Laboratory Phonology
Journal of the Association for Laboratory Phonology

Guo, Z.-C., & Smiljanic, R. 2021 Speaking clearly improves speech segmentation by statistical learning under optimal listening conditions. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 12(1): 14, pp. 1–24. DOI: https://doi.org/10.5334/labphon.310

JOURNAL ARTICLE

Speaking clearly improves speech segmentation by statistical learning under optimal listening conditions

Zhe-chen Guo and Rajka Smiljanic
Department of Linguistics, the University of Texas at Austin, Austin, TX, US
Corresponding author: Zhe-chen Guo (zcadamguo@utexas.edu)

This study investigated the effect of speaking style on speech segmentation by statistical learning under optimal and adverse listening conditions. Similar to the intelligibility and memory benefits found in previous studies, enhanced acoustic-phonetic cues of the listener-oriented clear speech could improve speech segmentation by statistical learning compared to conversational speech. Yet, it could not be precluded that hyper-articulated clear speech, reported to have less pervasive coarticulation, would result in worse segmentation than conversational speech. We tested these predictions using an artificial language learning paradigm. Listeners who acquired English before age six were played continuous repetitions of the 'words' of an artificial language, spoken either clearly or conversationally and presented either in quiet or in noise at a signal-to-noise ratio of +3 or 0 dB SPL. Next, they recognized the artificial words in a two-alternative forced-choice test. Results supported the prediction that clear speech facilitates segmentation by statistical learning more than conversational speech but only in the quiet listening condition. This suggests that listeners can use clear speech acoustic-phonetic enhancements to guide speech processing dependent on domain-general, signal-independent statistical computations. However, there was no clear speech benefit in noise at either signal-to-noise ratio. We discuss possible mechanisms that could explain these results.

Keywords: Clear speech; speech segmentation; statistical learning; noise; artificial language learning

1. Introduction

An important task in understanding spoken language is the segmentation of continuous speech into discrete words. Adult listeners are able to discover word boundaries and extract possible word forms via statistical learning—that is, by tracking transitional probabilities across syllables in the speech stream (Saffran, Newport, & Aslin, 1996). For example, two syllables that frequently co-occur are likely to be perceived as word-internal, whereas two syllables with a low co-occurrence are interpreted as spanning a word boundary. This form of learning may occur without attention to the speech signal (Fernandes, Kolinsky, & Ventura, 2010; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997) and has been observed in infants (Saffran, Aslin, & Newport, 1996; Thiessen & Saffran, 2003, 2007). Such learning occurs in the visual modality as well (Kirkham, Slemmer, & Johnson, 2002; Turk-Browne, Jungé, & Scholl, 2005), suggesting that speech segmentation by statistical learning is rooted in a domain-general learning mechanism.

Listeners also exploit a variety of sub-lexical acoustic-phonetic cues for segmentation (for reviews, see Cutler, 2012; Davis, Marslen-Wilson, & Gaskell, 2002). One sub-segmental signal-driven cue that listeners rely on is coarticulation, or the articulatory overlap between neighboring segments (see Farnetani & Recasens, 2010, for a review on this topic). In general, segments are more coarticulated when they occur within a word or across a lower prosodic boundary than when they occur across words or span a higher-level prosodic boundary (Byrd, 1996; Byrd & Saltzman, 1998; Cho, 2004; Fougeron & Keating, 1997). Fernandes and colleagues demonstrated that coarticulatory information facilitated segmentation via statistical learning (Fernandes et al., 2010; Fernandes, Ventura, & Kolinsky, 2007). Using an artificial language learning paradigm, Fernandes et al. (2007) presented listeners with long speech streams containing strings of repeating syllables. Listeners were required to track the transitional probabilities across syllables for successful word segmentation. The results indicated that segmentation improved when segments within the words in the speech streams were coarticulated, compared with a condition where coarticulation cues were absent or incongruent with the transitional probabilities. Furthermore, when the speech signal was degraded by adding noise, the role of coarticulation was diminished and listeners relied more on the signal-independent statistical properties. The results showed that the signal-driven coarticulatory information enhances segmentation by statistical learning above and beyond domain-general statistical computations under optimal listening conditions.

One way in which speech signal itself can be enhanced is through conversational-to-clear speech modifications. Within the H&H theory (Lindblom, 1990), these modifications are understood as arising from movement along a continuum of hypo- and hyperspeech in response to talker- and listener-oriented forces in the communicative context. Clear speech occurs when the speaker shifts toward the hyperspeech end of the continuum in order to make their speech more intelligible for the listener. Conversational speech constitutes the speaker's shift toward the hypospeech end in optimal listening situations in order to conserve articulatory effort.

There are a number of such goal-oriented modifications in which talkers adapt their output in response to communication challenges, including noise-adapted speech, infant-directed speech, or speech produced in response to vocoded speech (for clear speech reviews, Pichora-Fuller et al., 2010; Smiljanić, 2021; Smiljanić & Bradlow, 2009). In line with previous work, we focus on the clear speech aimed at enhancing intelligibility for adult interlocutors with perceptual difficulties arising from hearing loss or low proficiency. Even though the various speaking styles share a number of features, clear speech modifications involve different acoustic-phonetic goals from, for instance, modifications aimed at increasing affective prosody to capture children's attention. The clear speech modifications typically include a decrease in speaking rate, an increase in vocal levels, a wider F0 range and higher F0 mean, more salient stop releases, vowel and consonant contrast enhancement, greater obstruent root-mean-square (RMS) energy, and increased energy at higher frequencies (Bradlow, Kraus, & Hayes, 2003; Ferguson & Kewley-Port, 2002, 2007; Gilbert, Chandrasekaran, & Smiljanic, 2014; Granlund, Hazan, & Baker, 2012; Hazan & Baker, 2011; Krause & Braida, 2004; Liu et al., 2004; Maniwa, Jongman, & Wade, 2009; Picheny, Durlach, & Braida, 1986; Pichora-Fuller et al., 2010; Smiljanić & Bradlow, 2005; Van Engen, Chandrasekaran, & Smiljanić, 2012), and so on.

It is well-established that clear speech acoustic-phonetic adjustments benefit perception for various communicative situations and listener populations (for reviews, see Pichora-Fuller et al., 2010; Smiljanić, 2021; Smiljanić & Bradlow, 2009; Uchanski, 2005). The ways by which they aid the listener include augmenting the speech signal, enhancing language-specific phoneme distinctions, and facilitating linguistic processing and cognitive functioning associated with speech perception and auditory memory (Bradlow & Bent, 2002; Cooke et al., 2013; Ferguson, 2012; Keerstock & Smiljanić, 2018, 2019; Krause & Braida, 2002; Payton et al., 1994; Picheny et al., 1985; Schum, 1996). There is also evidence that clear speech reduces lexical competition and enhances lexical access. Scarborough and Zellou (2013), for example, showed that words with many phonological neighbors are produced with greater hyper-articulation in clear speech compared to words with few phonological neighbors and this facilitated lexical decisions for high-neighborhood-density words. Van Engen (2017) similarly found that clear speech was helpful in reducing lexical competition for high-neighborhood-density words in noise for older adults. Finally, using a visual word recognition paradigm, van der Feest, Blanco, and Smiljanic (2019) showed that clear speech increased speed of word recognition for high-predictability sentences in quiet and in noise for young adult listeners. While none of these studies looked at speech segmentation, documented processing benefits suggest that clear speech could also make it easier for the listener to track statistical dependencies and improve segmentation.

Work by Palmer and Mattys (2016) provides evidence that at least some of the acoustic-phonetic adjustments typically associated with clear speech are beneficial for segmentation by statistical learning. Using the artificial learning paradigm, they showed that listeners' statistical segmentation of trisyllabic nonwords (e.g., *pabiku*) from speech streams in which they were presented continuously was facilitated when the speech rate was slower than when it was faster. As they argue, this may be because a slower rate allows more time for representations stored in working memory to be refreshed before they are displaced by incoming input. Given that clear speech is characterized by a slower speaking rate (Behrman, Ferguson, Akhund, & Moeyaert, 2019; Ferguson & Kewley-Port, 2002, 2007; Liu, Del Rio, Bradlow, & Zeng, 2004; Picheny et al., 1986; Smiljanić & Bradlow, 2005, 2009; Uchanski, Choi, Braida, Reed, & Durlach, 1996), Palmer and Mattys's finding supports the prediction that it will have an advantage over conversational speech in segmentation by statistical learning.

With regard to coarticulation and speaking style adaptations, the findings are mixed. Under the H&H framework, coarticulation is viewed as a low-cost articulatory behavior (Farnetani & Recasens, 2010). Therefore, similar to other hyper-articulated forms such as strengthening in prosodically strong positions (Cho, 2004), more exaggerated listener-oriented speech is expected to show coarticulatory resistance. Moon and Lindblom (1994) found that vowels in the /w__l/ frame produced in a hyper-articulated style by English speakers showed greater F2 displacements relative to those produced without attention to clarity, suggesting less coarticulatory influence from the neighboring consonants. Analogous findings have been obtained by studies comparing citation-form and spontaneous speech in French, Swedish, Spanish, and Catalan (Duez, 1992; Krull, 1989; Poch-Olivé et al., 1989). In contrast, Bradlow (2002) found that for English and Spanish speakers, clearly produced CV syllables exhibited neither exaggerated nor reduced coarticulation compared to those that were conversationally produced. Similar maintenance of coarticulation across speaking styles was reported by Matthies et al. (2001). Furthermore, unlike the above-referenced studies, in which speakers followed instructions for different styles and might not speak with a true communicative intent, Scarborough and Zellou (2013) investigated anticipatory nasal coarticulation in vowel-nasal sequences under several communicative contexts, including one in which speakers completed an interactive task with a real listener. They found that compared to read speech elicited with instructions to speak clearly or imagine someone hard of hearing, spontaneous speech in this real-listener context showed greater hyper-articulation as evident in vowel space expansion and *greater* vowel-nasal coarticulation.

