items), with Verb Tense, presence/absence of Like, and Context as factors, gave failure to converge warnings, even with a minimal random effects structure, but did indicate significant interactions among all three factors. Even though these models had warnings, the significant interactions indicate that the effect of Context was not the same at all levels of other factors.¹ We therefore split the data into subsets to test just the Context factor. The same was true for a model of all plural verb data (are/were, with 'like' items excluded) with Verb Tense and Context as factors. Therefore, for further analysis, the data were split into six subsets: 'is' items without 'like', 'was' items without 'like', 'is' items with 'like', 'was' items with 'like', 'are' items without 'like', and 'were' items without 'like'. For each subset the predictor variable Context was tested, with Limited context as the reference level, using the same methods of model selection as described above.²

For the Singular 'is' targets not followed by 'like', the Limited context shows significantly higher accuracy than the tokens in isolation ($\beta = -0.28$, $z = -2.62$, $p < 0.01$), and Full context shows higher accuracy than Limited context ($\beta = 0.56$, $z = 4.50$, $p < 0.001^3$). For singular 'was' not followed by 'like', neither Isolation nor Full context had significantly different identification from Limited context (Isolation: $\beta = 0.05$, $z = 0.47$, $p > 0.05$; Full: $\beta = 0.04$, $z = 0.39$, $p > 0.05$). Thus, for 'is/was' targets not followed by 'like', both types of context facilitate the perception of 'is', but no type of context helps listeners to recognize 'was'. This result for 'was' differs from the native English listeners, who benefited in their perception from the addition of Limited context but did not gain additional information

from the Full context. To verify that the difference in use of Limited context between the native and non-native listeners is significant, we conduct a post hoc comparison of the data from (Warner et al., 2022) and this paper for only the ‘was’ without ‘like’ condition, testing the factors Context (Limited as reference level) and Native Language (English as reference level) and their interaction. This model showed a significant interaction between Native language and Context for the Isolation context only ($\beta = 0.34, z = 2.14, p < 0.04$).⁴