Nevertheless, while the presence of the listener in Scarborough and Zellou's (2013) interactive task introduced a communicative intent, it did not imply a communicative difficulty that would explicitly motivate the speakers to adopt a more effortful, listener-oriented speaking style. Recently, Guo and Smiljanić (to appear) applied a whole-spectrum measure of coarticulation (Cychosz, Edwards, Munson, & Johnson, 2019; Gerosa et al., 2006) to analyze speech from the various communicative conditions of the LUCID corpus (Baker & Hazan, 2011), which include read speech elicited with instructions to speak clearly, spontaneous speech from an interactive task where pairs of speakers communicated in good listening condition, and spontaneous speech from the same task but with explicitly imposed communicative barriers (e.g., one speaker's voice was vocoded or mixed with talker babble). Results indicated that the speakers coarticulated less when completing the task under any condition with communicative barriers compared to the no-barrier condition. Moreover, compared to read speech obtained with instructions to speak casually, read clear speech showed a similar extent of coarticulatory resistance as spontaneous speech in most of the barrier-present conditions. Taken together, the findings on the relationship between coarticulation and speaking style, though still mixed, suggest that clear speech could be less coarticulated than conversational speech. If this is the case, the prediction is that the relatively greater coarticulation in conversational speech may facilitate word segmentation by statistical learning (cf. Fernandes et al., 2007, 2010) compared to clear speech with relatively less coarticulation.

This prediction gains additional support from research investigating the effect of speech rate on statistical learning. Unlike Palmer and Mattys (2016), Emberson, Conway, and Christiansen (2011) found that when listeners were presented with speech streams consisting of nonwords (e.g., *meep*) that formed statistically coherent triplets, speeding up presentation rate facilitated their discovery of which nonwords constituted a triplet. This suggests that reducing the temporal distances between elements in speech promotes auditory grouping of the elements into larger units. Since conversational speech is characterized by a faster speaking rate and hence reduced temporal distances between syllables, it could lead to better grouping of syllables into discrete units than clear speech. Given the contradictory predictions, the main goal of this study was to assess the role of speaking style on segmentation by statistical learning.

We furthermore wanted to explore whether speaking style variation affects statistical learning under adverse listening conditions, specifically in noise. It is well-documented that clear speech enhances word recognition in noise (e.g., Ferguson & Kewley-Port, 2007; Gilbert, Chandrasekaran, & Smiljanic, 2014; Liu et al., 2004; Picheny et al., 1986; Smiljanić & Bradlow, 2005) and this benefit may become larger as the signal-to-noise ratio (SNR) decreases (Bradlow et al., 2003; Bradlow & Bent, 2002; Payton et al., 1994). Work by Fernandes and colleagues showed that the contribution of coarticulation to segmentation by statistical learning is largest in quiet and becomes smaller or even absent in noise. If noise masks the beneficial coarticulatory cues, and particularly if coarticulation is more pervasive in conversational speech, segmentation may be better for clear compared to conversational speech even if a different speaking style effect is found in quiet.

However, noise may increase cognitive load and affect speech segmentation beyond covering the beneficial acoustic-phonetic cues (Francis, Love, & Boutin, 2019; McCoy et al., 2005; Peelle, 2018; Pichora-Fuller et al., 2016; Rabbitt, 1968, 1991; Rönnberg et al., 2013; Rönnberg, Rudner, Foo, & Lunner, 2008; Schneider, Bernarding, Francis, Hornsby, & Strauss, 2019; Van Engen & Peelle, 2014; Winn, Edwards, & Litovsky, 2015; Zekveld, Kramer, & Festen, 2010, 2011). Palmer and Mattys (2016) showed that the benefit of decreased speaking rate for segmentation by statistical learning was reduced in a dual-task paradigm, where listeners completed a concurrent task while performing speech segmentation. It made no difference whether the concurrent task involved phonological processing or non-linguistic visual processing. This led them to conclude that the benefit of a slower speech rate is supported by domain-general central processing resources within working memory which are taxed more when performing a dual-task than a single task. Similar to the effect of a dual task, increased cognitive load and listening effort due to signal degradation by noise may result in fewer resources available for speech segmentation (e.g., Miles, Jones, & Madden, 1991; Pichora-Fuller, Schneider, & Daneman, 1995; Rönnberg, Rudner, Lunner, & Zekveld, 2010; Zekveld, Kramer, & Festen, 2011). The increased processing load in noise would affect speech segmentation of both speaking styles equally though the adverse effect could be larger for clear speech as any benefit of its slower speaking rate would be reduced (cf. Palmer & Mattys, 2016). These findings suggest that the clear speech benefit for segmentation in quiet, if any, will be reduced or absent in noise. The second goal of the present study was thus to examine the role of speaking style adaptations in segmentation by statistical learning in noise and to gain a better understanding of the processes underlying successful segmentation.

We addressed the above questions using an artificial language learning experiment (e.g., Saffran, Newport, et al., 1996). Such an experiment uses nonsense speech material and therefore prevents listeners from relying on lexical knowledge to parse the speech signal. This was important for this investigation as we wanted to draw listeners' attention to the signal-dependent acoustic-phonetic cues and the relatively signal-independent statistical properties and away from the lexical cues (Mattys & Bortfeld, 2017; Mattys, White, & Melhorn, 2005). Similar to previous artificial language learning studies, our experiment consisted of a learning phase and a subsequent test phase. During the learning phase, listeners heard long speech streams in which syllables co-occurred with varying probability giving rise to the artificial language 'words.' Next, they identified the words in a two-alternative (word versus partword) forced-choice test. Higher recognition accuracy in the test is assumed to reflect better segmentation performance during the learning. To examine the effect of speaking style, listeners heard either a clear or a conversational version of the artificial language. To examine the effect of the listening condition, the learning-phase speech streams were either presented in quiet or masked with noise at increasing levels of difficulty: + 3 dB SPL SNR and 0 dB SPL SNR.

2. The experiment

2.1 Participants

One hundred and eighty speakers of American English (97 female, 79 male, two non-binary, and two who declined to identify gender) participated in the study, equally and randomly assigned to the six conditions (see below). They self-reported to have started learning English before the age of six and their mean age was 20.4 years (range: 18–46).[1] All participants signed written informed consent and filled out a detailed language background questionnaire adapted from the LEAP-Q questionnaire (Marian, Blumenfeld, & Kaushanskaya, 2007). The questionnaire indicated that 51 participants were functionally bilingual as they had also learned one or more languages other than English before the age of six and self-reported to be fluent in those languages. However, the data of the monolingual and bilingual participants were collapsed since there was no evidence that

[1] Instead of using the term 'native speaker/listener,' we opted to define our participants as those who have acquired English before the age of six, including those who may have acquired another language at the same time. The term 'native speaker' is inconsistently conceptualized across studies, is often imbued with prescriptivist notions and linguistic privilege, and may be connected with discriminatory practices (cf. Cheng et al., 2021 and Grammon & Babel, 2021). For similar reasons, we use 'language learners' instead of 'non-native' listeners.

the two groups performed differently in the experiment.[2] All participants passed a pure-tone screening test bilaterally at 25 dB hearing level for 1000, 2000, and 4000 Hz, and received course credit or a small honorarium. Four participants who failed the hearing screening were excluded and four additional participants were recruited for a total of 30 participants in each test condition.

2.2 Stimuli

Three vowels (/a, i, u/) and six consonants (/p, t, k, m, n, l/) were combined to form six trisyllabic CVCVCV sequences, which served as the 'words' of the artificial language: /pakila/, /timani/, /kutupi/, /mikamu/, /nuluta/, and /lipuna/. The three peripheral vowels were chosen to allow us to estimate the effect of speaking style on vowel space area. The consonants all occur in English. Another six 'partwords'—/pinulu/, /tamika/, /kilati/, /mutima/, /nakutu/, and /lalipu/—were created as three-syllables substrings across 'word' boundaries (e.g., /kutupi/ + /nuluta/ = /pinulu/) and used as distractors in the test phase. A 21-year-old female speaker who acquired English from birth and who had some phonetic training recorded the stimuli in a sound-attenuating room. Each target word or partword was written in the International Phonetic Alphabet (e.g., /pakila/), embedded in a carrier sentence (i.e., The word I said was ___), and presented on a PowerPoint slide. The recording consisted of two parts. In the first part, the speaker was instructed to read the target-bearing carrier sentences aloud in a conversational speaking style, "as if she was talking to her friends or someone familiar with her voice." Additionally, she was instructed not to reduce any vowels. In the second part, which followed after a 10-minute break, the speaker read the same sentences in the clear speaking style "as if she were talking to someone who cannot follow her conversationally or someone with hearing loss." Similar instructions have been used in previous studies to elicit clear and conversational speaking styles (see Pichora-Fuller, Goy, & Van Lieshout, 2010 and Smiljanić & Bradlow, 2009). The recordings were made with a MOTU UltraLite-MK3 Hybrid recorder and a Shure SM10A head-mounted microphone, digitized at a sampling rate of 44.1 kHz, and saved in WAV format.

The 24 stimuli—six words and six partwords for each style—were excised from the carrier sentences and manipulated with Praat (Boersma & Weenink, 2018). First, their F0 contours were flattened to the mean F0 for each style, calculated by averaging the F0 values across all the voiced intervals in the words and partwords. This rendered the learning-phase speech streams (described below) monotonous. The goal was to prevent listeners from using F0 contours for word segmentation (Endress & Hauser, 2010; Shukla et al., 2007) as F0 was not the focus of this study and to prevent ceiling performance on the artificial language learning task. The flattened F0 contour was at 232 Hz for the clear stimuli and at 210 Hz for the conversational ones, consistent with the finding that clear speech has higher overall F0 (Bradlow, Kraus, & Hayes, 2003; Hazan & Baker, 2011). The stimuli were equalized for the average RMS amplitude.

For each speaking style, the six resynthesized words were concatenated without pauses to form six long speech streams for the learning phase. Each stream contained 20 repetitions of each word with the repetitions occurring pseudo-randomly such that the same word did not follow itself. Thus, listeners would be exposed to 120 repetitions of each artificial language word. The total duration of the six speech streams was about

[2] To assess whether the monolingual and bilingual participants might perform differently, we fitted the Bayesian mixed-effects model described in Section 3 but also included Group (monolingual versus bilingual) and all its interactions with Style and Listening Condition as fixed effects. We found that the 95% credible intervals of all the group-related fixed-effects included zero, providing no evidence that bilingual experience had an effect on this task.

7.6 minutes for the clear speaking style and 5.9 minutes for the conversational speaking style. Successful segmentation during the learning phase depends on tracking statistical regularities, typically expressed as transitional probabilities between syllables. Using the formula of Saffran, Newport, and Aslin (1996), we found that the average between-syllables transitional probability was one for all the words and ranged between 0.55 and 0.60 (mean: 0.58) for the partwords in each speech stream.

Two noise-masked versions of the six speech streams were created for each speaking style by mixing them with a Gaussian noise shaped to match the long-term average spectrum of the words. In each style, two noise conditions were created with an increasing level of difficulty: +3 and 0 dB SPL SNR. These SNRs were selected based on previous literature examining clear speech intelligibility benefit (Smiljanić, 2021; Smiljanić & Bradlow, 2009) and pilot testing. All the speech streams were faded in and out with five-second logarithmic ramps. There were six listening conditions total, resulting from a 2 × 3 between-subjects design with style (clear and conversational) and listening condition (quiet, +3 dB SNR, and 0 dB SNR) as independent factors.

The test phase consisted of a two-alternative forced-choice test. In each trial, two stimuli—a word of the artificial language and a partword—were presented with a 500-milisecond interstimulus interval. There were 36 trials, formed by all possible pairings of the six words with the six partwords. The orders of the two stimuli in a trial (i.e., partword first or partword second) were counterbalanced. All the test-phase stimuli were presented in quiet. The test lasted approximately eight minutes and contained the same speaking style as the learning phase to prevent responses based on acoustic matching. E-prime 2.0 (Psychology Software Tools, 2012) controlled response recording and stimulus presentation.

2.3 Procedure

Tested individually in a sound-attenuated room, participants first listened to the six speech streams of the artificial language, presented via Sennheiser HD570 headphones. They were not given any information about the artificial language and were instructed to listen to the recordings carefully and pay as much attention as they could. They were made aware of a subsequent test and those assigned to the +3 dB and 0 dB SNR conditions were additionally warned about the noise in the speech streams. Each participant only heard the six speech streams from one style and one listening condition (e.g., clear speech in quiet or conversational speech in +3 dB SNR). After the learning phase, participants immediately proceeded to the test consisting of words and partwords in the same speaking style. They were told that they would hear two stimuli in each trial and asked to press the button labeled “1” on a response box if they thought the first stimulus was a word of the artificial language and the button labeled “2” if they thought the second one was. They had five seconds to respond after the second stimulus was presented and the next trial would be automatically initiated after they made a response or timed out. Accuracy and reaction times (RTs) were calculated for each response.

2.4 Acoustic analysis

Acoustic analyses were performed on the artificial language words to confirm that the speaker produced two distinct speaking styles. Segment boundaries were located by searching for acoustic landmarks such as nasal murmurs and stop closures in the spectrogram and waveform in Praat. When the onsets or offsets of these landmarks could not be pinpointed, as was often the case with the /la/ sequence, the mid-point in the formant transition was defined as the boundary between the segments.

We focused on the acoustic measures typically found to distinguish clear and conversational speech: consonant and vowel duration, speech rate, and vowel space area (Smiljanić, 2021; Smiljanić & Bradlow, 2009). Speech rate was calculated as the total number of the syllables of the six words (i.e., 18) divided by their total duration in seconds. Vowel space area was calculated as the triangular area formed by the average mid-point F1 and F2 values of the three vowel categories. We also measured syllable duration to examine whether our speaker produced any syllables longer, allowing listeners to use word-initial stress as a segmentation cue (Cutler & Butterfield, 1992; Cutler & Norris, 1988). We examined coarticulation, which listeners may use to facilitate word segmentation and which may differ across conversational and clear speech productions (e.g., Moon & Lindblom, 1994; Scarborough & Zellou, 2013). We quantified coarticulation using a technique introduced by Gerosa, Lee, Giuliani, and Narayanan (2006) and validated by Cychosz et al. (2019). Unlike methods such as locus equation (e.g., Lindblom & Sussman, 2012; Sussman, McCaffrey, & Matthews, 1991), which measures only F2 and can thus discard coarticulatory information in other resonant frequencies, this technique takes the whole spectrum of a segment into account and can be applied to any segment types. Following Cychosz et al. (2019), we divided each segment into frames of 25.6 milliseconds, applied short-time Fourier transformation to each frame to obtain its spectrum, and convolved a Mel-frequency filter bank over the spectrum. The results were then averaged across the frames of each segment to derive the average log Mel-frequency spectral vector, in which each value could be thought of as representing the segment's intensity in a frequency band. Next, we calculated Euclidean distances between the spectral vectors of each diphone sequence in each word and obtained an overall spectral distance for that word by taking the average of the distances. Greater distances indicate more spectral difference between adjacent segments and hence less coarticulation. This analysis was done with a custom Python script using functions from the LibROSA library for audio processing (McFee et al., 2015).

Table 1 lists the mean syllable durations (ms), speech rate (syllables per second), and the vowel space areas for clear and conversational styles. As expected, overall speaking rate was slower for clear than for conversational speech. The clear syllables were on average longer than the conversational ones.[3]

Figure 1 shows the vowel space areas formed by /a, i, u/ in two speaking styles. The clear vowels (connected by dark gray solid lines) were generally more peripheral and their vowel space area was larger than that of the conversational vowels (367,632 Hz^2 versus 191,048 Hz^2). Finally, coarticulation results showed that segments were less coarticulated in the clear speaking style: The overall spectral distances between adjacent segments in the clear artificial language words (mean: 7.25) were all greater than those of their conversational counterparts (mean: 5.71). The difference in mean overall spectral distance between the two styles (i.e., 1.54) was comparable to those found by Guo and Smiljanić's (to appear) investigation of the LUCID corpus, which ranged between 0.59 and 2.64. In sum, the clearly produced artificial language words were acoustically distinct from the conversational ones, both temporally and spectrally and in the expected direction. Similar style differences were found for the partwords (see Appendix), which were used only in

[3] Since the final syllables in the artificial language words were the longest, it is unlikely that our English listeners would perceive the word-initial syllables as stressed based on duration and apply the stress-based segmentation strategy (Cutler & Butterfield, 1992; Cutler & Norris, 1988) to locate word beginnings. Instead, it is possible that they used the longer final syllables to locate word-final positions, as with the English listeners reported in previous research (Hay & Diehl, 2007; Tyler & Cutler, 2009; White, Benavides-Varela, & Mády, 2020). Yet, as the third syllables were the longest in both clear and conversational speech, such use should improve segmentation regardless of speaking style.

Table 1: Mean syllable durations (in milliseconds) divided by syllabic position and averaged across the positions, speech rates (syllables per second), vowel space areas (VSA, Hz^2), and mean spectral distances for the artificial language words produced clearly and conversationally, along with the clear-to-conversational ratio (CL/CO) of each measure.

		Clear	Conversational	CL/CO
Duration	1st syllable	163	113	1.44
	2nd syllable	198	138	1.44
	3rd syllable	271	243	1.11
	Average	211	165	1.28
Speech rate		4.75	6.07	0.78
VSA		367,632	191,048	1.92
Spectral distance		7.25	5.71	1.27

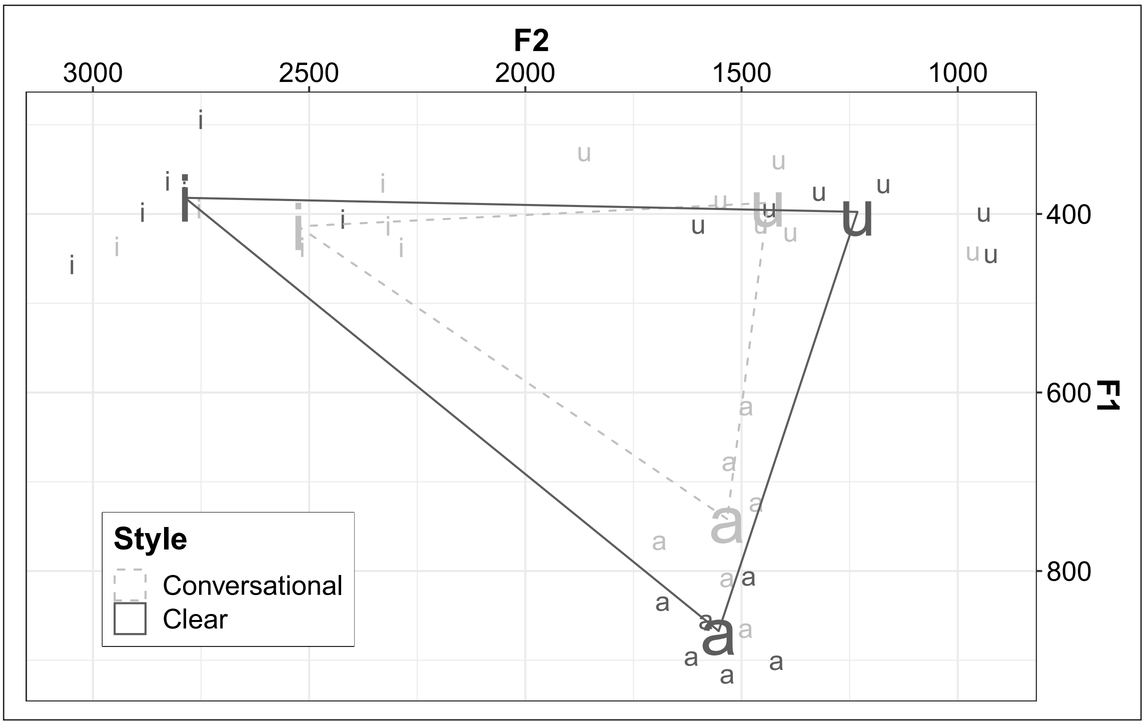

Figure 1: Vowel space areas formed by the mean mid-point F1 and F2 values (larger labels) of /a, i, u/ in the artificial language words for clear (dark gray) and conversational (light gray) speaking styles, along with F1 and F2 values of individual vowels (smaller labels).