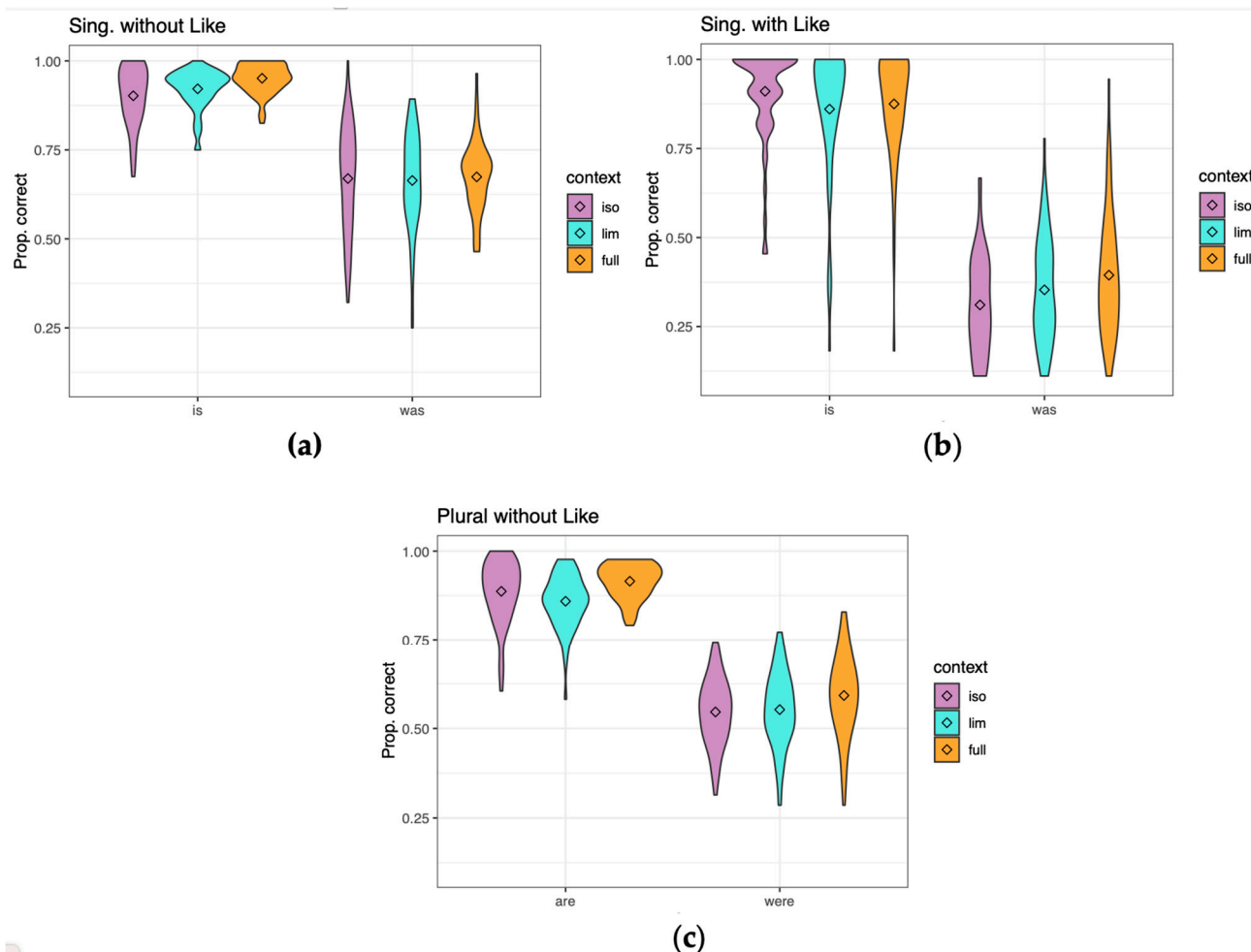


Figure 2. Results for listeners hearing targets with various amounts of context (iso = Isolation, lim = Limited, and full = Full utterance context, showing the distribution of listeners’ averages over items). Dots indicate the means for each condition. (a) ‘Is’ and ‘was’ targets not followed by ‘like’. (b) ‘Is’ and ‘was’ targets followed by ‘like’. (c) ‘Are’ and ‘were’ targets not followed by ‘like’.

For the ‘is’ targets followed by ‘like’, listeners performed significantly better in Isolation than with Limited context ($\beta = 0.57, z = 3.11, p < 0.005$). This apparent negative effect of added context on identification will be discussed below and matches the native listeners’ results. Full context led to no difference relative to Limited context ($\beta = 0.15, z = 0.86, p > 0.05$). For ‘was’ targets followed by ‘like’, Limited context was significantly more accurate than Isolation ($\beta = -0.30, z = -2.63, p < 0.01$), and Full context provided an additional benefit ($\beta = 0.31, z = 2.80, p < 0.01$)⁵. This significant improvement from Limited to Full context differs from the results for native listeners, who only benefited from limited context in the ‘was like’ condition, with no additional improvement with full context. However, the post hoc test of the language–context interaction did not show significance in the ‘was like’ condition (Isolation vs. Limited: $\beta = 0.24, z = 1.34, p > 0.05$; Limited vs. Full: $\beta = 0.05, z = 0.30, p > 0.05$).

For the plural 'are/were' (tokens followed by 'like' excluded), the 'are' targets showed Limited context having significantly worse identification than either the Isolation or Full (Limited vs. Isolation: $\beta = 0.28$, $z = 3.23$, $p < 0.005$; Full vs. Limited: $\beta = 0.63$, $z = 6.71$, $p < 0.001$). This is unexpected, as it seems to indicate that in this case, like the 'is like' case above, additional information lowers accuracy rather than raising it. However, as identification was near ceiling in these conditions, these differences may not be meaningful. For the 'were' targets, Limited context provided no benefit relative to Isolation ($\beta = -0.04$, $z = -0.49$, $p > 0.05$), but Full context led to significantly improved perception ($\beta = 0.23$, $z = 3.12$, $p < 0.005$). This result for perception of 'were' is the opposite of the native listeners' results in this condition: native listeners benefited from Limited context, but gained no additional accuracy from Full context. A post hoc test of Native Language by Context was also carried out for this condition ('were' without 'like'). It showed a significant interaction of Native Language by context, only for the Isolation condition relative to the Limited context reference level ($\beta = 0.56$, $z = 5.12$, $p < 0.001$). The non-native listeners cannot use the Limited context to improve their accuracy of identification of 'were', but they do gain additional information from Full context.