the test. The results of acoustic analyses show that our speaker was able to implement the clear speech modifications even for nonsense words (e.g., Ferguson & Quené, 2014; Moon & Lindblom, 1994; Picheny, Durlach, & Braida, 1986; Rosen et al., 2011; Smiljanić & Bradlow, 2005).

3. Results

Responses in the test phase were analyzed. Failures to respond within the allotted time (i.e., five seconds) were recorded as time-outs and discarded. As in some artificial language learning studies (e.g., Frank, Goldwater, Griffiths, & Tenenbaum, 2010; Palmer & Mattys, 2016), participants whose accuracy rates were more than two standard deviations (*SD*s) below the mean of their condition were excluded as outliers. Four participants (2% of all participants: two from clear in quiet, one from conversational in quiet, and one from clear with noise at 0 dB SNR) were excluded. **Figure 2** shows the accuracy rates of individual participants and the boxplots for two speaking styles and all listening conditions.

A Bayesian mixed-effects logistic regression model was fitted to the responses using the brms package (Bürkner, 2017) of R (R Core Team, 2020). We opted to use the Bayesian approach as it offers several advantages over traditional frequentist methods based on null-hypothesis significance testing (Kruschke, Aguinis, & Joo, 2012; Lee, 2011; Wagenmakers et al., 2018). For example, it allows the incorporation of prior knowledge in model estimation. Its output is a probability distribution of plausible parameter values, which has a natural interpretation like "There is 95% probability that the true parameter value falls between *x* and *y*, given the model and data." As data are usually collected with the goal of obtaining insight about parameters, this is conceptually more appealing than point estimates like the *p*-value in frequentist analysis, which is the probability of the data given that the null hypothesis is true (see, inter alia, Kruschke, 2014, for more details about Bayesian methods and Vasishth, Nicenboim, Beckman, Li, & Kong, 2018, for a tutorial using brms). The data and script for the statistical analysis are available on the Open Science Framework (https://osf.io/9v7k4/).

The fixed effects of primary interest in our model were Style (baseline: conversational), Listening Condition (LisCond; baseline: quiet), and their interaction. In addition, following Ou and Guo (2021), we log-transformed and scaled response time (LogRT) and included it as a covariate to factor out any potential relationship between response latency and accuracy. The random effects were random intercepts for participant, word, and partword as well as by-word random slopes for all fixed factors and a by-partword random slope for Style. Since there were no previous attempts to investigate the impact of speaking style on statistical segmentation, we used weakly informative, regularizing priors that have a normal distribution with zero mean and *SD* of 10 on all fixed-effect coefficients. The *SD* parameters for the random effects had the same normal distribution except they included only positive values. As a standard option, we used the LKJ(2) prior (Lewandowski, Kurowicka, & Joe, 2009) on the correlation parameters between the random intercepts and slopes (Vasishth et al., 2018). The joint posterior distribution was sampled using four Markov chain Monte Carlo (MCMC) chains, each with 2,000 iterations and a warm-up of

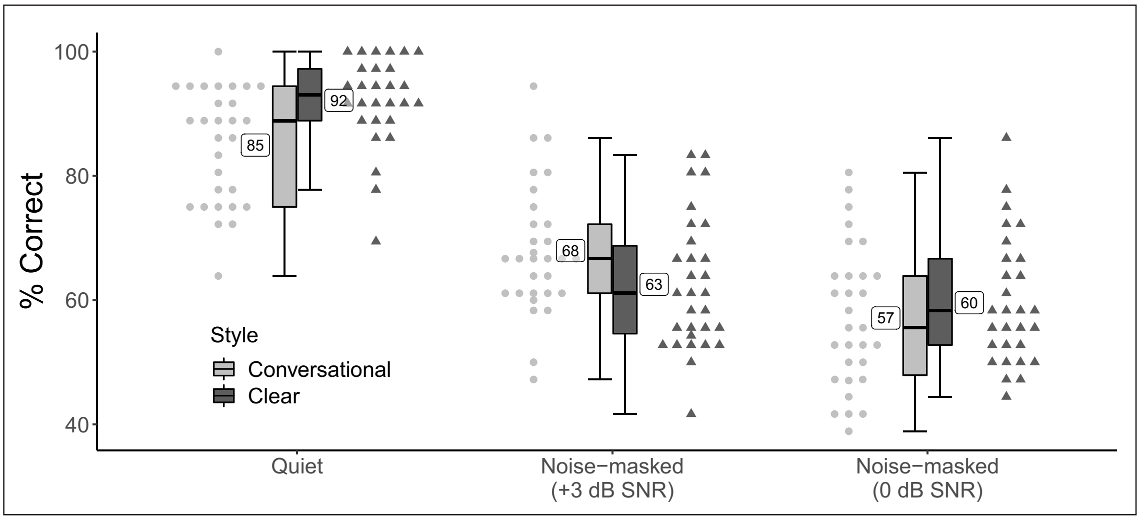

Figure 2: Accuracy rates of individual participants with boxplots for two speaking styles and three listening conditions. The horizontal line in each box represents the median and the whiskers indicate the median ± 1.5 × interquartile range. The number beside each box represents the mean accuracy.

1,000. The R-hat statistic (i.e., the ratio of between- to within-chain variance) was one for all parameters, indicating that the chains converged and could be representative of the underlying posterior distribution.

Figure 3 shows the marginal posteriors on all the fixed-effect parameters. A criterion for deciding whether a factor had a significant effect or whether two groups differ significantly in Bayesian inference is to examine the 95% credible interval (CI) of a posterior distribution—namely, the interval between the 2.5th and 97.5th percentiles of the distribution. Exclusion of zero from the interval is taken to be evidence for a significant effect or difference. For instance, the 95% CI was between –0.37 and –0.20 (mean: –0.29) for LogRT, meaning that given the model and data, there is a 95% probability that the parameter value of LogRT is between –0.37 and –0.20. The interval excluded zero and fell in the negative region, suggesting that there was an inverse relationship between response latency and accuracy: Responses with longer latencies were less likely to be correct. Such a relationship has been found in the artificial language learning experiments of Ou and Guo (2021). As they argue, this may reflect listeners' response certainty or confidence. Recall that the two stimuli in each test trial were separated by a pause. Thus, listeners might be ready to respond as soon as they heard the first stimulus and could determine whether it was a word of the artificial language or a partword. But when they were indecisive, they might think longer and guess, resulting in slower and less accurate responses.

It was also possible to draw inferences based on 95% CI for the other parameter values in the same way. Nevertheless, these raw parameter values were on a log-odds scale and the resulting inferences might not be immediately interpretable. Therefore, we chose to first compute the log-odds ratio for each condition at each iteration of the MCMC sample and

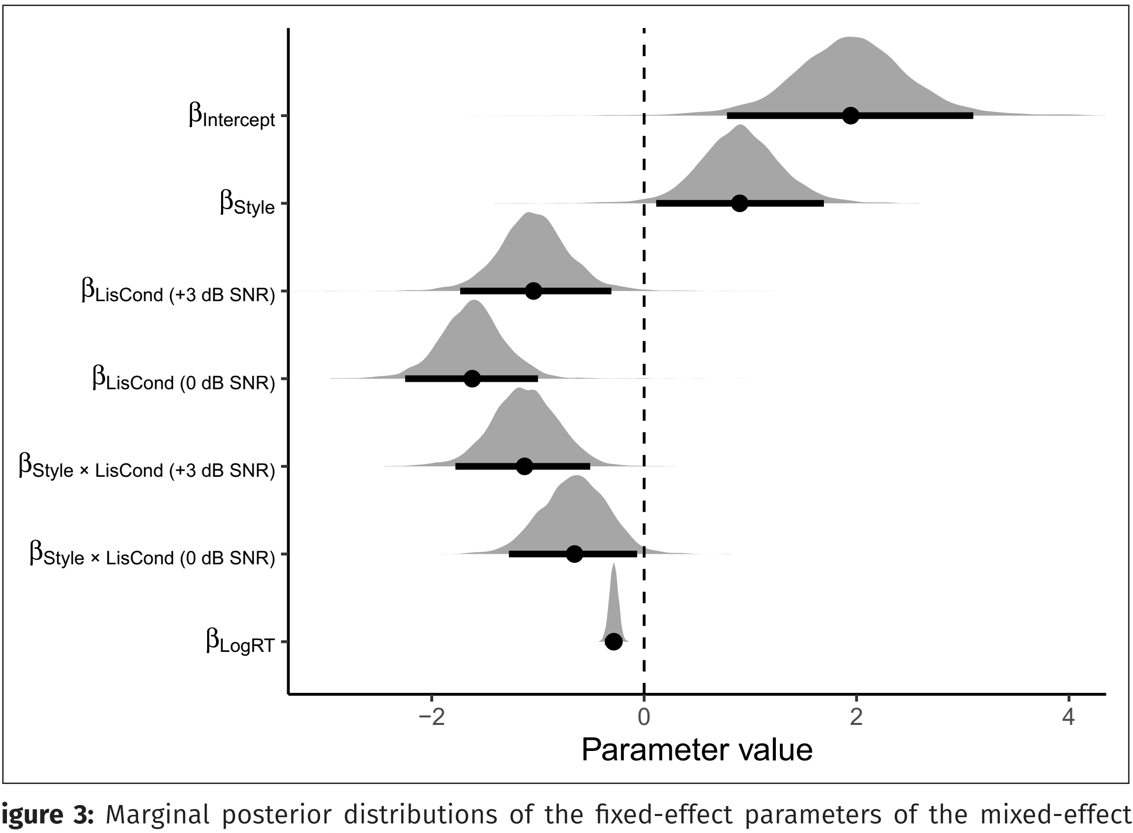

Figure 3: Marginal posterior distributions of the fixed-effect parameters of the mixed-effects logistic regression model (in which the dependent variable was the binary response outcome [correct or incorrect] in the test phase, modeled as log odds), along with means (black dots) and 95% credible intervals.

convert it to accuracy rate (in percentages) to obtain posterior distributions of accuracy for the six conditions. Next, we drew inferences about the effect of a particular factor by computing differences in posterior accuracy between the relevant conditions. Apparent from **Figure 2** is that regardless of speaking style, performance dropped substantially when the artificial language speech streams were noise-masked. This was confirmed by the posterior distributions of accuracy drop from the quiet condition, calculated by subtracting the accuracy of the quiet condition from that of a noise condition. For conversational style, the posterior accuracy drop for +3 dB SNR had a mean of –16% and a 95% CI from –33% to –3%: That is, given the model and data, there was a 95% probability that mixing the conversational artificial language speech with speech-shaped noise at +3 dB SNR would decrease performance by 3% to 33% and chance that noise had no effect or enhanced performance was very low. Since the 95% CI excluded zero, it was assumed that for the conversational style, accuracy dropped significantly from the quiet to +3 dB SNR condition. Likewise, the posterior accuracy drop for the conversational style at 0 dB SNR had a 95% CI excluding zero (mean: –29%; 95% CI: [–45%, –12%]), suggesting that noise at 0 dB SNR significantly reduced performance. Similar significant reduction in accuracy in noise was shown for the clear style at +3 dB SNR (mean: –28%; 95% CI: [–52%, –9%]) and 0 dB SNR (mean: –31%; 95% CI: [–54%, –11%]).

To further explore how the adverse effect of noise interacted with style, we computed posterior difference between the two styles in accuracy drop from quiet to each noise condition. For +3 dB SNR, the 95% CI of the posterior difference (mean: –12%; 95% CI: [–29%, 2%]) did not exclude zero, but given the model and data, there was a high (i.e., 96%) probability that the performance decrease was larger for the clear style than for the conversational one. Yet, in the 0 dB SNR condition, there was no evidence that the accuracy drop from quiet differed across the two styles (mean –2%: 95% CI: [–18%, 12%]): The probability that the drop was greater for clear speech was only 60%.

Our main objective was to examine whether clear speech improved segmentation by statistical learning relative to conversational speech and, if so, whether the clear speech segmentation benefit persisted in noise. These questions were addressed by calculating the posterior distribution of 'clear speech benefit'—which was the difference in posterior accuracy between a clear condition and a conversational one—for each level of the listening condition factor. The results are shown in **Figure 4**. The posterior clear speech benefit in quiet had a 95% CI excluding zero and falling in the positive region (mean: 7%; 95% CI: [1%, 18%]). That is, compared to hearing the conversational style, hearing the clear style significantly increased the possibility of accurate word segmentation when the artificial language speech streams were presented in quiet (mean: 7%; 95% CI: [1%, 18%]). In addition, given the data and model, the probability that the clear speech benefit was zero or negative was less than 2%. These findings support the idea that clear speech improves segmentation by statistical learning relative to conversational speech.[4] In contrast, the posterior distributions of clear speech benefit for the +3 dB SNR (mean: –5%; 95% CI: [–22%, 11%]) and 0 dB SNR (mean: 5%; 95% CI: [–13%, 22%]) were relatively flat and both their 95% CIs did not exclude zero, suggesting much uncertainty about the effect of speaking style. There is thus no evidence for a significant effect of style under either noise

[4] A sensitivity analysis was conducted to examine how the evidence for the clear speech benefit in quiet would be impacted by the choice of priors. The model presented above was rerun using four different priors for all the fixed effects: normal distributions with a mean of zero and standard deviations of 0.1, 1, 25, and 50. The resulting 95% CIs of posterior clear speech benefit in quiet for the four priors were [–4%, 5%], [0%, 18%], [1%, 18%], and [1%, 17%], respectively. That is, the 95% CIs failed to exclude zero only when we used the priors with standard deviations of 0.1 and 1, which represented a rather strong belief that the fixed effects were likely around zero.

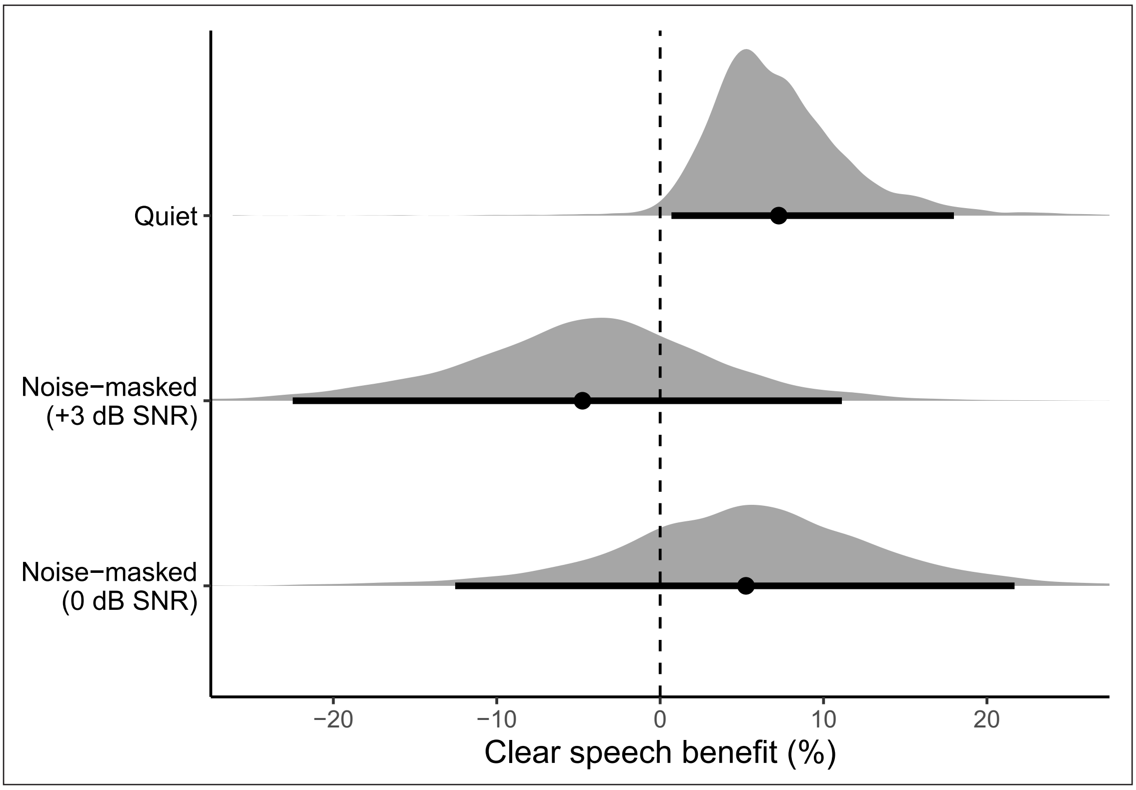

Figure 4: Posterior distributions of clear speech benefit in quiet, in noise at +3 dB SNR, and in noise at 0 dB SNR, with means (black dots) and 95% credible intervals.

condition. Finally, the posteriors of clear speech benefit of the two adverse conditions trended in opposite directions (see **Figure 4**) but the difference (mean: 10%; 95% CI: [–4%, 25%]) was not significant.

4. Discussion and conclusions

The current study investigated how speaking style variation impacted speech segmentation by statistical learning in quiet and in noise. A review of clear speech and speech segmentation research suggested that both conversational and clear speech could improve statistical segmentation, leading to competing predictions. The current work aimed to assess the contradictory predictions and to provide new evidence of the processes underlying successful segmentation. We used an artificial language learning experiment, in which English-speaking listeners recognized the words of a made-up language after hearing speech streams containing uninterrupted repetitions of the language's words. Results indicated that regardless of speaking style, recognition accuracy of artificial language words in noise was worse than that in the quiet condition, showing that noise reduces segmentation performance. This was true for both the noise levels (+ 3 and 0 dB SPL SNRs). Crucially, when presented in quiet, the clear version of the speech streams led to higher accuracy than the conversational version, supporting the prediction that clear speech improves segmentation by statistical learning. However, in noise, recognition accuracy was equivalent for both speaking styles.