3.2. Relationship to English Proficiency

We examined correlations between the accuracy of identification and the various measures related to English proficiency and the amount of exposure to English in particular for accuracy in the 'was' without 'like' and 'were' conditions. Because of bias toward present tense responses, the listeners' accuracy for the 'is' and 'are' conditions may be too close to the ceiling to be useful. Because of the low detectability before 'like', we focused on the conditions without 'like' to answer the question about the relationship to proficiency. Accuracy (for 'was' and 'were' without 'like', for the Full-context condition) did not correlate significantly with most measures (e.g., the age of onset of studying English or years of English studied, which have a limited range due to the standardization of the school system). Accuracy for 'was' (without 'like', full context) correlated significantly with self-reporting of English proficiency ($r = 0.318$, $p < 0.02$, $N = 60$), and correlated marginally with listeners' estimation of how often they watch TV in English ($r = 0.257$, $p < 0.05$, $N = 60$). Accuracy for 'were' (full context, without 'like') showed only a non-significant trend toward correlation with self-reported proficiency ($r = 0.227$, $p = 0.081$, $N = 60$) and score on the lexical decision task ($r = 0.234$, $p = 0.071$, $N = 60$). Thus, there is some suggestion that higher L2 proficiency in English could be related to better ability to comprehend reduced function words, but it may be that the Dutch university students tested did not vary enough in English proficiency to show this effect strongly.

3.3. Detectability and Bias

The listeners show a bias toward the present tense response for all three verb pairs (singular without 'like', singular with 'like', and plural without 'like') in this experiment. This is reflected in Figure 2 by the extremely high proportion correct for 'is' and 'are' relative to the corresponding past tense forms, which does not indicate that listeners are especially good at recognizing present tense verbs, but rather that, given any stimulus, they are more likely to choose the present than the past response. In order to examine how well the listeners can distinguish between the present and past verbs within each tense pair, we use signal detection measures to separate detectability of the tense (d') from bias. These results appear in Table 1.

Table 1. Signal detection measures d' (detectability) and β (bias), and average proportion correct across the present and past verb of the pair, for the tense distinction for each pair of conditions. Positive β indicates bias toward the past response, negative β indicates bias toward the present response.

Condition	Context	d'	β	Avg. Prop. Correct
is/was, no 'like'	Isolation	1.733	−0.739	0.786
	Limited	1.842	−0.917	0.793
	Full	2.106	−1.267	0.813
is/was, with 'like'	Isolation	0.843	−0.780	0.609
	Limited	0.694	−0.512	0.605
	Full	0.884	−0.626	0.635
are/were, no 'like'	Isolation	1.326	−0.726	0.717
	Limited	1.209	−0.570	0.706
	Full	1.605	−0.914	0.754

It is noteworthy here that the non-native listeners' average proportion correct (which abstracts away from bias) is not much lower than the native listeners in (Warner et al., 2022): the Dutch listeners' average proportion correct ranges from 0.605 in their lowest-scoring pair to 0.813 in their highest-scoring pair. The native listeners' comparable scores range from 0.639 to 0.878.

The d' results show that listeners have more ability to detect the difference between present vs. past verbs when those verbs are not followed by 'like'. Although the listeners are biased toward the present tense response ('is/are') in all conditions, regardless of following 'like', when there is a 'like' present, it seems that bias is almost the only thing listeners use, with rather little detectability. These results are similar to the results for native listeners in Experiment 3 of (Warner et al., 2022).

4. Discussion

4.1. Comparison to Native Listeners' Results

To facilitate comparison, the results for the effect of context in the current study of non-native listeners (Dutch L1, English L2) are summarized in Table 2 alongside the results for native English listeners performing the same experiment (Warner et al., 2022).

Notably, these non-native listeners are nearly as successful as native listeners overall at recovering the verb tense from these reduced, ambiguous tokens (compare Table 1 above to Table 1 in (Warner et al., 2022)). The average proportion correct across the past and present form in a given tense pair, which abstracts away from bias, is only a few percentage points lower for non-native listeners, as discussed above. This may be taken as an encouraging finding for adult learners of non-native languages: even learning to parse highly reduced spontaneous, ambiguous speech is not out of reach for advanced learners, at least for cases like Dutch and English where the languages have many phonological similarities.

It is also informative to compare the non-native and native listeners' average proportion correct and detectability (d'), specifically for the isolation condition (Table 1 of this paper and of Warner et al., 2022). While the non-native listeners' proportion correct and d' are slightly lower than native listeners' are for the same conditions, even non-native listeners' measures in the Isolation condition are higher than native participants' measures in the two control conditions (Orthographic information only and auditory presentation with the target word(s) replaced by a beep, Experiments 1–2 of that paper). The two control conditions (which only native listeners participated in) provide participants with all of the syntactic and semantic information from the entire rest of the utterance, but do not provide them with the speech signal for the target word(s) ('he's, we were', etc.) themselves. As

mentioned above, the rest of the utterance is not always sufficient to clarify what the tense of the target word is, since both present and past are grammatical in some of the stimuli. Even the non-native listeners perceived the distinction between present and past tense more accurately based on just the acoustic information in the target word(s) themselves (Isolation condition, no context given) than the native participants did based on all the information in the entire rest of the utterance, as shown by average proportion correct and d' in the Isolation condition in all three tense pairs. That is, even non-native listeners can derive more useful information from the acoustic signal of these highly reduced words that average only 262 ms in duration than native listeners can from the entire rest of the utterance. This suggests that like native listeners, proficient non-native listeners are more strongly influenced by acoustic cues within a word itself than by semantic/syntactic context. This finding aligns well with past work by Ernestus et al. (2017a), who found that non-native listeners relied less heavily on syntactic and semantic cues than native listeners did in perceiving reduced speech.