The finding from the quiet condition is consistent with the clear speech perceptual benefit well-documented in the literature. A shift from conversational to clear speaking style is associated with a number of acoustic-phonetic adjustments such as enhanced consonant and vowel contrasts and has been shown to assist listeners in phoneme identification and various linguistic and cognitive processes (e.g., Bradlow & Bent, 2002; Cooke et al., 2013;

Keerstock & Smiljanić, 2019; Krause & Braida, 2002, 2004; Maniwa et al., 2009; Picheny et al., 1985). Our findings suggest that in an optimal listening situation, these adjustments are similarly beneficial for tracking statistical regularities in an unfamiliar language. Part of the clear speech benefit may be attributed to its slower speaking rate, which, as Palmer and Mattys (2016) argue, improves segmentation by statistical learning by allowing more time for syllable sequences to be stored in working memory and to be refreshed. On the other hand, conversational speech is relatively disadvantaged in terms of its support for statistical segmentation in quiet despite its more pervasive coarticulation cues.

It is important to note that conversational speech still led to rather high accuracy. Our participants reached a mean accuracy rate of 85% in the quiet conversational condition and the Bayesian mixed-effects analysis revealed a mean posterior accuracy of 86% (95% CI: [69%, 96%]) for this condition. These percentages are well above chance and relatively higher compared to those reported in many artificial language learning studies (e.g., Ordin, Polyanskaya, Laka, & Nespor, 2017; Saffran, Newport, et al., 1996; Tyler & Cutler, 2009), which were mostly below 80%. Our participants' mean accuracy is close to the mean accuracy of 82–85% observed by Fernandes et al. (2007) in the condition where the artificial language speech streams were presented in quiet with congruent coarticulatory and statistical cues. Therefore, conversational speech with its greater coarticulation supports successful segmentation, leading to an above 80% accuracy. However, clear speech provided an even greater benefit for segmentation accuracy. This suggests that, despite the less pervasive coarticulation, other acoustic-phonetic features associated with clear speech contributed significantly to segmentation. This result seems to be at odds with Fernandes et al.'s findings. However, since our speaker produced each artificial word as a whole, rather than as isolated syllables, word-internal segments were coarticulated to some extent even in clear speech. This is different from Fernandes et al.'s baseline condition in which syllables of artificial language words were concatenated without natural coarticulatory transition and against which the condition with congruent or incongruent coarticulation was compared. Further research is needed to precisely determine the relative strength of each acoustic-phonetic cue associated with speaking style variation as it contributes to segmentation.

In contrast to the quiet listening condition, the results showed equivalent segmentation performance for conversational and clear speech in noise at both SNR levels. The overall lower performance in noise is not surprising as a number of beneficial cues to segmentation are masked. In addition, noise increases listening effort, detracting cognitive resources from the segmentation task (e.g., Rönnberg et al., 2010; Zekveld et al., 2011). Unexpected though is the lack of the clear speech perceptual advantage in segmentation considering the oft-reported findings that clear speech significantly improves phoneme and word recognition relative to conversational speech in noise at these or similar SNRs, sometimes even to a greater extent as the SNR becomes more challenging (e.g., Bradlow et al., 2003; Ferguson & Kewley-Port, 2007; Payton et al., 1994; Picheny et al., 1986; Smiljanić & Bradlow, 2005). It is possible that the increased cognitive load in noise depleted the processing resources in a way that the slower speaking rate in clear speech became less beneficial (cf. Palmer & Mattys, 2016), though this possibility remains speculative as the drop in accuracy from the quiet to noise conditions was not significantly greater for clear speech.

Another factor that could play a role in the observed lack of clear speech benefit relates to the relative weight of prosody as a segmentation cue in noise. Recall that our stimuli had monotone F0 contours, neutralizing pitch cues for the listeners. In addition to slower speaking rate and enhanced phonemic categories, clear speech is characterized by an increase in pitch average and range (e.g., Bradlow et al., 2003; Picheny et al., 1986; Smiljanić & Bradlow, 2005). When speech is degraded by noise, beneficial segmental cues are masked and their role in segmentation may be diminished. In contrast, prosodic cues like stress become particularly robust and useful in the adverse listening conditions. English listeners are known to treat stressed syllables as word beginnings (stress-based segmentation: e.g., Cutler & Norris, 1988; Tyler & Cutler, 2009), reflecting the fact that most English words are stressed initially (Cutler & Carter, 1987). This segmentation strategy seems to play a role when the listening conditions are optimal. According to Mattys, White, and Melhorn's (2005) model, descending weights are assigned to lexical, segmental (e.g., coarticulation), and prosodic/metrical cues in speech segmentation. The hierarchy is not fixed as listeners dynamically adjust cue weights in response to the listening condition. In noise-degraded speech, listeners rely on prosodic information, if well-preserved, more than on segmental information (Mattys & Bortfeld, 2017). Since our listeners could not rely on lexical knowledge in the artificial language learning tasks, it is possible that stress could have played an important role in the noise-masked conditions (cf., Fry, 1958; Gay, 1978; Lieberman, 1960; Sluijter & van Heuven, 1996). However, in our stimuli, vowels in the initial syllables were not longer and the words were synthesized with flattened F0, thus eliminating stress-based segmentation (see note 2). As F0 contours are more exaggerated in clear speech, F0 flattening may have affected segmentation in clear speech more than in conversational speech precisely in the listening conditions that would have favored prosodic cues. Future work should examine whether the presence of exaggerated stress cues would contribute to better segmentation performance for clear speech relative to conversational speech in noise. Another important future direction is to determine how noise impacts speech signal in combination with its effect on the domain-general processing resources to determine speech segmentation.

The present study showed that speaking clearly relative to conversationally improves segmentation by statistical learning under optimal listening conditions. This finding suggests that as far as listening in quiet is concerned, a shift towards a hyper-articulated listener-oriented speaking style improves not only recognition of meaningful words or sentences, but also processes that depend on domain-general cognitive abilities such as tracking statistical regularities. Listeners are able to use acoustic-phonetic enhancements to track syllable co-occurrences and improve segmentation by statistical learning. It is important though to acknowledge the limitations of the current study. The clear speech benefit for segmentation found here was rather small. It is therefore crucial to replicate these findings with a different group of listeners with the same language profile. Similarly, it cannot be taken for granted that the current results would generalize to more naturalistic communicative contexts or to other 'clear speeches.' Scarborough and colleagues (Scarborough et al., 2007) found that speech to a real listener (e.g., a foreigner) was faster with less expanded vowel space than that to an imagined one. In a later study (Scarborough & Zellou, 2013), they showed that relative to read clear speech, real listener-directed speech had greater vowel-nasal coarticulation. As mentioned, however, Guo and Smiljanić's (to appear) whole-spectrum analysis suggested that clear speech exhibited coarticulatory resistance when communicative barriers were explicitly imposed. Real listener-directed speaking styles also differ in F0 and vowel space or formant range depending on the interlocutor (Burnham et al., 2002; Uther et al., 2007) or the nature of the communicative barrier (e.g., vocoded versus mixed with talker babble) (Hazan & Baker, 2011). It is important then to see how the speaking style variation encountered in the real-world communicative situations contributes to speech segmentation. Our speculation is that the various acoustic-phonetic enhancements consistently found across hyper-articulated clear speaking styles (e.g., slower speaking rate) will improve segmentation by statistical learning relative to the casual hypoarticulated forms, though the effect size may vary across clear speaking styles elicited in different ways.

It is also of interest to examine whether the clear speech segmentation benefit observed at least in quiet extends to the listeners who are learning another language later in life. Clear speech enhancements reflect language-specific properties and gain in intelligibility is greater for the listeners who acquired the target language from birth than for the language learners (Bradlow & Alexander, 2007; Bradlow & Bent, 2002). Using signal-driven cues for segmentation is in part a language-specific skill that requires proficiency or knowledge in the target language (Tremblay & Broersma, 2019; Tremblay, Coughlin, Bahler, & Gaillard, 2012). It is possible that language learners' segmentation by statistical learning is enhanced through the language-independent clear speech enhancements such as slower speaking rate and exaggerated F0 contours. However, they may not be able to take advantage of other signal-driven cues such as vowel and consonant contrast enhancements, which might be available only to those with extensive experience with the sound patterns of the target language (Smiljanić & Bradlow, 2011). Finally, we need to extend this line of work to examine whether speech segmentation of real words is enhanced by clear speech in quiet and in noise differently than what was found for artificial language learning.

Additional File

The additional file for this article can be found as follows:

- **Appendix.** PDF file containing the acoustic analyses of the partwords. DOI: https://doi.org/10.5334/labphon.310.s1

Acknowledgements

We would like to thank the research assistants at the UT Sound Lab for their assistance with collecting data.

Competing Interests

The authors have no competing interests to declare.