Table 2. Comparison of significance and direction of effect for context effects in non-native vs. native listeners. Native listeners' results are from Warner et al. (2022) Experiment 3. "Worse" indicates significantly lower accuracy of identification of the verb relative to the reference condition (Limited context), "better" indicates significantly more accurate identification. "Iso." refers to the Isolation condition, "Lim." to Limited context, and "Full" to Full utterance context. * indicates that the interaction with Language was significant, "n.s." indicates that the interaction was not significant. Interactions with Language were only tested where motivated as post hoc comparisons.

Condition	Context Comparison	Native Listeners	Non-Native Listeners
is, no 'like'	Iso. vs. Lim. Lim. vs. Full	Iso. worse Full better	Iso. worse Full better
was, no 'like'	Iso. vs. Lim. * Lim. vs. Full n.s.	Iso. worse non-sig.	non-sig. non-sig.
is, with 'like'	Iso. vs. Lim. Lim. vs. Full	Iso. better non-sig.	Iso. better non-sig.
was, with 'like'	Iso. vs. Lim. n.s. Lim. vs. Full n.s.	Iso. worse non-sig.	Iso. worse Full better
are, no 'like'	Iso. vs. Lim. Lim. vs. Full	Iso. worse Full better	Iso. better Full better
were, no 'like'	Iso. vs. Lim. * Lim. vs. Full n.s.	Iso. worse non-sig.	non-sig. Full better

Finally, the direction of bias in Table 1 above shows a further similarity of the non-native listeners to the natives: both groups have a bias toward present tense responses in all conditions. This likely stems from the acoustic reduction of the stimuli because they were produced in spontaneous, casual conversational speech. Reduction overall shortens the duration of words, and in the case of these words/phrases, a shorter duration is a potential cue for the present tense. Native and non-native listeners both display this direction of bias across the board.

4.2. Use of Various Types of Context Information When Not Followed by 'Like'

The most obvious pattern one could expect, if all types of information are useful to listeners, is that listeners would show higher accuracy in perceiving the target word/phrase in Limited context relative to Isolation, and would show yet higher accuracy in Full context relative to Limited context. This is what we see for both native and non-native listeners for

the verb 'is' when it is not followed by 'like'. This suggests that the case of 'is' is not further reduced by being in the high-frequency phrase 'is like'. Information from the speech rate (Limited context) and the semantics/syntax of the rest of the utterance (Full context) are helpful to both native and non-native listeners. However, this is not the case for the other five verb conditions.

Most notably, for both 'was' and 'were' without 'like', the native listeners show significant improvement in identification when they hear Limited context, relative to hearing the target in Isolation. Speech rate information, and possibly also coarticulation information, provides cues that native listeners can use to recover the speaker's intended form in reduced speech. One can think of this as the native listener adjusting the category boundary for how long a speech signal has to be to be classified as the category 'he was' instead of the shorter 'he's'. If the surrounding speech is fast, then the listener expects the speech in 'he's' or 'he was' to be fast too, and with the addition of Limited context (out to the edge of the surrounding vowels) the listener recognizes the string as 'he was' because its duration is too long to be just the shorter 'he's' at that speech rate. This is similar to previous results showing that listeners can use the speech rate to adjust category boundaries for both individual segment perception and the perception of reduced words (Brown et al., 2012; Miller & Volaitis, 1989; Volaitis & Miller, 1992; Gottfried et al., 1990; Dilley & Pitt, 2010; Niebuhr & Kohler, 2011; Heffner et al., 2013). The non-native listeners, however, do not show this benefit of Limited context relative to Isolation, either in the 'was' or 'were' conditions (without 'like').

Furthermore, the native listeners do not show any significant additional benefit of full context in these two conditions. The native listeners recover somewhat from the reduced speech in these conditions based on just the Limited context, and additional syntactic/semantic information in the sentence does not help. The non-native listeners, however, do show a benefit from Full context (syntactic/semantic information) in identifying 'were', but not for 'was', (without 'like'). Even these rather high-proficiency non-native listeners who hear English around them in their own country frequently cannot use speech rate information to recover longer strings (e.g., 'he was/we were') from highly reduced spontaneous speech, but they do use syntactic/semantic context.

There are several possible explanations for why non-native listeners might fail to use surrounding speech rate information to recover reduced words. Morrill et al. (2016) and Baese-Berk and Bradlow (2021) find that non-native speakers' speech rates are more variable utterance to utterance than those of natives' in read speech, but less variable than natives' in unscripted speech such as storytelling. While these effects are somewhat small, if non-native speakers' speech rate is also less variable than natives' speech rate in spontaneous conversation, this could lead to non-native speakers having more difficulty adjusting perceptual boundaries for duration cues when hearing spontaneous speech from native speakers. Alternatively, non-native listeners may simply have too much cognitive load while parsing spontaneous, casual conversational speech in their L2 to apply the surrounding speech rate cue, or they could have a slower processing speed, making it difficult to integrate this low-level acoustic cue from neighboring words and apply it to the comprehension of the target word. Counter to the current results, both Dilley et al. (2013) and Baese-Berk et al. (2016) find evidence that, under some circumstances, non-native listeners can use the surrounding utterance speech rate as a cue to the presence/absence of potentially reduced speech sounds. However, both of these studies provide speech rate cues over the entire (resynthesized) utterance, while ours provides these cues in natural, non-synthesized speech, but for only two partial syllables. Thus, our method provides far less speech rate information. Anecdotally, we can report that both native and non-native listeners seem to hear the stimuli with something like categorical perception. When we play

the stimuli in the various context levels for native and non-native listeners to audiences or to our own students, listeners are very sure on a given token of which they have heard (past or present verb), even when their judgement differs from others in the room. The integration or the lack of integration of the speech rate cue happens at an automatic, unconscious stage of processing. The difference between native and non-native processing that leads to the difference we observed happens before reaching the perception of a category. The exact mechanism of this will be left to future research.

The remaining condition ('are' without 'like') shows a surprising effect for non-native listeners of the Limited-context condition being significantly worse than either of the other two levels of Context. However, since identification in this condition is near ceiling, this may not be meaningful.

4.3. Effects of Following 'Like'

The situation is somewhat different for 'is/was' followed by 'like'. Here, both native and non-native listeners rely heavily on bias, showing lower detectability for the tense distinction (Table 1). Statistically, the significant interaction of Like with Tense in the overall analysis, which motivated the splitting of the data into subsets for further testing, shows that both native and non-native listeners behave similarly overall with regard to presence vs. absence of 'like'. As explained in (Warner et al., 2022), 'is/was'+ 'like' is an extremely common collocation in the spontaneous, casual conversation of many of the young adult speakers whose conversations provided the stimuli. As such, sequences such as 'he was like' and 'she's like' are extremely reduced in their speech, even more so than the same words when followed by something other than 'like'. Because of the extreme reduction, many of these tokens have extremely short durations, even for the longer past tense form. Short duration is a cue to the present tense form, since 'he's' is shorter than 'he was'. Thus, a great many of these tokens sound like the shorter present tense form to listeners, regardless of which tense was actually produced. This is reflected as very high accuracy for the 'is' condition before 'like', and very low accuracy for the 'was' condition before 'like', because both native and non-native listeners tend to respond with 'is' regardless of the tense of the stimulus (they show bias toward the present tense). This bias must stem from the acoustic properties of the target word/phrase (such as reduction), not from listeners' expectations about 'like', because this strong bias toward 'is' and low performance on 'was' are present even in the isolation condition, where listeners could not hear that the token was followed by 'like'. Thus, the collocation with 'like' must affect the acoustics of the preceding verb in a way that causes this interaction, most likely by causing stronger reduction on the preceding verb. Both native and non-native listeners are misled by the acoustic cues to misperceive 'is' instead of the speaker's intended 'was' when a following 'like' in the original recording (which they do not hear) causes greater reduction.

This bias results in an apparently surprising context effect: for 'is like', listeners' accuracy is actually lower with Limited context than in Isolation. Adding contextual information seems to lower accuracy instead of improving it. We believe this shows that both native and non-native listeners rely even more heavily on bias (toward present tense, based on short durations) in the Isolation condition, while the Limited context helps them to move slightly away from bias. Thus, it is not that listeners become worse at perceiving 'is like' when provided with context, rather it is that the context cues allow them to move slightly away from their bias for the present response. It is possible that they can recognize the word 'like' after the verb with Limited context for many of these tokens (based on the /laɪ/ portion included in Limited context) and can perhaps use the knowledge that this is the 'is/was like' collocation to adjust their expectation slightly for how reduced the verb should be. Correspondingly, both native and non-native listeners

show improvement in the perception of ‘was like’ with the addition of Limited context, also reflecting a shift away from relying on bias. As in the conditions without ‘like’, the native listeners gain no additional benefit from Full (syntactic/semantic) context, while the non-native listeners do.⁶

5. Conclusions

Overall, the current study shows that highly proficient non-native listeners can recover information from ambiguous highly reduced speech almost as well as native listeners do. However, they use a different type of information to achieve this, specifically in the case of longer words that have been shortened by reduction (‘we were’, ‘he was’, etc.): native, but not non-native, listeners are able to use the speech rate of surrounding speech to recognize that something must have been deleted. Native listeners shift the boundary between shorter ‘we’re’ and longer ‘we were’ depending on the surrounding speech rate, similar to how listeners shift the boundary between segmental distinctions based on speech rate (Miller & Volaitis, 1989; Dilley & Pitt, 2010; and others discussed above). One can think of this as a subconscious process of realizing that, if the speech rate is that fast, the duration of this stretch of the signal is too long to represent just ‘we’re’ at this speech rate, and therefore something must have been deleted, so the intended words might be ‘we were’. However, the non-native listeners in this study do not act in this way. The addition of speech rate information in Limited-context stimuli does not help the non-native listeners to improve their identification of reduced longer forms at all. On the other hand, non-native listeners do make progress toward identifying the reduced longer forms when they hear the syntactic/semantic information in the rest of the utterance. The addition of syntactic/semantic information does not help native listeners beyond what they are already able to recover from the speech rate (Limited) information. This pattern indicates that non-native listeners rely more heavily on the meaning and syntactic context of the rest of the utterance to perceive difficult portions of the speech stream, whereas native listeners do not. For native listeners, acoustic cues are more important than syntactic/semantic cues in recovering reduced longer forms.

However, the non-native listeners in the current study perceive reduced speech very much like the native listeners do in other ways. Notably, they show the same strong bias toward present tense responses and low detectability for stimuli before ‘like’ as natives. As was true for the native listeners, this bias is present even when hearing the target words in the Isolation condition, where the listener cannot know that the following word is going to be ‘like’. Non-native listeners also match the native listeners’ results in being biased toward present tense responses for all conditions. Finally, even non-native listeners show greater detectability for the tense distinction based just on the acoustics of the target word/phrase itself than native participants do when supplied with the entire surrounding utterance less the target word/phrase (control experiments in (Warner et al., 2022)). That is, averaged over all items, even non-native listeners obtained more accurate information from hearing [i] as a realization of ‘you’re’ in the example in Figure 1 than natives did from hearing ‘Oh, guess ___ gonna hafta go over there and mess with it, huh?’ This provides strong evidence of the dominance of acoustic cues in the speech itself over syntactic and semantic cues.

In (Warner et al., 2022), we argued that native listeners show four separate types of evidence that acoustic cues outweigh any other type of information in the speech signal: (1) greater detectability of targets in isolation than in the control experiments that supplied the entire surrounding utterance, (2) improvement in the perception of reduced past tense forms (without ‘like’) based on the speech rate (Limited context), without further improvement based on syntactic/semantic information, (3) the strength and direction of bias when verbs are followed by ‘like’, even in the Isolation condition (where listeners are

unaware that a ‘like’ follows), and (4) consistent direction of bias toward shorter present tense responses in all conditions. The current results show that three of these same results hold for highly proficient non-native listeners. Of these four types of evidence, the only one that differs for the non-natives is the type of information they use to recover longer reduced forms that are not followed by ‘like’, where they use the syntactic/semantic information of the surrounding utterance, or fail to use either type of information, and show no benefit based on speech rate (Limited context) information.

Overall, these rather proficient non-native listeners, for whom English is frequently present as part of their daily life in their country, are similar to native listeners in their use of information for reduced speech perception in most ways, but they differ in not being able to utilize speech rate information to adjust their perception of reduced word categories. The Dutch listeners’ overall rather good ability to distinguish these reduced speech forms may reflect the prevalence of English in the mass media in the Netherlands (where most TV comes from the U.S., and it is subtitled rather than dubbed in Dutch). Even though speech in television and movies is produced by professional speakers (actors), reduction is common in such speech. As discussed above, English in the Netherlands is still a foreign language rather than a second language (Gerritsen et al., 2016), with Dutch speakers gaining substantial exposure to two native dialects (British and American), as well as to Dutch English as an expanding-circle variety. This may contribute to the Dutch listeners’ overall relatively good perception of the reduced speech targets in this study, but perhaps also to their inability to use speech rate to recover from reductions.

Author Contributions: Conceptualization, M.E., N.W. and B.V.T.; methodology, M.E., N.W. and B.V.T.; software, D.B.; validation, D.B. and N.W.; investigation, D.B. and N.W.; resources, N.W. and M.E.; data curation, D.B. and N.W.; writing—original draft preparation, N.W.; writing—review and editing, all; visualization, N.W. and B.V.T.; project administration, N.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by University of Arizona Institutional Review Board (Human Subjects Protection Program of the University of Arizona Project #03-0704-00, approved on 20 October 2003).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: We wish to acknowledge the help of the student assistants at the Radboud University who ran participants in the experiment and maintained data files. We also wish to thank all of the participants in the experiments, as well as the speakers who produced the stimulus recordings. We thank the reviewers for their helpful comments.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

¹ We also analyzed all of the results using traditional ANOVAs, which gave similar results for significance/non-significance in nearly all cases and provided confirmation of the significant interactions. The major difference in outcomes was that some of the pairwise comparisons of Limited Context to the other two contexts that are significant in the LMEs were only significant by subjects or by items, and were at $p < 0.10$ on the other test, in the ANOVAs.

² The model chosen for *is without like*, *is with like*, *was with like*, *are without like*, and *were without like*: Correct ~ Context + (1|Subject) + (1|Item); *was without like*: Correct ~ Context + (1+Context|Subject) + (1|Item).

- ³ Model for both ‘was’ and ‘were’ without ‘like’ as well as ‘was like’ (below): `glmer(Correct ~ Context * Language + (1 + Context | Subject) + (1 | Item))`. For ‘was’, the improvement relative to the same model without interaction is only significant at $p = 0.082$, however. For ‘were’ the improvement is fully significant ($p < 0.001$).
- ⁴ A reviewer suggests applying a Bonferroni correction for these tests of simple effects in order to be more conservative, even though they are motivated by statistically significant interactions. There are 12 tests of simple effects in all, so Bonferroni correction requires $p < 0.00417$ for each test to reach significance. This result thus remains significant with correction, as do all other significant results unless otherwise noted.
- ⁵ These two comparisons, for the ‘was like’ conditions, do not reach significance with Bonferroni correction.
- ⁶ The non-native listeners do not show improvement with either type of context in this ‘was like’ condition under the stricter criterion of $\alpha = 0.00417$ with Bonferroni correction. Under either criterion, their perception does not improve with the Limited context.

References

- Baese-Berk, M. M., & Bradlow, A. R. (2021). Variability in speaking rate of native and nonnative speech. In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 312–334). Cambridge University Press.
- Baese-Berk, M. M., Morrill, T. H., & Dilley, L. C. (2016, May 31–June 3). *Do non-native speakers use context speaking rate in spoken word recognition* [Paper presentation]. 8th International Conference on Speech Prosody (SP2016) (Vol. 979983,), Boston, MA, USA.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. [[CrossRef](#)]
- 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. [[CrossRef](#)] [[PubMed](#)]
- Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America*, 112(1), 272–284. [[CrossRef](#)] [[PubMed](#)]
- 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, 1240–1259. [[CrossRef](#)]
- Brouwer, S., Mitterer, H., & Huettig, F. (2012). Speech reductions change the dynamics of competition during spoken word recognition. *Language and Cognitive Processes*, 27(4), 539–571. [[CrossRef](#)]
- Brown, M., Dilley, L. C., & Tanenhaus, M. K. (2012, August 1–4). *Real-time expectations based on context speech rate can cause words to appear or disappear*. 34th Annual Meeting of the Cognitive Science Society (Vol. 34, No. 34, pp. 1374–1379), Sapporo, Japan.
- Dilley, L. C., Morrill, T. H., & Banzina, E. (2013). New tests of the distal speech rate effect: Examining cross-linguistic generalization. *Frontiers in Psychology*, 4, 1002. [[CrossRef](#)] [[PubMed](#)]
- Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, 21(11), 1664–1670. [[CrossRef](#)]
- E. F. Education First Ltd. (2019). *EF EPI: EF english proficiency index: A ranking of 100 countries and regions by english skills*. EF Education First.
- Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, 81(1–3), 162–173. [[CrossRef](#)] [[PubMed](#)]
- Ernestus, M., Dikmans, M. E., & Giezenaar, G. (2017a). Advanced second language learners experience difficulties processing reduced word pronunciation variants. *Dutch Journal of Applied Linguistics*, 6(1), 1–20. [[CrossRef](#)]
- Ernestus, M., Kouwenhoven, H., & Van Mulken, M. (2017b). The direct and indirect effects of the phonotactic constraints in the listener’s native language on the comprehension of reduced and unreduced word pronunciation variants in a foreign language. *Journal of Phonetics*, 62, 50–64. [[CrossRef](#)]
- Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, 39(3), 253–260. [[CrossRef](#)]
- Gerritsen, M., Van Meurs, F., Planken, B., & Korzilius, H. (2016). A reconsideration of the status of English in the Netherlands within the kachruvian three circles model. *World Englishes*, 35(3), 457–474. [[CrossRef](#)]
- Gottfried, T. L., Miller, J. L., & Payton, P. E. (1990). Effect of speaking rate on the perception of vowels. *Phonetica*, 47(3–4), 155–172. [[CrossRef](#)] [[PubMed](#)]
- Greenberg, S. (1999). Speaking in shorthand—A syllable-centric perspective for understanding pronunciation variation. *Speech Communication*, 29, 159–176. [[CrossRef](#)]
- Heffner, C. C., Dilley, L. C., McAuley, J. D., & Pitt, M. A. (2013). When cues combine: How distal and proximal acoustic cues are integrated in word segmentation. *Language and Cognitive Processes*, 28(9), 1275–1302. [[CrossRef](#)]
- 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* (pp. 29–54). The National International Institute for Japanese Language.

- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, 44, 325–343. [[CrossRef](#)]
- Marcoux, K., Cooke, M., Tucker, B. V., & Ernestus, M. (2022). The Lombard intelligibility benefit of native and non-native speech for native and non-native listeners. *Speech Communication*, 136, 53–62. [[CrossRef](#)]
- Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception and Psychophysics*, 46, 505–512. [[CrossRef](#)]
- Morano, L., Bosch, L. T., & Ernestus, M. (2019). Looking for exemplar effects: Testing the comprehension and memory representations of reduced words in Dutch learners of French. In S. Fuchs, J. Cleland, & A. Rochet-Capellan (Eds.), *Speech perception and production: Learning and memory* (pp. 245–277). Peter Lang.
- Morrill, T., Baese-Berk, M., & Bradlow, A. (2016). Speaking rate consistency and variability in spontaneous speech by native and non-native speakers of English. *Proceedings of the International Conference on Speech Prosody, 2016*, 1119–1123.
- Niebuhr, O., & Kohler, K. J. (2011). Perception of phonetic detail in the identification of highly reduced words. *Journal of Phonetics*, 39(3), 319–329. [[CrossRef](#)]
- Nijveld, A., Bosch, L. T., & Ernestus, M. (2022). The use of exemplars differs between native and non-native listening. *Bilingualism: Language and Cognition*, 25(5), 841–855. [[CrossRef](#)]
- Podlubny, R. G., Nearey, T. M., Kondrak, G., & Tucker, B. V. (2018). Assessing the importance of several acoustic properties to the perception of spontaneous speech. *The Journal of the Acoustical Society of America*, 143(4), 2255–2268. [[CrossRef](#)] [[PubMed](#)]
- Smiljanić, R., & Bradlow, A. R. (2011). Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness. *The Journal of the Acoustical Society of America*, 130(6), 4020–4031. [[CrossRef](#)]
- van de Ven, M., & Ernestus, M. (2018). The role of segmental and durational cues in the processing of reduced words. *Language and Speech*, 61(3), 358–383. [[CrossRef](#)] [[PubMed](#)]
- van de Ven, M., Ernestus, M., & Schreuder, R. (2012). Predicting acoustically reduced words in spontaneous speech: The role of semantic/syntactic and acoustic cues in context. *Laboratory Phonology*, 3(2), 455–481. [[CrossRef](#)]
- van de Ven, M., Tucker, B. V., & Ernestus, M. (2010, September 26–30). *Semantic facilitation in bilingual everyday speech comprehension* [Paper presentation]. 11th Annual Conference of the International Speech Communication Association (Interspeech) (pp. 1245–1248), Makuhari, Japan.
- van de Ven, M., Tucker, B. V., & Ernestus, M. (2011). Semantic context effects in the comprehension of reduced pronunciation variants. *Memory and Cognition*, 39, 1301–1316. [[CrossRef](#)] [[PubMed](#)]
- Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *The Journal of the Acoustical Society of America*, 92(2), 723–735. [[CrossRef](#)]
- Warner, N., Brenner, D., Tucker, B. V., & Ernestus, M. (2022). Native listeners' use of information in parsing ambiguous casual speech. *Brain Sciences*, 12(7), 930. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.