References

Baker, R., & Hazan, V. (2011). DiapixUK: Task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods*, *43*(3), 761–770. DOI: https://doi.org/10.3758/s13428-011-0075-y

Behrman, A., Ferguson, S. H., Akhund, A., & Moeyaert, M. (2019). The effect of clear speech on temporal metrics of rhythm in Spanish-accented speakers of English. *Language and Speech*, *62*(1), 5–29. DOI: https://doi.org/10.1177/0023830917737109

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer [Computer program]*. http://www.praat.org/

Bradlow, A. R. (2002). Confluent talker- and listener-oriented forces in clear speech production. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 241–273). Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110197105.241

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. DOI: https://doi.org/10.1121/1.2642103

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. DOI: https://doi.org/10.1121/1.1487837

Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking clearly for children with learning disabilities: Sentence perception in noise. *Journal of Speech, Language, and Hearing Research*, *46*(1), 80–97. DOI: https://doi.org/10.1044/1092-4388(2003/007)

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1). DOI: https://doi.org/10.18637/jss.v080.i01

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science*, *296*(5572), 1435. DOI: https://doi.org/10.1126/science.1069587

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, *24*(2), 209–244. DOI: https://doi.org/10.1006/jpho.1996.0012

Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics*, *26*(2), 173–199. DOI: https://doi.org/10.1006/jpho.1998.0071

Cheng, L. S. P., Burgess, D., Vernooij, N., Solís-Barroso, C., McDermott, A., & Namboodiripad, S. (2021). Problematizing the native speaker in Psycholinguistics: Replacing vague and harmful terminology with inclusive and accurate measures. *PsyArXiv*. DOI: https://doi.org/10.31234/osf.io/23rmx

Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics*, *32*(2), 141–176. DOI: https://doi.org/10.1016/S0095-4470(03)00043-3

Cooke, M., Mayo, C., Valentini-Botinhao, C., Stylianou, Y., Sauert, B., & Tang, Y. (2013). Evaluating the intelligibility benefit of speech modifications in known noise conditions. *Speech Communication*, *55*(4), 572–585. DOI: https://doi.org/10.1016/j.specom.2013.01.001

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. The MIT Press. DOI: https://doi.org/10.7551/mitpress/9012.001.0001

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*(2), 218–236. DOI: https://doi.org/10.1016/0749-596X(92)90012-M

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, *2*(3–4), 133–142. DOI: https://doi.org/10.1016/0885-2308(87)90004-0

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 113–121. DOI: https://doi.org/10.1037/0096-1523.14.1.113

Cychosz, M., Edwards, J. R., Munson, B., & Johnson, K. (2019). Spectral and temporal measures of coarticulation in child speech. *The Journal of the Acoustical Society of America*, *146*(6), EL516–EL522. DOI: https://doi.org/10.1121/1.5139201

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 218–244. DOI: https://doi.org/10.1037/0096-1523.28.1.218

Duez, D. (1992). Second formant locus-nucleus patterns: An investigation of spontaneouos French speech. *Speech Communication*, *11*(4–5), 417–427. DOI: https://doi.org/10.1016/0167-6393(92)90047-B

Emberson, L. L., Conway, C. M., & Christiansen, M. H. (2011). Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Quarterly Journal of Experimental Psychology*, *64*(5), 1021–1040. DOI: https://doi.org/10.1080/17470218.2010.538972

Endress, A. D., & Hauser, M. D. (2010). Word segmentation with universal prosodic cues. *Cognitive Psychology*, *61*(2), 177–199. DOI: https://doi.org/10.1016/j.cogpsych.2010.05.001

Farnetani, E., & Recasens, D. (2010). Coarticulation and connected speech processes. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences*. Second Edition (pp. 316–352). Blackwell Publishing. DOI: https://doi.org/10.1002/9781444317251.ch9

Ferguson, S. H. (2012). Talker differences in clear and conversational speech: Vowel intelligibility for older adults with hearing loss. *Journal of Speech, Language, and Hearing Research, 55*(3), 779–790. DOI: https://doi.org/10.1044/1092-4388(2011/10-0342)

Ferguson, S. H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America, 112*(1), 259–271. DOI: https://doi.org/10.1121/1.1482078

Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research, 50*(5), 1241–1255. DOI: https://doi.org/10.1044/1092-4388(2007/087)

Ferguson, S. H., & Quené, H. (2014). Acoustic correlates of vowel intelligibility in clear and conversational speech for young normal-hearing and elderly hearing-impaired listeners. *The Journal of the Acoustical Society of America, 135*(6), 3570–3584. DOI: https://doi.org/10.1121/1.4874596

Fernandes, T., Kolinsky, R., & Ventura, P. (2010). The impact of attention load on the use of statistical information and coarticulation as speech segmentation cues. *Attention, Perception, & Psychophysics*, *72*(6), 1522–1532. DOI: https://doi.org/10.3758/APP.72.6.1522

Fernandes, T., Ventura, P., & Kolinsky, R. (2007). Statistical information and coarticulation as cues to word boundaries: A matter of signal quality. *Perception and Psychophysics, 69*(6), 856–864. DOI: https://doi.org/10.3758/BF03193922

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America*, *101*(6), 3728–3740. DOI: https://doi.org/10.1121/1.418332

Francis, A. L., Love, J., & Boutin, M. (2019). Does noise sensitivity mediate physiological measures of listening effort? *The Journal of the Acoustical Society of America*, *146*(4), 3051–3051. DOI: https://doi.org/10.1121/1.5137574

Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition*, *117*(2), 107–125. DOI: https://doi.org/10.1016/j.cognition.2010.07.005

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech, 1*(2), 126–152. DOI: https://doi.org/10.1177/002383095800100207

Gay, T. (1978). Physiological and acoustic correlates of perceived stress. *Language and Speech, 21*(4), 347–353. DOI: https://doi.org/10.1177/002383097802100409

Gerosa, M., Lee, S., Giuliani, D., & Narayanan, S. (2006). Analyzing children's speech: An acoustic study of consonants and consonant-vowel transition. *2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings*, 393–396. DOI: https://doi.org/10.1109/ICASSP.2006.1660040

Gilbert, R. C., Chandrasekaran, B., & Smiljanic, R. (2014). Recognition memory in noise for speech of varying intelligibility. *The Journal of the Acoustical Society of America*, *135*(1), 389–399. DOI: https://doi.org/10.1121/1.4838975

Grammon, D., & Babel, A. (2021). *What does "native speaker" mean, anyway?* https://languagelog.ldc.upenn.edu/nll/?p = 51108

Granlund, S., Hazan, V., & Baker, R. (2012). An acoustic-phonetic comparison of the clear speaking styles of Finnish-English late bilinguals. *Journal of Phonetics*, *40*(3), 509–520. DOI: https://doi.org/10.1016/j.wocn.2012.02.006

Guo, Z.-C., & Smiljanić, R. (to appear). Speakers coarticulate less when facing real and imagined communicative difficulties: An analysis of read and spontaneous speech from the LUCID corpus. *Proceedings of INTERSPEECH 2021*.

Hay, J. S. F., & Diehl, R. L. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception and Psychophysics*, *69*(1), 113–122. DOI: https://doi.org/10.3758/BF03194458

Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *The Journal of the Acoustical Society of America*, *130*(4), 2139–2152. DOI: https://doi.org/10.1121/1.3623753

Keerstock, S., & Smiljanić, R. (2018). Effects of intelligibility on within- and cross-modal sentence recognition memory for native and non-native listeners. *The Journal of the Acoustical Society of America*, *144*(5), 2871–2881. DOI: https://doi.org/10.1121/1.5078589

Keerstock, S., & Smiljanić, R. (2019). Clear speech improves listeners' recall. *The Journal of the Acoustical Society of America*, *146*(6), 4604–4610. DOI: https://doi.org/10.1121/1.5141372

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), 4–5. DOI: https://doi.org/10.1016/S0010-0277(02)00004-5

Krause, J. C., & Braida, L. D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *The Journal of the Acoustical Society of America*, *112*(5), 2165–2172. DOI: https://doi.org/10.1121/1.1509432

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*, *115*(1), 362–378. DOI: https://doi.org/10.1121/1.1635842

Krull, D. (1989). Consonant-vowel coarticulation in spontaneous speech and in reference words. *Speech Transmission Laboratory Quarterly Progress and Status Report*, *30*(1), 101–105.

Kruschke, J. K. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press. DOI: https://doi.org/10.1016/B978-0-12-405888-0.00008-8

Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods*, *15*(4), 722–752. DOI: https://doi.org/10.1177/1094428112457829

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology*, *55*(1), 1–7. DOI: https://doi.org/10.1016/j.jmp.2010.08.013

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001. DOI: https://doi.org/10.1016/j.jmva.2009.04.008

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America*, *32*(4), 451–454. DOI: https://doi.org/10.1121/1.1908095

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H& H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403–439). Netherlands: Springer. DOI: https://doi.org/10.1007/978-94-009-2037-8_16

Lindblom, B., & Sussman, H. M. (2012). Dissecting coarticulation: How locus equations happen. *Journal of Phonetics*, *40*(1), 1–19. DOI: https://doi.org/10.1016/j.wocn.2011.09.005

Liu, S., Del Rio, E., Bradlow, A. R., & Zeng, F.-G. (2004). Clear speech perception in acoustic and electric hearing. *The Journal of the Acoustical Society of America*, *116*(4), 2374–2383. DOI: https://doi.org/10.1121/1.1787528

Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *The Journal of the Acoustical Society of America*, *125*(6), 3962–3973. DOI: https://doi.org/10.1121/1.2990715

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967. DOI: https://doi.org/10.1044/1092-4388(2007/067)

Matthies, M., Perrier, P., Perkell, J. S., & Zandipour, M. (2001). Variation in anticipatory coarticulation with changes in clarity and rate. *Journal of Speech, Language, and Hearing Research*, *44*(2), 340–353. DOI: https://doi.org/10.1044/1092-4388(2001/028)

Mattys, S. L., & Bortfeld, H. (2017). Speech segmentation. In M. G. Gaskell & J. Mirković (Eds.), *Speech Perception and Spoken Word Recognition* (pp. 55–75). Routledge.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, *134*(4), 477–500. DOI: https://doi.org/10.1037/0096-3445.134.4.477

McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *58*(1), 22–33. DOI: https://doi.org/10.1080/02724980443000151

McFee, B., Raffel, C., Liang, D., Ellis, D. P. W., Mcvicar, M., Battenberg, E., & Nieto, O. (2015). Librosa – audio processing Python library. *Proceedings of the 14th Python in Science Conference*, 18–25. http://conference.scipy.org/proceedings/scipy2015/pdfs/brian_mcfee.pdf

Miles, C., Jones, D. M., & Madden, C. A. (1991). Locus of the irrelevant speech effect in short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(3), 578–584. DOI: https://doi.org/10.1037/0278-7393.17.3.578

Moon, S. J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical Society of America*, *96*(1), 40–55. DOI: https://doi.org/10.1121/1.410492

Ordin, M., Polyanskaya, L., Laka, I., & Nespor, M. (2017). Cross-linguistic differences in the use of durational cues for the segmentation of a novel language. *Memory and Cognition*, *45*(5), 863–876. DOI: https://doi.org/10.3758/s13421-017-0700-9

Ou, S.-C., & Guo, Z.-C. (2021). The language-specific use of fundamental frequency rise in segmentation of an artificial language: Evidence from listeners of Taiwanese Southern Min. *Language and Speech*, *64*(2), 437–466. DOI: https://doi.org/10.1177/0023830919886604

Palmer, S. D., & Mattys, S. L. (2016). Speech segmentation by statistical learning is supported by domain-general processes within working memory. *Quarterly Journal of Experimental Psychology*, *69*(12), 2390–2401. DOI: https://doi.org/10.1080/17470218.2015.1112825

Payton, K. L., Uchanski, R. M., & Braida, L. D. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *The Journal of the Acoustical Society of America*, *95*(3), 1581–1592. DOI: https://doi.org/10.1121/1.408545

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing*, *39*(2), 204–214. DOI: https://doi.org/10.1097/AUD.0000000000000494

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing. I. Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*, *28*(1), 96–103. DOI: https://doi.org/10.1044/jshr.2801.96

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing. II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, *29*(4), 434–446. DOI: https://doi.org/10.1044/jshr.2904.434

Pichora-Fuller, M. K., Goy, H., & Van Lieshout, P. (2010). Effect on speech intelligibility of changes in speech production influenced by instructions and communication environments. *Seminars in Hearing, 31*(2), 77–94. DOI: https://doi.org/10.1055/s-0030-1252100

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing, 37*, 5S-27S. DOI: https://doi.org/10.1097/AUD.0000000000000312

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, *97*(1), 593–608. DOI: https://doi.org/10.1121/1.412282

Poch-Olivé, D., Fernandez-Guitierrez, N., & Martinez-Dauden, G. (1989). Some problems of coarticulation in CV stop syllables in Spanish and Catalan spontaneous speech. *Proceedings of Speech Resarch '89*, 111–115.

Psychology Software Tools. (2012). *E-Prime 2.0*.

R Core Team. (2020). *R: A language and environment for statistical computing* (4.0.1). R Foundation for Statistical Computing. https://www.r-project.org/

Rabbitt, P. M. A. (1968). Channel-capacity, intelligibility and immediate memory. *The Quarterly Journal of Experimental Psychology*, *20*(3), 241–248. DOI: https://doi.org/10.1080/14640746808400158

Rabbitt, P. M. A. (1991). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Oto-Laryngologica*, *111*(S476), 167–176. DOI: https://doi.org/10.3109/00016489109127274

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*, 1–17. DOI: https://doi.org/10.3389/fnsys.2013.00031

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology*, *47*(sup2), S99–S105. DOI: https://doi.org/10.1080/14992020802301167

Rönnberg, J., Rudner, M., Lunner, T., & Zekveld, A. (2010). When cognition kicks in: Working memory and speech understanding in noise. *Noise and Health*, *12*(49), 263. DOI: https://doi.org/10.4103/1463-1741.70505

Rosen, K. M., Folker, J. E., Murdoch, B. E., Vogel, A. P., Cahill, L. M., Delatycki, M. B., & Corben, L. A. (2011). Measures of spectral change and their application to habitual, slow, and clear speaking modes. *International Journal of Speech-Language Pathology*, *13*(2), 165–173. DOI: https://doi.org/10.3109/17549507.2011.529939

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. DOI: https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606–621. DOI: https://doi.org/10.1006/jmla.1996.0032

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and Learning) out of the corner of your ear. *Psychological Science*, *8*(2), 101–105. DOI: https://doi.org/10.1111/j.1467-9280.1997.tb00690.x

Scarborough, R., Brenier, J., Zhao, Y., Hall-Lew, L., & Dmitrieva, O. (2007). An acoustic study of real and imagined foreigner-directed speech. *Proceedings of the 15th International Congress of Phonetic Sciences*, 2165–2168. DOI: https://doi.org/10.1121/1.4781735

Scarborough, R., & Zellou, G. (2013). Clarity in communication: "Clear" speech authenticity and lexical neighborhood density effects in speech production and perception. *The Journal of the Acoustical Society of America*, *134*(5), 3793–3807. DOI: https://doi.org/10.1121/1.4824120

Schneider, E. N., Bernarding, C., Francis, A. L., Hornsby, B. W. Y., & Strauss, D. J. (2019). A quantitative model of listening related fatigue. *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, 619–622. DOI: https://doi.org/10.1109/NER.2019.8717046

Schum, D. J. (1996). Intelligibility of clear and conversational speech of young and elderly talkers. *Journal of the American Academy of Audiology*, *7*(3), 212–218.

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, *54*(1), 1–32. DOI: https://doi.org/10.1016/j.cogpsych.2006.04.002

Sluijter, A. M. C., & van Heuven, V. J. (1996). Acoustic correlates of linguistic stress and accent in Dutch and American English. *Proceedings of the Fourth International Conference on Spoken Language Processing (ICSLP)*, 630–633. DOI: https://doi.org/10.1109/ICSLP.1996.607440

Smiljanić, R. (2021). Clear speech perception. In L. C. Nygaard, J. Pardo, D. Pisoni, & R. Remez (Eds.), *The Handbook of Speech Perception* (2nd ed., pp. 177–205). Wiley. DOI: https://doi.org/10.1002/9781119184096.ch7

Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America*, *118*(3 Pt 1), 1677–1688. DOI: https://doi.org/10.1121/1.2000788

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

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. DOI: https://doi.org/10.1121/1.3652882

Sussman, H. M., McCaffrey, H. A., & Matthews, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *The Journal of the Acoustical Society of America*, *90*(3), 1309–1325. DOI: https://doi.org/10.1121/1.401923

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716. DOI: https://doi.org/10.1037/0012-1649.39.4.706

Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development*, *3*(1), 73–100. DOI: https://doi.org/10.1080/15475440709337001

Tremblay, A., & Broersma, M. (2019). Foreign-language knowledge enhances artificial-language segmentation. *Interspeech 2019*, 2658–2662. DOI: https://doi.org/10.21437/Interspeech.2019-2446

Tremblay, A., Coughlin, C. E., Bahler, C., & Gaillard, S. (2012). Differential contribution of prosodic cues in the native and non-native segmentation of French speech. *Laboratory Phonology*, *3*(2), 385–423. DOI: https://doi.org/10.1515/lp-2012-0018

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552–564. DOI: https://doi.org/10.1037/0096-3445.134.4.552

Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *The Journal of the Acoustical Society of America*, *126*(1), 367–376. DOI: https://doi.org/10.1121/1.3129127

Uchanski, R. M. (2005). Clear speech. In D. Pisoni & R. Remez (Eds.), *The Handbook of Speech Perception* (pp. 207–235). Blackwell. DOI: https://doi.org/10.1002/9780470757024.ch9

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., & Durlach, N. I. (1996). Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech, Language, and Hearing Research*, *39*(3), 494–509. DOI: https://doi.org/10.1044/jshr.3903.494

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication*, *49*(1), 2–7. DOI: https://doi.org/10.1016/j.specom.2006.10.003

van der Feest, S. V. H., Blanco, C. P., & Smiljanic, R. (2019). Influence of speaking style adaptations and semantic context on the time course of word recognition in quiet and in noise. *Journal of Phonetics*, *73*, 158–177. DOI: https://doi.org/10.1016/j.wocn.2019.01.003

Van Engen, K. J. (2017). Clear speech and lexical competition in younger and older adult listeners. *The Journal of the Acoustical Society of America*, *142*(2), 1067–1077. DOI: https://doi.org/10.1121/1.4998708

Van Engen, K. J., Chandrasekaran, B., & Smiljanić, R. (2012). Effects of Speech Clarity on Recognition Memory for Spoken Sentences. *PLoS ONE*, *7*(9), e43753. DOI: https://doi.org/10.1371/journal.pone.0043753

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, *8*, 1–4. DOI: https://doi.org/10.3389/fnhum.2014.00577

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, *71*, 147–161. DOI: https://doi.org/10.1016/j.wocn.2018.07.008

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin and Review*, *25*(1), 35–57. DOI: https://doi.org/10.3758/s13423-017-1343-3

White, L., Benavides-Varela, S., & Mády, K. (2020). Are initial-consonant lengthening and final-vowel lengthening both universal word segmentation cues? *Journal of Phonetics*, *81*, 100982. DOI: https://doi.org/10.1016/j.wocn.2020.100982

Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear and Hearing*, *36*(4), e153–e165. DOI: https://doi.org/10.1097/AUD.0000000000000145

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and Hearing*, *31*(4), 480–490. DOI: https://doi.org/10.1097/AUD.0b013e3181d4f251

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing*, *32*(4), 498–510. DOI: https://doi.org/10.1097/AUD.0b013e31820512bb

How to cite this article: Guo, Z.-C., & Smiljanic, R. 2021 Speaking clearly improves speech segmentation by statistical learning under optimal listening conditions. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 12(1): 14, pp. 1–24. DOI: https://doi.org/10.5334/labphon.310

Submitted: 11 October 2020 **Accepted:** 01 July 2021 **Published:** 23 July 2021

]u[*Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS