Journal of Phonetics 73 (2019) 158–177

Contents lists available at ScienceDirect

Journal of Phonetics

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

Research Article

Influence of speaking style adaptations and semantic context on the time course of word recognition in quiet and in noise

Suzanne V.H. van der Feest [a,b,*], Cynthia P. Blanco [a,c], Rajka Smiljanic [a]

[a] Department of Linguistics, The University of Texas at Austin, 305 E. 23rd Street B5100, Austin, TX 78712, USA
[b] Linguistics Program, The Graduate Center, City University of New York, 365 Fifth Avenue, Room 7407, New York, NY 10016, USA
[c] Duolingo, 5900 Penn Avenue, Pittsburgh, PA 15206, USA

ARTICLE INFO

Article history:
Received 22 January 2018
Received in revised form 23 December 2018
Accepted 25 January 2019

Keywords:
Word recognition
Speech perception in noise
Clear Speech
Infant Directed Speech
Eye-tracking

ABSTRACT

This study examines the effects of different listener-oriented speaking styles and semantic contexts on online spoken word recognition using eyetracking. In Experiment 1, different groups of listeners participated in a word-identification-in-noise and in a pleasantness-rating task. Listeners heard sentences with high- versus low-predictability semantic contexts produced in infant-directed speech, Clear Speech, and Conversational Speech. Experiment 2 (in silence) and 3 (in noise) investigated the time course of visual fixations to target objects when participants were listening to different speaking styles and contexts. Results from all experiments show that relative to conversational speech, both infant-directed speech and Clear Speech improved word recognition for high-predictability sentences, in quiet as well as in noise. This indicates that established advantages of infant-directed speech for young listeners cannot be attributed only to affect; the acoustic enhancements in infant-directed speech benefit adult speech processing as well. Furthermore, in silence (Experiment 2) lexical access was facilitated by contextual cues even in conversational speech; but in noise (Experiment 3) listeners reliably focused the target only when a combination of contextual cues and listener-adapted acoustic–phonetic cues were available. These findings suggest that both semantic cues and listener-oriented acoustic enhancements are needed to facilitate word recognition, especially in adverse listening conditions.

1. Introduction

Understanding spoken language requires the conversion of sound to meaning. During this process, listeners need to map the quickly-evolving and variable speech stream onto the lexicon, where a number of viable candidate words are activated and compete for selection (Gaskell & Marslen-Wilson, 2002; Luce & Pisoni, 1998; Marslen-Wilson & Zwitserlood, 1989; McClelland & Elman, 1986; McQueen, 2007; Norris, 1994). Recognizing spoken words is even more challenging in noise (Mattys, Brooks, & Cooke, 2009; Mattys, Davis, Bradlow, & Scott, 2012), which can lead to reduced attentional and memory capacities and can result in word recognition failure (Assmann & Summerfield, 2004; Pichora-Fuller, Schneider, & Daneman, 1995; Rönnberg, Rudner, Foo, & Lunner, 2008).

The time course of lexical access is tied to the unfolding of the speech signal; nonetheless, lexical activation and recognition can occur prior to the offset of the target word when words are clearly enunciated (e.g. Allopenna, Magnuson, & Tanenhaus, 1998; Grosjean, 1980; Marslen-Wilson, 1984; see Dahan, 2010 for an overview). In the case of reduced conversational speech then, listeners often need to hear speech following the target word offset for word identification to occur, presumably to account for its hypo-articulated characteristics (Bard, Shillcock, & Altmann, 1988; Bard, Sotillo, Kelly, & Aylett, 2001). Noise can also alter the time course of spoken word processing. Pre-offset identification is even less likely to occur for spoken words presented in noise (Orfanidou, Davis, Ford, & Marslen-Wilson, 2011), and more lexical competition was found from onset competitors in noise than in quiet (Ben-David et al., 2010). In Brouwer and Bradlow (2015), listeners looked more to phonological competitors when presented with words in noise compared to words in quiet in a visual-world task, and Hintz and Scharenborg (2016) found

* Corresponding author at: Linguistics Program, The Graduate Center, City University of New York, 365 Fifth Avenue, Room 7407, New York NY 10016, USA.
E-mail address: svanderfeest@gc.cuny.edu (S.V.H. van der Feest).

https://doi.org/10.1016/j.wocn.2019.01.003
0095-4470/

delayed fixations to targets and phonological competitors in noise compared to in quiet. Furthermore, in noise listeners fixated the target picture less and showed fixations to phonological competitors for longer after hearing the target word. This reveals a processing cost in noise even for words that were identified correctly in the end. Results from these studies show that the dynamics of spoken word recognition are affected by speech clarity and by noise. However, no work to date tested whether deliberate, listener-oriented speaking style modifications and semantic context modulate the time course of spoken word recognition.

The present study focuses on two such speaking style modifications: Clear Speech and Infant-Directed Speech (IDS). It is well-documented that Clear Speech improves word recognition in noise for a number of listener groups: adult listeners with hearing impairment (e.g. Ferguson & Kewley-Port, 2002; Picheny, Durlach, & Braida, 1985, 1986), children with cochlear implants (Smiljanic & Sladen, 2013) and learning disabilities (Bradlow, Kraus, & Hayes, 2003), as well as non-native listeners (Bradlow & Bent, 2002; Smiljanić & Bradlow, 2005). This Clear Speech processing benefit has been extended to recognition memory for spoken sentences in quiet and in noise (Gilbert, Chandrasekaran, & Smiljanic, 2014; Keerstock & Smiljanic, 2018; Van Engen, Chandrasekaran, & Smiljanic, 2012) and to reduced lexical competition (Van Engen, 2017). Clear Speech typically involves speaking more slowly and more loudly and producing more salient stop releases, an expanded vowel space, greater pitch variation, and increased energy in 1000–3000 Hz range of long-term spectra (Ferguson & Kewley-Port, 2002; Maniwa, Jongman, & Wade, 2009; Picheny, Durlach, & Braida, 1986; Smiljanić & Bradlow, 2005; Smiljanic & Gilbert, 2017). Like Clear Speech, IDS is also a listener-oriented speaking style, produced by talkers when they are addressing young children (e.g. Cristia, 2013; Johnson, Lahey, Ernestus, & Cutler, 2013). It has been argued that IDS aids various aspects of social-emotional and affective development in young children (e.g. Cristia, 2013; Kaplan, Goldstein, Huckeby, & Panneton Cooper, 1995; Werker & McLeod, 1989). There is also evidence that IDS facilitates overall language development (Fernald, 1984, 1989; Kemler-Nelson, Hirsh-Pasek, Jusczyk, & Cassidy, 1989; Kuhl, 2007; Soderstrom, 2007; Werker et al., 2007). Specifically, it has been argued that it facilitates the creation of perceptual sound categories (Cristia & Seidl, 2014; Kuhl et al., 1997; Kuhl, 2007), sound discrimination (Liu, Kuhl, & Tsao, 2003), speech segmentation (e.g. Schreiner & Mani, 2017), and word learning (Graf Estes & Hurley, 2013; Song, Demuth, & Morgan, 2010). Many acoustic–phonetic adjustments described as typical of IDS are similar to those found in Clear Speech. They include suprasegmental changes, such as more frequent pauses, a wider pitch range, more prosodic repetitions, and slower overall speaking rate (Cooper & Aslin, 1990, 1994; Cristia, 2013; Fernald & Mazzie, 1991; Knoll, Scharrer, & Costall, 2009; Wang, Seidl, & Cristia, 2016; Grieser & Kuhl, 1988; but see also Martin, Igarashi, Jincho, & Mazuka, 2016, who argue the slower speaking rate may be mostly attributed to shorter average utterance length). They also include segmental modifications, like more pronounced voicing contrasts, a stretching of the vowel space (the point vowels, specifically), phonetic enhancement of sibilants as well as the maintenance of the connected speech assimilation processes (Cristià, 2010; Englund, 2005; Kuhl et al., 1997; Sundberg & Lacerda, 1999; Fish, García-Sierra, Ramírez-Esparza, & Kuhl, 2017; Buckler, Goy, & Johnson, 2018). Recently, a shift of vocal timbre in IDS has also been reported (Piazza, Iordan, & Lew-Williams, 2017). A number of studies have cast doubt on the contributions of these acoustic-articulatory modifications to improved word learning or recognition in children (for an overview, see Golinkoff, Can, Soderstrom, & Hirsh-Pasek, 2015; Soderstrom, 2007; Eaves, Feldman, Griffiths, & Shafto, 2016). Specifically, it has been argued that observed segmental changes in IDS may be unintended consequences of other properties of IDS, such as its speaking rate and prosodic characteristics (e.g. Benders, 2013; McMurray, Kovack-Lesh, Goodwin, & McEchron, 2013). Despite these advancements in our understanding of the characteristics of Clear Speech and IDS and the intelligibility gains associated with Clear Speech, very little is known about how speaking style adaptations affect the time course of spoken word recognition in quiet and in noise. The main goal of this study was to test whether Clear Speech and IDS modulate the temporal dynamics of spoken word recognition, for the first time comparing both to the "baseline" of conversational speech in the same study. We predict that both speaking styles will be beneficial for the listeners, but given that adults are not the intended "target audience" for IDS, we may see a larger benefit for Clear Speech for this group of listeners. However, another possibility is that the specific affect and typical pitch characteristics of IDS make this speaking style even more beneficial for a listener, regardless of whether that listener is considered part of the target audience.

The second goal of this paper was to examine how semantic context interacts with speaking style modifications in spoken word processing. In contexts like noisy environments, where acoustic–phonetic cues are masked, listeners may rely on higher-level linguistic structural and contextual information (lexical, semantic, and syntactic) in order to recover from losses at the perceptual level (Kalikow, Stevens, & Elliott, 1977; Miller, Heise, & Lichten, 1951; Nittrouer & Boothroyd, 1990). For example, McCoy et al. (2005) showed that adults with poor hearing recalled significantly fewer of the non-final words in word lists without semantic-contextual cues compared to word lists where target words were predictable from the prior words. Previous work also showed that supportive discourse information improves the recognition of reduced target words in spontaneous conversational speech (Bouwer, Mitterer, & Huettig, 2013). In addition, there is evidence that listeners derive significant benefits from semantic context and speech clarity (Clear Speech), and that these two cues are mutually enhancing in their effects on speech recognition (Bradlow & Alexander, 2007; Smiljanic & Sladen, 2013): Clear Speech appears to be of greater benefit to listeners when semantic context is also available. Combined, these studies suggest that high-predictability contexts may increase the likelihood of successful identification of the target word by decreasing the number of potential word candidates and thus facilitating word recognition (Friederici, Steinhauer, & Frisch, 1999; Mattys, White, & Melhorn, 2005). Presence of contextual

cues could also reduce the perceptual burden on listeners' processing resources in noise and so aid recall and recognition (McCoy et al., 2005; Van Engen et al., 2012). Here, we extend this work by investigating the effect of sentential context in combination with the speaking style adaptations on the time course of spoken word recognition. If speaking style by itself provides benefits to the listener regardless of context, the benefits of listener-oriented speaking styles should be uniform across both high- and low-predictability contexts.

Finally, the third goal was to examine whether the acoustic–phonetic characteristics of IDS as well as of Clear Speech aid word recognition for young adult listeners in more challenging listening conditions, namely in noise. While Clear Speech improves word recognition in noise for both adults and children, it is not known whether the IDS acoustic–phonetic adjustments result in similar processing advantages for all listeners. Several studies have considered similarities between IDS and different types of listener-oriented speaking styles, such as Lombard speech, foreigner-directed speech, and read speech (Martin et al., 2014; Scarborough, Dmitrieva, Hall-Lew, Zhao, & Brenier, 2007; Tang, Xu Rattanasone, Yuen, & Demuth, 2017). The results revealed a number of similarities between these speaking styles as well as a number of differences in, for instance, vowel space, tone space, and pitch expansion.

Similar cross-style acoustic-articulatory enhancements could thus lead to the similar processing benefit in noise. In contrast, the studies emphasizing the positive affect of IDS as crucial for infant attention and learning, while phonetic cues are modified incidentally along with slower speaking rate and different prosodic structure would predict IDS not to be beneficial for a non-target audience (e.g. McMurray et al., 2013). The affect of IDS, which makes it attractive to young listeners, may be a strong signal to adult listeners that they are not the target audience for this speech, and this could make IDS less attractive or even annoying for adult listeners, and in this case the prediction would be that it is less beneficial for word recognition in noise.

In a series of three experiments, the current study examined whether listener-oriented speaking style modifications and contextual-semantic information facilitate spoken word processing in quiet and in noise for young adult listeners. In all experiments, baseline conversational sentences were compared with Clear Speech and IDS sentences, and within each speaking style sentences with low versus high semantic predictability were compared.

In Experiment 1a, we tested 'offline' word recognition in which listeners heard conversational, Clear Speech, and IDS sentences in high-predictability and low-predictability contexts mixed with noise and then provided their written responses. In Experiment 1b, listeners rated the pleasantness of the sentences across speaking styles and semantic contexts. This was done in order to assess whether prosody and affect features of IDS, intended for children, are perceived as less attractive or less pleasant to adult listeners; if this was in fact the case, these features could interfere with the perceptual benefit of IDS in word recognition.

In Experiments 2 and 3, we tested 'online' spoken word recognition in a 'looking-while-listening' task, in which listeners' eye gaze is tracked continuously as they listen to speech and see pictures of objects on a screen (e.g. Cooper, 1974; McMurray, Clayards, Tanenhaus, & Aslin, 2008; Tanenhaus, Magnuson, Dahan, & Chambers, 2000). Crucially, eye movements are closely related to the speech input and can reveal, over time, which lexical candidate listeners believe the input supports. This paradigm relies on the tendency of people to fixate a named image, even when not explicitly instructed to do so (Fernald, Pinto, Swingley, Weinbergy, & McRoberts, 1998). In this way, subtle differences in the speed of word recognition as a result of available contextual and acoustic–phonetic cues can be compared. This will allow us to more closely investigate the time course of the facilitatory effects of semantic context and speaking style on spoken word recognition in quiet (Experiment 2) and in noise (Experiment 3). The examination of the locus of the speaking style and context processing benefit will provide a deeper understanding of the mechanisms underlying spoken word recognition in challenging listening conditions.

2. Experiments 1a and 1b – Intelligibility and pleasantness

2.1. Method

2.1.1. Participants

Two groups of adult listeners were recruited from the University of Texas at Austin Linguistics Department subject pool. Eighteen listeners (9 female, age range 19–24 years) participated in an intelligibility-in-noise task. Eighteen listeners (13 female, age range 18–32 years) participated in a pleasantness rating task. All participants were native, monolingual speakers of American English. All were undergraduate students at the University of Texas at Austin and received class credit for their participation. All passed a pure-tone hearing screening administered bilaterally at 25 dB hearing level at 500, 1000, 2000, and 4000 Hz.

2.1.2. Stimuli

Seventy-two sentences were recorded by a 27-year-old female native speaker of American English. The sentences were simple and short as they were developed specifically for testing children's ability to use contextual-semantic cues in speech recognition in noise (Fallon, Trehub, & Schneider, 2002). In half (36) of the sentences the final word occurred in a High Predictability semantic context (e.g. "Mice like to eat *cheese"*), and in the other half the sentence-final word occurred in a Low Predictability semantic context (e.g. "He looked at the *cheese"*). In both sentence types, the final, monosyllabic word served as a target word to score participants' accuracy of word recognition.

The recordings were made using a Shure SM10A head-mounted microphone and a MOTU UltraLite-MK3 Hybrid recorder. The recordings were made in a sound-attenuated booth over the course of two sessions. During the first session, the talker read sentences that were displayed on PowerPoint slides one at a time on a computer screen. First, all sentences were recorded in a Conversational speaking style. For Conversational sentences, the talker was instructed to read in a casual manner as if she was talking to someone familiar with her voice and speech patterns. Next, all sentences were

recorded in Clear Speech. For the Clear Speech sentences, the talker was asked to read as if she was talking to a listener with hearing loss or to a non-native speaker, since previous studies have shown that listener-oriented Clear Speech can be elicited successfully in this way in a laboratory setting (Harnsberger, Wright, & Pisoni, 2008; Smiljanić & Bradlow, 2009). In the second recording session, the talker was instructed to read the same sentences, but this time as if she was talking to an infant. The PowerPoint slides with the target sentences now also included stock photographs of infants. The total set included 216 recorded sentences (36 High Predictability + 36 Low Predictability in each of three speaking styles: Conversational, Clear Speech, and IDS). The recorded sentences were segmented into individual files and equalized for RMS amplitude.

a. Intelligibility-in-noise. Each file was digitally mixed with speech-shaped noise (SSN) at a signal-to-noise ratio (SNR) of −5 dB sound pressure level using Praat (Boersma & Weenink, 2012), following previous studies by e.g. Smiljanić and Bradlow (2005, 2008), Gilbert et al. (2014) and Mattys et al. (2009). This noise level has been classified as "moderate" in those previous studies, and was chosen to ensure that listeners would not perform at ceiling in an 'easy' listening condition or at the floor with a more difficult SNR. This was further verified by looking at the overall performance of our first subjects, who were indeed performing in the expected reported range for this noise level of 45–46% average intelligibility in the baseline Conversational Speech in quiet. Each stimulus file consisted of a 400 ms silent lead, followed by 500 ms of noise before the onset of the target sentence, and ended with 500 ms of noise after the offset of the target sentence. Each listener heard 36 unique sentences counterbalanced for speaking style and semantic predictability: 18 High Predictability and 18 Low Predictability sentences, six sentences in each of six blocks, and for each semantic predictability six sentences were presented in Conversational, six in Clear Speech, and six in IDS. Each listener heard stimuli in one of six presentation orders. Each block included six sentences from a single semantic context and speaking style (e.g. sentences 1–6 were all High Predictability and all produced in Conversational). Sentence order within each block was fixed. The order of the blocks was pseudo-randomized across conditions to minimize order effects. Participants never heard the same sentence twice. Each sentence was heard in all speaking styles and contexts across participants. The blocks may have made the task slightly "easier" for the listener, but since style and context changed every three sentences, whatever expectations were formed changed quickly and would impact all conditions equally.[1]

b. Pleasantness ratings. A subset of sentences from each style and semantic context was used to gauge whether young adult listeners found different speaking styles more or less pleasant. Participants heard a total of 60 sentences presented in quiet: 10 High Predictability and 10 Low Predictability sentences were presented in each of the three speaking styles (Conversational, Clear Speech, and IDS), in blocks of 3 sentences with the same style/predictability, to match the setup of Experiment 2 and 3 as explained below.

2.1.3. Acoustic analyses of stimuli

A series of acoustic analyses were performed to confirm that the two listener-oriented speaking style adaptations (Clear Speech and IDS) differed in acoustic-articulatory characteristics from the Conversational stimuli and from each other. In order to perform the analyses, sentences were manually annotated using Praat textgrids (Boersma & Weenink, 2012). Praat scripts were then run in order to obtain acoustic values automatically from the annotated sentences. The specific acoustic–phonetic features were: speaking rate (syllables per second), energy in 1–3 kHz range (dB), and F0 range and mean (Hz). We focused on these features, as they are typically found in conversational-to-clear and adult-to-child-oriented modifications (see e.g. Cristia, 2013; Smiljanić & Bradlow, 2009).

Speaking rate was calculated as the number of syllables produced per second after the pauses were excluded. A pause was defined as a period of silence of at least 100 ms in duration, excluding silent periods before word-initial stop consonants where it would be impossible to determine the end of a pause and the beginning of the stop closure (see Smiljanić & Bradlow, 2005). Energy in the 1–3 kHz range was measured by averaging the long-term average spectrum energy between 1 and 3 kHz across each sentence. Pitch was measured as F0 mean and range (the difference between the highest and lowest F0 points in the sentence). Table 1 below summarizes the acoustic characteristics of the stimuli.

The measurements illustrate that the three speaking styles differed from each other as intended; Conversational sentences (target word durations and speaking rate) were faster than Clear Speech and IDS (Clear Speech and IDS being similar); IDS showed higher energy levels than Clear Speech (Clear Speech and Conversational being similar); Conversational sentences showed the least pitch variability, and IDS the most (i.e. for F0 range and mean, Conversational < Clear < IDS).

2.1.4. Procedure

a. Intelligibility-in-noise. Participants sat at a computer monitor in a sound-attenuated booth in the UT SoundLab at the University of Texas at Austin. The stimuli were presented over Sennheiser HD570 or Sony MDR-CD780 headphones at a comfortable listening level using EPrime (Schneider, Eschman, & Zuccolotto, 2002). Listeners were instructed to type out as much as they could of the sentence they had just heard. After each trial, the participant pressed a button on the keyboard to move onto the next trial. Each trial was presented only once, but participants could take as much time as they wished to write down the sentences. In order to familiarize listeners with the task, they heard two practice sentences produced by a different talker and not used in the test. Practice sentences were mixed with SSN at +2 dB SNR.

b. Pleasantness ratings. All participants were tested in the same set-up in the same sound attenuated booth as in the intelligibility-in-noise task. A computerized visual-analog scale

[1] Presenting the sentences in blocks of three made it possible to use this same design in a follow up study with 3- and 4-year-olds; in addition to make the task slightly less "hard" by not switching speaking style every sentence (or every other sentence), we wanted to maximize the chances of getting a data point for each participant, for each sentence type with young children who in this type of task typically have less focused looking behavior, and a typically higher number of lost trials per participant.

Table 1
Acoustic characteristics of the auditory stimuli recorded in Conversational Speech (Conv), Clear Speech (CS), and Infant-Directed Speech (IDS), in High Predictability (HP) versus Low Predictability (LP) sentences (Average target word duration in ms. with standard deviations; speaking rate in syllables per second; Root Mean Square for slopes; mean energy in the 1–3 kHz rang; F0 range and mean).

	Target word duration, ms. (SD)	Speaking rate (syllables/sec)	RMS for slope	Mean energy, dB 1–3 kHz (SD)	F0 range, mean (min–max)
Conv-LP	430 (0.08)	5.99	9.70	19.54 (3.6)	116.3, 164.9 (125–241)
Conv-HP	410 (0.10)	5.58	9.34	21.39 (3.5)	104.6, 161.8 (122–227)
CS-LP	570 (0.08)	3.30	9.05	20.16 (3.1)	178.6, 165.9 (128–306)
CS-HP	610 (0.09)	2.75	9.55	21.20 (3.0)	162.2, 164.1 (108–270)
IDS-LP	670 (0.10)	3.32	10.29	26.1 (2.3)	253.1, 225.4 (135–388)
IDS-HP	740 (0.10)	2.67	10.11	27.8 (2.3)	246, 223.5 (123–370)

(VAS) was presented with EPrime, and listeners judged the pleasantness of each sentence by clicking anywhere on a horizontal line presented in the middle of the screen. The line was labeled "most pleasant" on the right endpoint (higher scores) and "least pleasant" on the left endpoint (lower scores). The line was not labeled at any other intermediate point and no line divisions were visible. Each listener rated all 60 sentences. The order of sentences was randomized for each listener, and listeners heard each sentence only once. Three practice sentences were included at the beginning to familiarize listeners with the task. For each sentence, the click location in pixels was logged.

2.2. Results and discussion

2.2.1. Experiment 1a – Intelligibility-in-noise

We adopted a strict scoring criterion. A keyword was counted as correct only if all morphemes of the target word were present and transcribed correctly, e.g. if the target word was "keeping," "keep, keeps, or kept" were scored as incorrect. Obvious alternate homophone spellings were counted as correct (e.g. *their* for *there*). Since the target words were very frequent short words, orthographic mistakes were virtually nonexistent.

All statistical analyses reported in this paper were conducted using R (version 3.3.0; R Core Team, 2016) and RStudio (1.0.136; RStudio Team, 2016), and the packages *lme4* (Bates, Maechler, Bolker, & Walker, 2015) and *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2016). A mixed-effects logistic regression model (Baayen, Davidson, & Bates, 2008) was fit to the data from the intelligibility-in-noise task to model the accuracy of listeners' transcription of the final target word of each sentence (1 = correct, 0 = incorrect). Speaking style (Conversational, Clear Speech, IDS) and sentence context (High Predictability, Low Predictability) were entered as fixed effects, and participant and target word were included as random intercepts. Results are illustrated in Fig. 1.

Increasingly complex nested models were compared via ANOVAs to determine the simplest model that best fit the data. The best-fitting model revealed significant main effects of speaking style and sentence context (both ps < 0.001), and a significant interaction between the two ($p < 0.001$). Pairwise comparisons were made by releveling the model with the six different possible conditions as the intercept. (Note that with the current sample size, the models did not converge with the addition of random slopes.) There was a greater likelihood of accurate transcription for IDS than for Clear Speech or Conversational (IDS vs. CS: $\beta = -0.98$, $z(640) = -1.946$, $p = 0.05$; IDS vs. Conv: $\beta = -4.56$, $z(640) = -8.632$, $p < 0.001$), for Clear Speech than for Conversational ($\beta = -3.58$, $z(640) = -8.113$, $p < 0.001$), and for High Predictability than for Low Predictability sentences ($\beta = -2.01$, $z(640) = -4.950$, $p < 0.001$).[2] For the interaction of these factors, there was no difference in likelihood of accurate transcription between High Predictability and Low Predictability contexts for Conversational ($p > 0.05$), but listeners benefited from High Predictability context over Low Predictability when more exaggerated acoustic cues were present, i.e., in the two listener-oriented speaking styles (Clear Speech: $\beta = -2.01$, $z(640) = -4.950$, $p < 0.001$; IDS: $\beta = -2.31$, $z(640) = -4.752$, $p < 0.001$). See Table 2 for model summary. Results showed that listeners used speaking style modifications and sentence context in combination to enhance their word recognition in noise. Interestingly, there was a marginally significant effect indicating that young adult listeners found IDS acoustic–phonetic adjustments to be even more beneficial for word recognition in noise than Clear Speech.

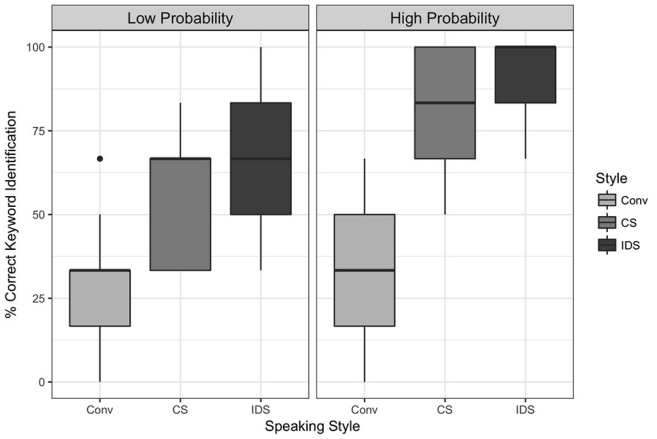


Fig. 1. Percentage correct keyword identification, with whiskers from 1.5 times the interquartile range below and above the first and third quartiles, respectively, for Conversational Speech (Conv), Clear Speech (CS), and Infant-Directed Speech (IDS). Dots represent outliers. Data for the Low Predictability Context sentences are in the left panel and data for the High Predictability Context sentences are in the right panel.

2.2.2. Experiment 1b – Pleasantness ratings

A mixed-effects linear regression model was fit to the pleasantness ratings to model listeners' perceived pleasantness of

[2] Due to the structure of the regression models uses here, the estimates reported here for the speaking style differences are the estimates for the high probability sentences, and the estimates reported for the probability contexts are for the CS sentences. The direction and magnitude of effects are representative of the combined class (e.g., representative of both HP and LP, even though the in-text estimate is for HP sentences).

Table 2
Experiment 1a: Summary of model fitting transcription accuracy. Intercept represents log odds of accurately transcribing target words in high predictability sentences in Clear Speech (CS).

	Estimate	Std. Error	*z* value	*p* value
Intercept (CS, HP)	2.2965	0.4470	5.137	<0.001
LP	−2.0103	0.4061	−4.950	<0.001
Conv	−3.5756	0.4407	−8.113	<0.001
IDS	0.9799	0.5036	1.946	0.05167
LP*Conv	1.9790	0.5404	3.662	<0.001
LP*IDS	−0.2954	0.6081	−0.486	0.62713

Random effects:

	Variance	Std. Deviation
Word	2.8192	1.6790
Participant	0.2239	0.4731

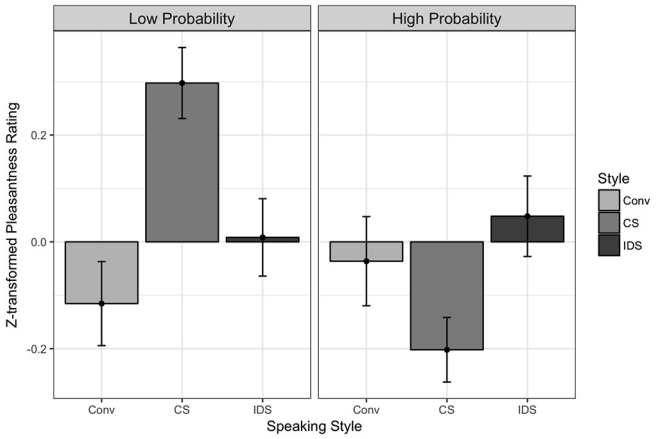


Fig. 2. Average z-transformed pleasantness ratings, with standard errors, for Conversational Speech (Conv), Clear Speech (CS), and Infant-Directed Speech (IDS). Dots represent outliers. Data from Low Predictability (LP) context sentences are in the left panel and data from High Predictability (HP) context sentences are in the right panel.

the different speaking styles and sentence types. Pleasantness ratings were converted to z-scores to account for differences in how raters use the available continuum, so a score of 0 indicates speech rated as neither particularly pleasant nor as particularly unpleasant, positive scores indicate more pleasant speech, and negative scores indicate less pleasant speech. Speaking style (Conversational, Clear Speech, IDS) and sentence context (High Predictability, Low Predictability) were tested as fixed effects, and sentence was included as a random intercept. The results are illustrated in Fig. 2.

As in Experiment 1a, models were tested against each other using ANOVA, and the referent (intercept) of the best fitting model was releveled for pairwise comparisons. A model with an interaction between speaking style and sentence context and with sentence as a random intercept provided a significantly improved fit to the data compared to a model without the interaction term ($\chi^2 = 18.384$, df = 2, $p < 0.001$); see Table 3 for the model summary. IDS and Conversational were not rated as significantly different from 0 (neutral) in either Low Predictability or High Predictability semantic contexts (all p ns > 0.05). In contrast, both Low Predictability and High Predictability sentences in Clear Speech were significantly different from 0. In Low Predictability contexts, Clear Speech was rated as significantly more pleasant ($\beta = 0.30$, $t(1074) = 4.062$, $p < 0.001$) while in High Predictability contexts, Clear Speech was rated as significantly less pleasant ($\beta = -0.20$, $t(1074) = -2.759$, $p < 0.01$).

The apparent neutral pleasantness ratings of IDS to adult listeners may be due in part to large differences across how listeners rated the pleasantness of IDS. By examining the distribution of raw scores for each sentence type, we see that most listeners rated Conversational and Clear Speech near the center of the scale, near neutral pleasantness (see Appendix A). In contrast, the ratings for IDS sentences were more distributed across the scale. This difference in the distribution of ratings across sentence types can be captured in the kurtosis of each distribution, which measures how much of the data is found in the tails of a distribution. Distributions with negative excess kurtosis values indicate more data in the tails than are found in the normal distribution, and distributions with negative values have as much data in the tails as the peak of the distribution (DeCarlo, 1997). While the kurtosis values for all six

Table 3
Experiment 1b: Summary of model fitting z-transformed pleasantness ratings. Intercept represents z-transformed rating for high predictability sentences in Clear Speech (CS).

	Estimate	Std. Error	*t* value	*p* value
Intercept (CS, HP)	−0.20196	0.07319	−2.759	<0.01
LP	0.49927	0.10351	4.823	<0.001
Conv	0.16579	0.10351	1.602	0.1095
IDS	0.24991	0.10351	2.414	<0.05
LP*Conv	−0.57863	0.14639	−3.953	<0.001
LP*IDS	−0.53884	0.14639	−3.681	<0.001

Random effects:

	Variance	Std. Deviation
Sentence	1.3×10^{-14}	0.0000001
Residual	0.9643	0.9820

sentence types are negative (the tails have more data than would occur in a normal distribution), the IDS kurtosis scores are much more negative than the scores for the other styles (see Table 4). These more widely distributed ratings show that young adult listeners were less consistent in their ratings of IDS sentences, regardless of semantic context, than for the other adult-oriented speaking styles.

Combined results of Experiments 1a and 1b have established that acoustic–phonetic enhancements of both Clear Speech and IDS along with semantic enhancements contributed significantly to the improved word recognition in noise for young adults and that this benefit did not arise from the listeners' preference of Clear Speech over IDS, as seen in the pleasantness ratings. Furthermore, the direct comparison of the two listener-oriented speaking styles revealed a larger intelligibility benefit of IDS compared to Clear Speech, suggesting that the acoustic-articulatory modifications aimed at two different listener groups led to different intelligibility gain for the listeners. Next, we examine the time course of word recognition for the three speaking styles and two semantic contexts using an online word-recognition task (e.g. Fernald et al., 1998; Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987).

Table 4
Kurtosis of the distribution of responses to each sentence type.

	Kurtosis
Conv, HP	−0.52
Conv, LP	−0.41
CS, HP	−0.43
CS, LP	−0.44
IDS, HP	−1.13
IDS, LP	−1.18

3. Experiment 2 – Online word recognition in quiet

3.1. Method

3.1.1. Participants

Thirty-six adult native, monolingual speakers of American English (22 female, age range 19–25 years) participated in Experiment 2. All participants were undergraduate students at the University of Texas at Austin and received class credit for their participation. They were different individuals than in Experiment 1. All passed a pure-tone hearing screening, administered bilaterally at 25 dB HL at 500, 1000, 2000, and 4000 Hz.

3.1.2. Stimuli

The *auditory stimuli* consisted of a subset of 36 sentences from Experiment 1, including the same 18 unique sentence-final words in both High Predictability and Low Predictability sentences. Sentences were selected for having sentence-final target words that were easily depictable objects (i.e., targets such as *ball* were selected rather than targets such as *snow*). All selected sentences had target words that were monosyllabic; most had a CVC structure (e.g. *bed*, $n = 13$), four targets had a word-final consonant cluster (*corn, fork, pants, horse*), one had a CV structure (*bee*), and one had a CCV structure (*tree*). These design restrictions were implemented to ensure that this paradigm can be used with young children (Van der Feest, Blanco, & Smiljanic, 2016).

The *visual stimuli* consisted of photographs of the target objects, which were edited to be similar in size and brightness. The objects measured about 6–7 inches on the screen, and were presented on a 27-inch screen. Target object pictures appeared in pairs, side-by-side on the screen, and separated by about 8–9 inches. As much as possible, objects were paired based on basic semantic properties (e.g. animate objects were paired with other animate objects, a bus was paired with a car).

Each participant was presented with 18 unique sentences (with 18 different target words) evenly divided between High Predictability and Low Predictability sentences. For each semantic context (Low Predictability and High Predictability), three sentences each were presented in Conversational, Clear Speech, and IDS styles. Sentences were always presented in blocks of three with the same semantic context and speaking style combination (e.g. High Predictability Conversational) within each block. The subset of sentences was counterbalanced across subjects for speaking style and semantic predictability, resulting in six different test orders (see Appendix C). Each sentence was heard in all speaking styles and contexts across the different orders. The order of the semantic context and speaking style combination blocks was counterbalanced across the six conditions. Participants were presented with each sentence twice over the course of the experiment, for a total of 36 trials per participant. The order of the semantic context and speaking style blocks was the same in the first and the second half of each condition, but the order of the sentences within each block was different in the first versus the second half of the test.

3.1.3. Procedure

All participants were tested in a sound-proof booth in the Perception, Production and Processing lab at the University of Texas at Austin Speech and Hearing Center. Participants sat in front of a 27-inch iMac monitor. The pre-recorded audio was played over external speakers, and a remote-controlled video camera (Sony EVI-D100) located directly below the screen recorded the participants' faces. Participants were instructed to just watch the video, and the experimenter explicitly stated that there were no hidden tasks.

Each trial consisted of a 2-s lead where the pictures were presented on the screen in silence, followed by the target sentence, and ending again with the pictures presented in silence so that each trial was exactly 5.5 s in duration. The side on which the labeled target object appeared was counterbalanced across the different test orders, such that each target object appeared in on the left in three orders and on the right in three orders. The same two objects always appeared together in all conditions, alternating which object was the named target on a particular trial within each order. After each trial, a white flashing star on a black background appeared, to direct participants' eye gaze towards the center of the screen between trials. After every nine trials, short animations were played (of a duck bouncing, a fish swimming, or a bird flying across the screen). The duck, fish, and bird were not named (cf. Van der Feest & Johnson, 2016).

3.1.4. Coding and analyses

Participants' eye movements were coded off-line by trained coders who analyzed the silenced video using the SuperCoder program (Hollich, 2005). Eye movements were coded for each 33.3 ms frame (30 frames/second): the coders indicated whether the participant was looking at the left picture, at the right picture, or was shifting between pictures or looking away from the screen. The beginning and end of each trial was clearly indicated on the video by a change in background light (between the white background behind the target objects and the dark background behind the blinking star and the short animations). The coder was blind to target side and test order. Coder reliability was determined by comparing the decisions of two different coders for twenty percent of the total dataset. The mean agreement between coders was 97%.

Following previous studies (e.g. Fernald et al., 1998; Swingley & Aslin, 2000), we assessed fixations to the target object (versus the distractor object) for each 33.3 ms frame of each trial. We first assessed target word recognition and checked for inherent preferences of target objects by calculating the proportion of the total looking time to the screen on which participants fixated the target object, for the first second of each trial as well as for the first second starting at target word onset. Next, we analyzed fixations to the target object

over time during the one-second window of analysis starting at target word onset. While planning an eye-movement in response to an auditory stimulus is estimated to take perhaps as little as 100 ms for adults (Altmann, 2011) and up to 365 ms in young children (e.g. Fernald et al., 1998; Johnson, McQueen, & Huettig, 2011; Swingley, 2009), here we measure fixations from target word onset because in the High Predictability sentences, participants arguably already had enough information before the target word onset to predict the sentence-final target word. Thirteen trials (about 2% of all trials across all subjects) were excluded because the participants were briefly distracted and did not look at the screen (e.g. because of sneezing, briefly changing their exact position in their chair, or prolonged blinks).

3.2. Results and discussion

Proportions of target object fixations are illustrated for each of the different trial types before and after target word onset in Fig. 3.

Two-tailed *t*-tests were conducted to compare looks to chance or 50%; looks to the target 'before' (during the first second of each trial) were at chance level for all six sentence types, illustrating that no inherent biases or picture preferences were found (Conversational–Low Predictability $t(35) = 0.5$, p = n.s.; Conversational-High Predictability $t(35) = -0.4$, p = n.s.; Clear-Low Predictability $t(35) = 0.2$, p = n.s.; Clear-High Predictability $t(35) = 0.2$, p = n.s.; IDS-Low Predictability $t(35) = 1.7$, p = n.s.; IDS-High Predictability $t(35) = 1.1$, p = n.s.). Looks to the target 'after' (during a one-second window beginning at target word onset) were all significantly different from 0, indicating the target words in silence were always recognized (as expected). (Conversational-Low Predictability $t(35) = 11.3$, $p < 0.0001$; Conversational–High Predictability $t(35) = 13.3$, $p < 0.0001$; Clear-Low Predictability $t(35) = 8.9$, $p < 0.0001$; Clear-High Predictability $t(35) = 11.3$, $p < 0.0001$; IDS-Low Predictability $t(35) = 5.9$, $p < 0.0001$; IDS-High Predictability $t(35) = 10.5$, $p < 0.0001$).

Next, we conducted two-tailed *t*-tests comparing looks to the target picture to chance at target word onset, i.e. the first frame visible in Fig. 4, to make a first assessment of the effect of semantic context and speaking style. We find that at target word onset, the proportion of looks to the target is already significantly above chance for all sentence types with High Predictability semantic context (Conversational $t(35) = 4.2$, $p < 0.0001$; Clear $t(35) = 10.2$, $p < 0.0001$; IDS $t(35) = 7.9$, $p < 0.0001$). This indicates that the semantic context in itself is already helpful for target word recognition, regardless of speaking style. For the sentences with Low Predictability semantic contexts, overall looks to the target at target word onset were not significantly different from chance for Clear sentences ($t(35) = 0.6$, p = n.s.) nor for IDS sentences ($t(35) = -0.6$, p = n.s.), and marginally significant for Conversational sentences ($t(35) = 2.0$, $p = 0.05$).

To further asses the looks to the target object over time and compare participants' looking behavior on the six different sentence types, a mixed-effects linear regression model was fit to the proportion of looks to the target object at each 33.33 ms time frame, during the first one-second window after target word onset. Speaking style (Conversational, Clear Speech, IDS), semantic context (High Predictability, Low Predictability), and time (frame) were tested as fixed effects, and participant was included as a random intercept. The results are illustrated in Fig. 4.

As above, models were tested against each other using ANOVA, and the referent (intercept) of the best fitting model was releveled for pairwise comparisons. The best-fitting regression model (summarized in Table 5) included a three-way interaction among speaking style, semantic content, and time. The proportion of fixations to the target were significantly higher for High Predictability sentences than for Low Predictability sentences in all three speaking styles (Clear: $\beta = -0.03$, $t(6649) = -18.879$, $p < 0.001$; Conversational: $\beta = -0.07$, $t(6649) = -4.826$, $p < 0.001$; IDS: $\beta = -0.29$, $t(6649) = -18.809$, $p < 0.001$)). Among the High Predictability sentences, the overall proportions of fixations to the target was greatest in Clear Speech, less in IDS (Clear vs. IDS: $\beta = -0.06$, $t(6649) = -3.605$, $p < 0.001$), and least in Conversational (Clear vs. Conversational: $\beta = -0.13$, $t(6649) = -8.520$, $p < 0.001$; IDS vs. Conversational: $\beta = -0.08$,

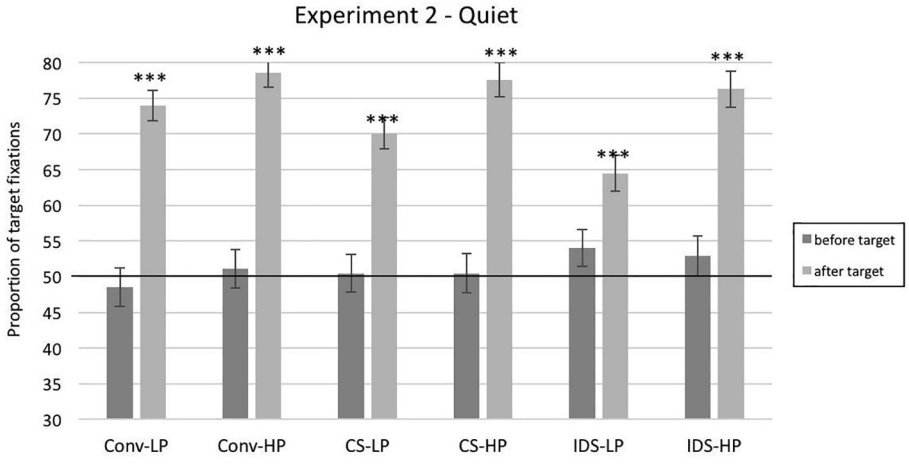


Fig. 3. Proportions of target fixations in quiet, as a function of the total looking time during a one second window of analysis, with standard errors. Bars illustrate target fixations during the first second of each trial (before target, dark grey) versus the first second of each trial starting at target word onset (after target, light grey). The line at 50% represents chance. Three stars indicate significant differences from chance at the $p < 0.001$ level. Data is broken down by speaking style (Conversational Speech (Conv), Clear Speech (CS), Infant-Directed Speech (IDS)) and by semantic context (High Predictability (HP), Low Predictability (LP)).

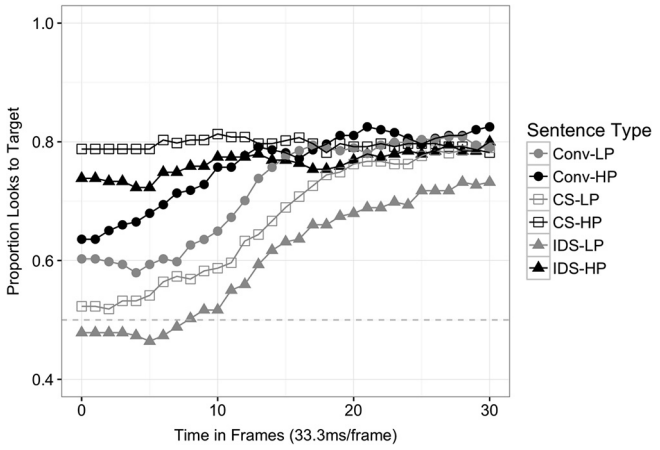


Fig. 4. Participants' target fixations, in quiet. Lines illustrate the proportion of looks to the target picture in a one-second window after target word onset, by showing the looks to the target picture calculated as the proportion of looks to the target compared to the total looks to the target and distractor picture for each 33.33 ms time frame. The dotted line at 0.5 represents chance. Data is broken down by speaking style (Conversational Speech (Conv), Clear Speech (CS), Infant-Directed Speech (IDS)) and by semantic context (High Predictability (HP), Low Predictability (LP)).

Table 5
Experiment 2: summary of model fitting proportion of fixations to target. Intercept represents proportion of fixations to target for High Predictability sentences in Clear Speech.

	Estimate	Std. Error	*t* value	*p* value
Intercept (CS, HP)	0.79360	0.02228	35.625	<0.001
LP	−0.29250	0.01549	−18.879	<0.001
Conv	−0.13200	0.01549	−8.52	<0.001
IDS	−0.05586	0.01549	−3.605	<0.001
Time (Frame)	−0.00012	0.00063	−0.189	0.85
LP*Conv	0.21780	0.02191	9.937	<0.001
LP*IDS	0.00108	0.02191	0.049	0.96
LP*Time	0.01121	0.00089	12.637	<0.001
Conv*Time	0.00681	0.00089	7.679	<0.001
IDS*Time	0.00202	0.00089	2.276	<0.05
LP*Conv*Time	−0.00909	0.00126	−7.242	<0.001
LP*IDS*Time	−0.00270	0.00126	−2.154	<0.05

Random effects:

	Variance	Std. Deviation
Participant	0.01354	0.1164
Residual	0.03514	0.1875

$t(6649) = -4.915$, $p < 0.001$). For the Low Predictability sentences, listeners fixated the target most in Conversational, less so in Clear Speech (Conversational vs. Clear Speech: $\beta = -0.09$, $t(6649) = -5.534$, $p < 0.001$), and least in IDS (Conversational vs. IDS: $\beta = -0.14$, $t(6649) = -9.069$, $p < 0.001$; Clear Speech vs. IDS: $\beta = -0.05$, $t(6649) = -3.535$, $p < 0.001$). The effect of the time variable indicated significant non-zero slopes for all conditions except for High Predictability Clear Speech (Clear Speech, High Predictability: $p > 0.05$; Conversational, Low Predictability: $\beta = 0.01$, $t(6649) = 17.683$, $p < 0.001$; IDS, High Predictability: $\beta = 0.002$, $t(6649) = 3.031$, $p < 0.01$; IDS, Low Predictability: $\beta = 0.01$, $t(6649) = 16.594$, $p < 0.001$; Conversational, High Predictability: $\beta = 0.007$, $t(6649) = 10.670$, $p < 0.001$; Conversational, Low Predictability: $\beta = 0.009$, $t(6649) = 14.057$, $p < 0.001$). Listeners were at ceiling for High Predictability Clear Speech sentences and did not change over time, but looks to target increased significantly for all other sentences in the first second after target word onset. Over the entire second, fixations to target increased more quickly (had steeper slopes) for the Low Predictability sentences than for the High Predictability sentences (Clear Speech: $\beta = -0.01$, $t(6649) = -12.637$, $p < 0.001$; IDS: $\beta = -0.009$, $t(6649) = 9.590$, $p < 0.001$; Conversational: $\beta = -0.002$, $t(6649) = -2.395$, $p < 0.05$). For the three styles of High Predictability sentences, fixations to target increased more rapidly for IDS than for Clear Speech ($\beta = -0.002$, $t(6649) = -2.276$, $p < 0.05$) and even more so for Conversational than for IDS ($\beta = -0.005$, $t(6649) = -5.402$, $p < 0.001$). Among the Low Predictability sentences, fixations increased faster for Clear Speech than for Conversational ($\beta = -0.002$, $t(6649) = -2.564$, $p < 0.05$), and the rate of change for IDS did not differ from the other styles ($ps > 0.05$).

A visual inspection of the proportions of looks to the target in the six conditions (Fig. 4) suggests that at least some conditions may have had non-linear trends of proportions of fixations to the target over time, which would not be captured with the linear time predictor in the model reported above. Therefore, changes in looks over time were investigated using multiple regression lines to fit the data in each condition instead of a single line per condition. First, we analyzed the data using a breakpoint linear regression analysis in order to test whether each condition could be better described statistically with multiple regression lines, one on either side of a "breakpoint." If there is such a breakpoint, a segment on each side of it could be described by a distinct equation and changes in slope and intercept between the segments could be compared. For each of the six conditions, each frame was tested as a possible breakpoint in a breakpoint linear regression or standard linear regression model. Models for only two conditions were improved by the inclusion of a breakpoint: Low Predictability sentences produced in Conversational (breakpoint at frame 3) and Low Predictability sentences produced in IDS (breakpoint at frame 4).

Since only two of the six conditions were better explained with a breakpoint regression, and since the breakpoints occurred in neighboring frames, an alternative approach was taken to compare differences in slopes and intercepts among the conditions at different time points. "Breakpoints" were uniformly assigned across the conditions by dividing the analysis window into three bins: two 10-frame bins before frame 20 (666 ms, per the breakpoint analysis) and one bin of the 11 frames after frame 20 (the remainder of the analysis window).

For each bin, a mixed-effects linear regression model was fit to the data with style (Conversational, Clear Speech, IDS), context (High Predictability, Low Predictability), and time (frame) tested as fixed effects and subject as a random intercept. Details of the results for each bin are included in Appendix B. In *Bin 1* (0–333 ms after target word onset), there was a significant interaction of speaking style and semantic context and a main effect of time, but no interactions between time and the other factors. That is, immediately after target word onset, looks to the target changed significantly over time, but the change in proportion of looks did not vary across the six conditions. In *Bin 2* (333–666 ms after target word onset), there was a three-way interaction between style, context, and time. The proportion of looks to the target varied significantly across the six conditions, and the increase in looks over time also varied across the conditions, with faster increases in

looks to target in the three Low Predictability conditions than in the three High Predictability conditions. In *Bin 3* (666–1000 ms after target word onset), there was an interaction between style and context, but no main effect of time and no interactions with time. By Bin 3, some differences between conditions remained: looks to target in Low Predictability IDS remained significantly lower than in High Predictability IDS and compared to the other Low Predictability conditions, and there were fewer looks to target for Low Predictability Clear Speech than for Low Predictability Conversational. However, by Bin 3, there were no further changes in proportion of fixations over time for any condition.

The results of Experiment 2 provide a more nuanced understanding of the intelligibility benefit reported in Experiment 1a where 'offline' word recognition scores showed the overall benefit of combined speaking style and context enhancements. Detailed analyses of the temporal dynamics of spoken word recognition revealed differences already at the onset of the target word. These differences were due to speech clarity and semantic information of the sentence portion preceding the target word. Listeners were faster to fixate the target picture for the High Predictability Clear Speech and High Predictability IDS sentences. They were also faster for Clear Speech sentences than IDS sentences. The beneficial effect of context and speaking style continued to affect rate of fixations as listeners heard more of the target word itself. The bin analyses showed that early on (Bin 1) listeners' looks to the target picture increased uniformly across all conditions, nearly reaching ceiling for the High Predictability conditions. The changes in the looks to the target for the sentences in two predictability contexts were distinct in Bin 2, 333 ms to 666 ms after target word onset. Here, listeners hearing Low Predictability sentences continued to make steep, rapid gains in the proportion of looks to the target. By Bin 3, participants had reached their final maximum looks to the target and there were no further changes over time. Overall, listeners performed at ceiling in this online word-recognition task in which they heard all sentences in quiet. Note that the fixation rates never reach 100%, even though this task in quiet is rather easy, likely due to the fact that the listeners were never instructed specifically to look at the target picture. Next, we examine whether the speaking style and context benefit for the time course of word recognition is maintained under adverse listening conditions, i.e., in noise.

4. Experiment 3 – Word recognition in noise

4.1. Method

4.1.1. Participants

Eighteen additional adult native monolingual speakers of American English (11 female, age range 20–24 years) participated in Experiment 3. All participants were undergraduate students at the University of Texas at Austin and received class credit for their participation. Two additional participants were tested but not included in the analyses because of equipment failure ($n = 1$) and extreme distractedness leading to the subject not looking at the screen and missing data on more than half of the trials ($n = 1$). All passed a pure-tone hearing screening, administered bilaterally at 25 dB HL at 500, 1000, 2000, and 4000 Hz.

4.1.2. Stimuli, apparatus, and procedure

The 36 sentences used in Experiment 2 were digitally mixed with speech-shaped noise (SSN) at an SNR of −5 dB SPL, the same noise and SNR as used in Experiment 1a. Each auditory stimulus in Experiment 3 consisted of a 1.5 second silent lead, followed by 500 ms of noise and the target sentence in noise, and ended with silence, so that all trials were (as in Experiment 2) exactly 5.5 s in duration. Visual stimuli, apparatus, and procedure were identical to Experiment 2.

4.2. Results and discussion

Proportions of target object fixations are illustrated for each of the different trial types before and after target word onset in Fig. 5.

Two-tailed *t*-tests were conducted to compare looks to chance or 50%; as in Experiment 2, looks to the target 'before' (during the first second of each trial) were at chance level for all six sentence types, again illustrating no inherent biases or picture preferences. (Conversational–Low Predictability $t(17) = 0.7$, p = n.s.; Conversational-High Predictability $t(17) = 1.2$, p = n.s.; Clear-Low Predictability $t(17) = 0.4$, p = n.s.; Clear-High Predictability $t(17) = 0.7$, p = n.s.; IDS-Low Predictability $t(17) = 2.1$, p = n.s.; IDS-High Predictability $t(17) = 0.7$, p = n.s.). Looks to the target 'after' (during a one-second window beginning at target word onset) were significantly different from 0 for all conditions, except for the Conversational sentences with Low Predictability semantic context. (Conversational–Low Predictability $t(17) = 1.3$, p = n.s.; Conversational-High Predictability $t(17) = 3.5$, $p = 0.01$.; Clear-Low Predictability $t(17) = 2.5$, $p < 0.05$; Clear-High Predictability $t(17) = 4.1$, $p < 0.0001$; IDS-Low Predictability $t(17) = 3.8$, $p < 0.05$.; IDS-High Predictability $t(17) = 5.3$, $p < 0.001$). This indicates that unlike in silence (Experiment 2), target words in noise in Low Probability Conversational sentences were not significantly recognized during the first second after target word onset. This sentence type arguably represents the most difficult condition in this experiment as it provides the least acoustic and semantic cues of all sentence types, as illustrated by these proportion of looking time analyses.

As for Experiment 2, we conducted two-tailed *t*-tests comparing looks to the target picture to chance at target word onset, i.e. the first frame visible in Fig. 6. We find that at target word onset, the proportion of looks to the target is significantly above chance for Clear and IDS sentences with High Predictability semantic context (Clear $t(17) = 4.2$, $p < 0.001$; IDS $t(17) = 3.7$, $p < 0.001$). Unlike in Experiment 2, at target word onset the proportion of target fixations was not significant yet for Conversational sentences with High Predictability context ($t(17) = 1.9$, p = n.s.); nor for Conversational or Clear sentences with Low Predictability semantic contexts (Conversational $t(17) = 1.9$, p = n.s.; Clear $t(17) = 0.2$, p = n.s.). Target fixations for IDS Low Predictability sentences were marginally significant from chance ($t(17) = 2.1$, $p = 0.05$).

To further asses the looks to the target object over time and compare participants' looking behavior on the six different

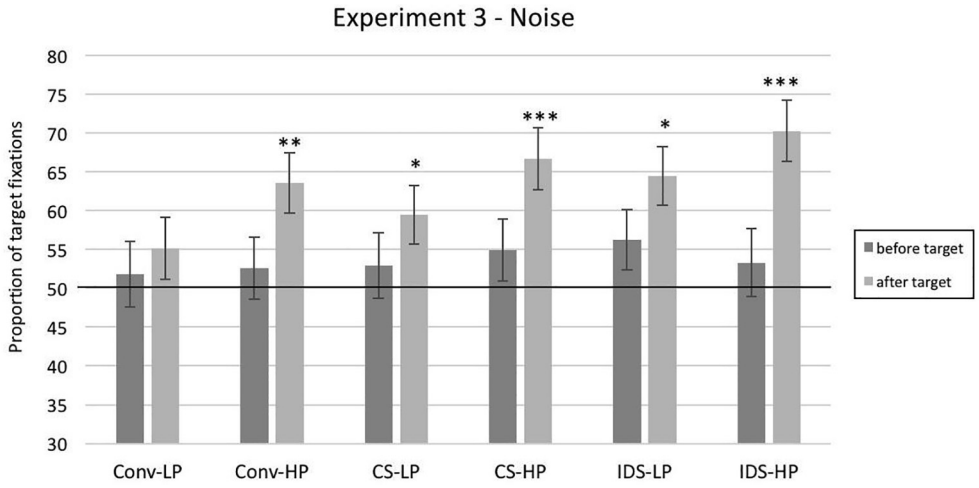


Fig. 5. Proportions of target fixations in noise, as a function of the total looking time during a one second window of analysis, with standard errors. Bars illustrate target fixations during the first second of each trial (before target, dark grey) versus the first second of each trial starting at target word onset (after target, light grey). The line at 50% represents chance. Stars indicate significant differences from chance (one star at the $p < 0.05$ level, two stars at the $p < 0.01$ level, three stars at the $p < 0.001$ level). Data is broken down by speaking style (Conversational Speech (Conv), Clear Speech (CS), Infant-Directed Speech (IDS)) and by semantic context (High Predictability (HP), Low Predictability (LP)).

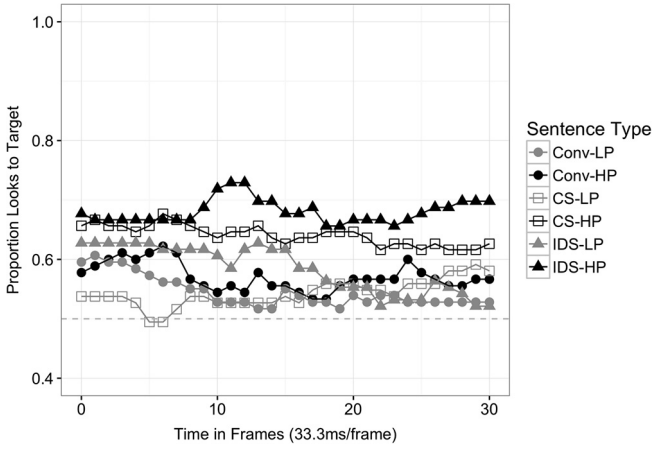


Fig. 6. Participants' target fixations, in noise. Lines illustrate the proportion of looks to the target picture in a one-second window after target word onset, by showing the looks to the target picture calculated as the proportion of looks to the target compared to the total looks to the target and distractor picture for each 33.33 ms time frame. The dotted line at 0.5 represents chance. Data is broken down by speaking style (Conversational Speech (Conv), Clear Speech (CS), Infant-Directed Speech (IDS)) and by semantic context (High Predictability (HP), Low Predictability (LP)).

Table 6
Experiment 3: summary of model fitting proportion of fixations to target. Intercept represents proportion of fixations to target for High Predictability sentences in Clear Speech.

	Estimate	Std. Error	*t* value	*p* value
Intercept (CS, HP)	0.67520	0.03737	18.067	<0.001
LP	−0.15660	0.02496	−6.275	<0.001
Conv	−0.08508	0.02496	−3.408	<0.001
IDS	−0.00249	0.02496	−0.1	0.92
Time (Frame)	−0.00160	0.00101	−1.58	0.11
LP*Conv	0.15850	0.03558	4.454	<0.001
LP*IDS	0.11640	0.03530	3.296	<0.001
LP*Time	0.00374	0.00143	2.613	<0.01
Conv*Time	0.00051	0.00143	0.359	0.72
IDS*Time	0.00180	0.00143	1.258	0.21
LP*Conv*Time	−0.00489	0.00204	−2.403	<0.05
LP*IDS*Time	−0.00703	0.00202	−3.476	<0.001

Random effects:

	Variance	Std. Deviation
Participant	0.01953	0.1398
Residual	0.04561	0.2136

sentence types, a mixed-effects linear regression model was fit to the proportion of looks to the target object at each 33.33 ms time frame, during the first one-second window after target word onset. Speaking style (Conversational, CS, IDS), semantic context (HP, LP), and time (frame) were tested as fixed effects, and participant was included as a random intercept. The results of Experiment 3 are summarized in Fig. 6.

As for Experiment 2, models were tested against each other using ANOVA, and the referent (intercept) of the best fitting model was releveled for pairwise comparisons. As was the case for Experiment 2, the best-fitting regression model for listening in noise included a three-way interaction among speaking style, semantic content, and time; the model summary is presented in Table 6. For the Clear Speech sentences, the proportion of fixations to target was greater in the High Predictability sentences than the Low Predictability sentences ($\beta = -0.16$, $t(3288) = -6.275$, $p < 0.001$); there was no difference between semantic contexts for IDS and Conversational ($p > 0.05$). Among the High Predictability sentences, there were significantly fewer fixations to target for Conversational than for either Clear Speech or IDS (Clear vs. Conversational: $\beta = -0.09$, $t(3288) = -3.408$, $p < 0.001$; IDS vs. Conversational: $\beta = -0.08$, $t(3288) = -3.309$, $p < 0.001$; CS vs. IDS: $p > 0.05$)). For the Low Predictability sentences, the pattern was different: listeners had significantly lower proportions of looks to the target for sentences produced in Clear Speech than in Conversational ($\beta = -0.07$, $t(3288) = 2.895$, $p < 0.01$), and Low Predictability IDS did not significantly differ from the other Low Predictability styles ($ps > 0.05$). There was a significant increase in looks over time only for the Low Predictability sentences in Clear Speech ($\beta = 0.002$, $t(3288) = 2.115$, $p < 0.05$); the slopes in all other conditions did not differ from zero ($ps > 0.05$). As such, looks to target in Low Predictability Clear Speech increased significantly faster than in the

other Low Predictability speaking styles (IDS: $\beta = -0.005$, $t(3288) = -3.658$, $p < 0.001$; Conversational: $\beta = -0.004$, $t(3288) = -3.021$, $p < 0.01$), and the increase in Low Predictability Clear Speech was faster than in High Predictability Clear Speech ($\beta = -0.004$, $t(3288) = -2.613 p < 0.01$). There were no other differences between conditions in terms of rate of increase in fixations ($ps > 0.05$). Since time was only significant for Low Predictability Clear Speech sentences, and since the increase in looks to target over time for Low Predictability Clear Speech was minimal (looks to target increased 0.2% in each frame), no further analyses (e.g. breakpoint regression or a binned analysis) were conducted to explore changes over time.

As in quiet, adult listeners benefited from semantic context in the enhanced acoustic–phonetic conditions in a more challenging listening condition. Fixation rates were higher for High Predictability Clear Speech and High Predictability IDS sentences than for the other conditions. Unlike in quiet, Clear Speech and IDS modifications contributed equally to the increased fixation rates for High Predictability sentences. The results revealed no context or speaking style effect for Conversational sentences. In the Low Predictability condition in general, acoustic–phonetic enhancements did not aid word recognition. Also, unlike in quiet, there was no evidence of further increases in fixation rates over the target word, i.e., looks to target did not change over the analysis window. In challenging listening conditions, the benefit of the context and speaking style that was maximally evident at the onset of the target word did not increase further even after the target word was heard.

5. General discussion

This study examined the influence of two listener-oriented speaking styles (Clear Speech and IDS) and the presence or absence of a high predictability semantic context on word recognition. Young adult listeners were tested in 'offline' and 'online' tasks, in quiet and in noise. We first tested the combined effects of IDS and Clear Speech and High- and Low-Predictability semantic contexts on word recognition in noise (Experiment 1a) and on pleasantness ratings (Experiment 1b). Results showed that both the speaking style modifications and High Predictability semantic context led to improved word recognition in noise. This is in line with evidence from a number of studies showing intelligibility benefit for Clear Speech and High Predictability context (Bradlow & Alexander, 2007; Smiljanic & Sladen, 2013; Van Engen et al., 2012). Here, we also showed that IDS in combination with High Predictability context resulted in the most accurate transcriptions of sentence-final target words masked with noise. The intelligibility benefit of IDS for young adult listeners suggests that the effect of IDS on infants' and young children's development cannot be attributed to affect and attractiveness alone (cf. Benders, 2013; McMurray et al., 2013). Rather the specific acoustic–phonetic modifications, segmental and suprasegmental, typically found in IDS likely contributed to the increased intelligibility found here (Adriaans & Swingley, 2017; Cristià, 2010, 2013; Eaves et al., 2016; Englund, 2005; Sundberg & Lacerda, 1999; Wang et al., 2016). It remains to be determined whether the same perceptual advantage is found for children listening to Clear Speech, which lacks the affective and prosodic characteristics responsible for children's heightened attention when listening to IDS.

Differences in intelligibility levels (Experiment 1a) as well as pleasantness ratings (Experiment 1b) suggest that the acoustic–phonetic modifications differed for the two listener-oriented styles elicited in the current study. This is in line with previous work established that talkers tailor their spoken output to meet the demands of the communicative situation, which results in different acoustic–phonetic modifications (Hazan & Baker, 2011; Lam & Tjaden, 2013; Lam, Tjaden, & Wilding, 2012; Smiljanic & Gilbert, 2017) and that IDS differs from Lombard speech, foreigner-directed speech, and read speech (Martin et al., 2014; Scarborough et al., 2007; Tang et al., 2017). The intelligibility advantage of the IDS over Clear Speech in Experiment 1a could be attributed to the effectiveness of some of the specific IDS modifications, such as enhanced stress and F0 cues, against the masking effect of the noise (Liss, Spitzer, Caviness, Adler, & Edwards, 1998; Mattys et al., 2005; Welby, 2007). Future work should compare acoustic characteristics of IDS and Clear Speech elicited for the same materials by more talkers to better understand in what ways the two modifications are similar and in what ways they are different.

The current study provides evidence that listeners were able to utilize some of these modifications to enhance speech recognition in noise; however, it is important to keep in mind that a direct link between any one acoustic–articulatory modification and increased intelligibility remains rather tenuous (Godoy, Koutsogiannaki, & Stylianou, 2014; Krause, 2001; Krause & Braida, 2004; Liu & Zeng, 2006; Picheny, Durlach, & Braida, 1989; Tjaden, Kain, & Lam, 2014; Uchanski, Choi, Braida, Reed, & Durlach, 1996). The current study was not designed to assess how much each of these individual acoustic–phonetic features contributes to intelligibility exactly, but rather to illustrate how conversational-to-clear and adult-to-infant speech modification affect overall speech processing in quiet and in noise.

The pleasantness results in Experiment 1b revealed that young adult listeners rated IDS as neutral, demonstrating that the exaggerated prosodic characteristics of IDS led to neither negative assessment nor diminished intelligibility. In contrast, adult-directed Clear Speech was judged as significantly less pleasant in the High Predictability context. It is possible that when listeners hear Clear Speech in a non-challenging, quiet acoustic environment they may perceive it as "unnecessary" and potentially even grating or condescending. Morgan and Ferguson (2017) found that both young adults with normal hearing and older adults with hearing impairment rated Clear Speech as angrier compared to Conversational speech. Some of the Conversational-to-clear speech modifications are similar to the expression of anger in speech (Banse & Scherer, 1996; Scherer, Johnstone, & Klasmeyer, 2003) and may contribute to the perceived unpleasantness especially in situations when the contextual cues already contributed to high intelligibility. It is also possible that IDS in High Predictability context was not rated as unpleasant as Clear Speech because adult listeners perceived the infant-directed style as misdirected, but not condescending. Pleasantness results did reveal more individual variation in the ratings of IDS sentences compared to Clear Speech and Conversational sentences. The varying degrees of subjective perception of IDS may be related to the listener's

experience interacting with young children and exposure to and usage of IDS. Recently, Smiljanic and Gilbert (2017) argued that acoustic–phonetic characteristics of the talker's speech, rather than the interaction between talker- and listener-related factors, determined, to a large extent, intelligibility variation for talkers of different ages (also Bradlow, Torretta, & Pisoni, 1996; Hazan & Markham, 2004). This suggests that the listeners' IDS pleasantness ratings in the current study could also be more related to the acoustic characteristics of IDS, rather than their experience with hearing IDS and talking to children. Future work should consider individual differences in pleasantness ratings for various speaking style adaptations and how these ratings relate to intelligibility variation. The perceived emotion of speaking styles should also be examined for sentences in noise so that the intelligibility-enhancing nature of Clear Speech is assessed in a more 'expected' setting.

With the established intelligibility benefit of speaking styles and context, Experiments 2 and 3 tested online word recognition in quiet and noise. In both environments, combined acoustic–phonetic and semantic enhancements contributed to improved word recognition. When listeners heard High Predictability Clear Speech and High Predictability IDS sentences in quiet, fixations to target were nearly at ceiling at the word onset. Enhanced acoustic cues allowed listeners to utilize semantic context to build up predictions about the upcoming final word, such that they were already fixating the target word at its onset. In fact, hearing the target word itself contributed little to their fixations further as evident in shallow fixation slopes across the target words in High Predictability IDS and High Predictability Clear Speech. Future work is needed to determine how early the facilitatory effect of speaking styles and context can be observed in a sentence. Lack of context, even when acoustic–phonetic cues were enhanced, provided little evidence to the listeners as to which picture to fixate, resulting in overall lower fixation rates on the target word at its onset. The target word itself contributed more significantly to word recognition in the contexts where preceding information was insufficient, as seen in greater fixation slope changes for Low Predictability sentences. Listeners thus benefited most from the exaggerated acoustic–phonetic cues on the target word when little semantic information was available to help them predict it. These findings are in line with the results from Experiment 1, in that adult listeners benefited from both listener-oriented speaking styles and contextual cues when processing speech in quiet. An interesting difference between the two experiments is that IDS showed an intelligibility advantage over Clear Speech in Experiment 1a, but Clear Speech contributed more to word recognition in Experiment 2. The main difference is that in Experiment 1a, listeners were hearing sentences mixed with noise, while in Experiment 2, they heard sentences in quiet. As mentioned above, it is possible that the specific acoustic–phonetic characteristics of the IDS sentences made them stand out more from the masking effect of the noise so that listeners were overall more accurate when identifying IDS words in noise.

As expected, the task of fixating one of the two pictures on the screen when listening to speech in noise became more difficult. As in Experiment 2 (and 1a), a combination of contextual cues and speaking style modifications (High Predictability sentences in Clear Speech and IDS) enabled the most reliable and rapid lexical access. However, listeners in Experiment 3 fixated the target pictures overall less than in Experiment 2. Even when hearing the High Predictability IDS and High Predictability Clear Speech sentences, listeners were less certain of which picture to look at by the time the target word was produced. The noise disrupted their ability to utilize contextual cues to the same extent as in quiet. Furthermore, unlike in quiet, there were no significant changes in proportions of fixations over the course of the target word for any sentence type. In noise, the beneficial effect of the acoustic–phonetic enhancements on the target word itself was diminished for Low Predictability sentences, precisely the context in which we saw biggest recognition gains in quiet. Unlike in Experiment 1, IDS did not provide a bigger processing advantage compared to Clear Speech. Even though listeners were ultimately more accurate in word identification in noise when hearing IDS sentences compared to Clear Speech sentences, this advantage was not evident in fixation rates on the target word during online processing. The results thus reveal a processing cost even for enhanced acoustic and semantic information when processing speech in noise, which is not evident when examining 'offline' accuracy alone.

The difficulty of speech processing in noise could arise from multiple sources. Processing degraded speech can be effortful (cf. effortfulness hypothesis, McCoy et al., 2005) and can tax working memory (Francis & Nusbaum, 2009; Francis, 2010; Rabbitt, 1968). Perceptual effort to correctly recognize words masked by noise may diminish the cognitive resources for building up the meaning over the course of the evolving High Predictability sentences. Similar difficulties in using semantic context in challenging listening situations were observed for children and non-native listeners (Bradlow & Alexander, 2007; Smiljanic & Sladen, 2013). The effect of noise can be evident at all levels of linguistic processing (Mattys et al., 2012). That is, the masking effect of noise can disrupt mapping of acoustic–phonetic features to segmental representations and access to suprasegmental information, such as F0 variation and the distribution of pauses, which indicates prosodic boundaries. This, in turn, can disrupt mapping to lexical representations leading to increased lexical uncertainty, selecting wrong lexical items, or failing to access a lexical item at all. While not implicating any one of these processes directly, the current findings demonstrate that noise disrupts processing both at the signal-related (accessing acoustic–phonetic cues) and higher-level linguistic structural (utilizing sentence context) levels. How exactly speaking style and semantic enhancements aid these different processes and what is the relative timing of these sources of enhancements remains a pressing goal for future work.

Taken together, the results from this study provide insights into the extent and limits of the influence of listener-oriented speaking styles and semantic context on offline word-identification in noise and online word recognition in quiet and noise. Overall the results demonstrate that spoken word processing is enhanced through speaking style modifications and the presence of high-predictability contextual information. The online word recognition results revealed fine-grained differences between processing of different speaking styles and semantic context in silence compared to noise, and provided novel insights into the locus of the intelligibility benefit. The results also provide evidence that the acoustic–phonetic modifications of IDS lead to improved word recognition for young

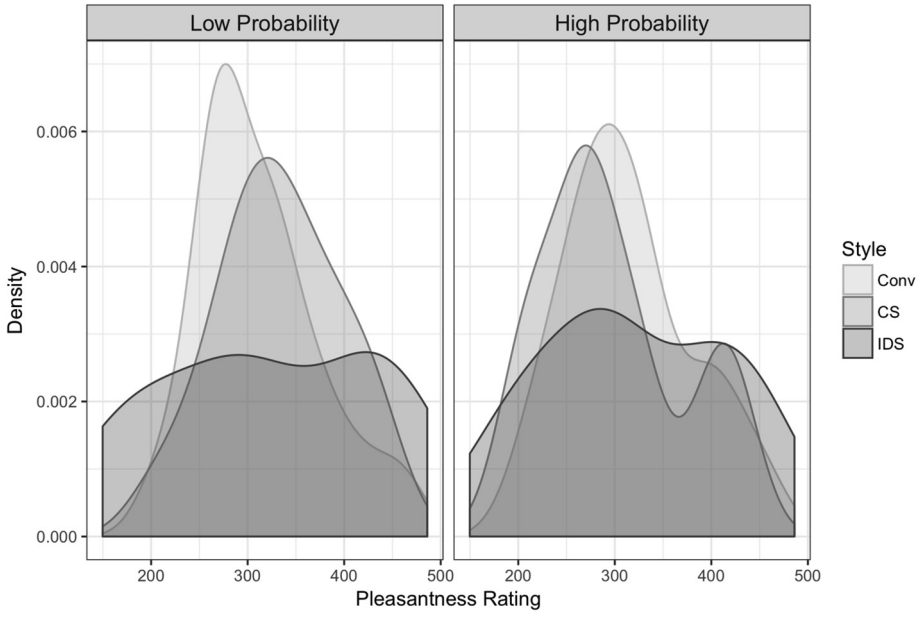
adult listeners. To fully understand the contribution of the various Clear Speech and IDS features on offline and online measures of word recognition in quiet and noise, children should be tested (preliminary findings are reported in Van der Feest et al., 2016;). Additional work is required to further clarify the relationships among intelligibility-enhancing cues and the way that children and young adults use these modifications to aid spoken word processing.

Acknowledgements

We thank members of the UT Sound Lab and Child Language Lab Austin for help with data collection and coding. Portions of this work were presented at Acoustical Society of America Meetings, Providence, Rhode Island, May 2014; Jacksonville, Florida, November 2015; and the Boston University Conference on Language Development, November 2016, Boston, Massachusetts.

Appendix A

Pleasantness ratings (Experiment 1b), distribution of raw scores. Distributions are expressed in density per raw pleasantness rating, for each sentence type: Conversational Speech (Conv), Clear Speech (CS), and Infant-Directed Speech (IDS).



Appendix B

Details of bin analyses (Experiment 2). For the binned analyses, the 31 frames (0–30) were divided into three bins at frames 9 and 19: frames 0–9 in Bin 1, frames 10–19 in Bin 2, and frames 20–30 in Bin 3.

Bin 1

Summary of model fitting proportion of fixations to target in frames 0–9. Intercept represents proportion of fixations to target for high predictability sentences in Clear Speech.

	Estimate	Std. Error	*t* value	*p* value
Intercept (CS, High)	0.5485	0.0407	13.469	<0.001
Low Predictability	−0.4958	0.0274	−18.092	<0.001
Conversational	−0.2313	0.0274	−8.439	<0.001
IDS	−0.0687	0.0274	−2.505	<0.05
Time (Frame)	0.01	0.00275	3.614	<0.001
Low*Conv	0.3497	0.0388	9.024	<0.001
Low*IDS	−0.0123	0.0388	−0.316	0.75

Random effects:

	Variance	Std. Deviation
Participant	0.04179	0.2044
Residual	0.13893	0.3727

Bin 2

Summary of model fitting proportion of fixations to target in frames 10–19. Intercept represents proportion of fixations to target for high predictability sentences in Clear Speech.

Model:

	Estimate	Std. Error	t-value	*p*-value
Intercept (CS, High)	0.6668	0.1042	6.399	<0.001
Low Predictability	−0.8296	0.1361	−6.095	<0.001
Conversational	−0.2385	0.1361	−1.752	0.0799
IDS	−0.04648	0.1361	−0.342	0.7327
Time (Frame)	−0.004090	0.006511	−0.628	0.5300
Low*Conv	0.4495	0.1925	2.335	0.0196
Low*IDS	−0.002649	0.1925	−0.014	0.9890
Time*Conv	0.01291	0.009207	1.402	0.1609
Time*IDS	−0.0002675	0.009207	−0.029	0.9768
Low*Time	0.04068	0.009207	4.418	<0.001
Low*Conv*Time	−0.01800	0.01302	−1.382	0.1670
Low*IDS*Time	−0.003487	0.01302	−0.268	0.7889

Random effects:

	Variance	Std. Deviation
Participant	0.05902	0.2429
Residual	0.12939	0.3597

Bin 3

Summary of model fitting proportion of fixations to target in frames 20–30. Intercept represents proportion of fixations to target for high predictability sentences in Clear Speech.

Model:

	Estimate	Std. Error	t-value	*p*-value
Intercept (CS, High)	0.5876	0.04769	12.321	<0.001
Low probability	−0.03049	0.02226	−1.370	0.17085
Conversational	0.03309	0.02226	1.487	0.13726
IDS	−0.007453	0.02226	−0.335	0.73777
Low*Conv	0.01509	0.03148	0.479	0.63163
Low*IDS	−0.09057	0.03148	−2.877	0.00405

Random effects:

	Variance	Std. Deviation
Participant	0.07498	0.2738
Residual	0.10081	0.3175

Appendix C

Test orders Experiment 2 and 3.

ORDER 1					
Trial #	*Left Image*	*Right Image*	*Predictability (H/L)*	*Style (Clear/Convl/IDS)*	*Auditory Stimulus, (# of syllables)*
1	ball	bed	H	clear	We played catch with the ball
2	horse	sock	H	clear	We put the shoe on after the sock
3	fish	fork	H	clear	I went to the pond and caught a fish
4	book	bus	H	Conversational	I like to read a book

(*continued*)

ORDER 1

Trial #	*Left Image*	*Right Image*	*Predictability (H/L)*	*Style (Clear/Convl/IDS)*	*Auditory Stimulus, (# of syllables)*
5	corn	cup	H	Conversational	I drink juice out of a cup
6	shoe	doll	H	Conversational	The girl played with her doll
7	cheese	chair	H	IDS	Mice like to eat cheese
8	fish	fork	H	IDS	I eat spaghetti with a fork
9	pants	pig	H	IDS	I fell and ripped my pants
	filler swimming fish				
10	cat	car	L	clear	She pointed at the car
11	sock	horse	L	clear	She pointed at the horse
12	pig	pants	L	clear	She pointed at the pig
13	corn	cup	L	IDS	Dad pointed at the corn
14	book	bus	L	IDS	Mom looked at the bus
15	bed	bal	L	IDS	We read about the bed
16	doll	shoe	L	Conversational	He pointed at the shoe
17	chair	cheese	L	Conversational	She talked about the chair
18	cat	car	L	Conversational	We pointed at the cat
	filler flying bird				
19	fish	fork	H	clear	I went to the pond and caught a fish
20	horse	sock	H	clear	We put the shoe on after the sock
21	ball	bed	H	clear	We played catch with the ball
22	cup	cup	H	Conversational	I drink juice out of a cup
23	doll	shoe	H	Conversational	The girl played with her doll
24	book	bus	H	Conversational	I like to read a book
25	fish	fork	H	IDS	I eat spaghetti with a fork
26	cheese	chair	H	IDS	Mice like to eat cheese
27	pig	pants	H	IDS	I fell and ripped my pants
	filler bouncing ducks				
28	car	cat	L	clear	She pointed at the car
29	pig	pants	L	clear	She pointed at the pig
30	sock	horse	L	clear	She pointed at the horse
31	cup	corn	L	IDS	He talked about the cup
32	book	bus	L	IDS	Mom looked at the bus
33	ball	bed	L	IDS	We read about the bed
34	chair	cheese	L	Conversational	She talked about the chair
35	doll	shoe	L	Conversational	He pointed at the shoe
36	cat	car	L	Conversational	We pointed at the cat

ORDER 2

Trial #	*Left Image*	*Right Image*	*Predictability (H/L)*	*Style (clear/Conv/IDS)*	*Auditory Stimulus*
1	ball	bed	H	Conversational	We played catch with the ball
2	horse	sock	H	Conversational	We put the shoe on after the sock
3	fish	fork	H	Conversational	I went to the pond and caught a fish
4	book	bus	H	IDS	I like to read a book
5	corn	cup	H	IDS	I drink juice out of a cup
6	shoe	doll	H	IDS	The girl played with her doll
7	cheese	chair	H	clear	Mice like to eat cheese
8	fish	fork	H	clear	I eat spaghetti with a fork
9	pants	pig	H	clear	I fell and ripped my pants
	filler swimming fish				
10	cat	car	L	Conversational	She pointed at the car
11	sock	horse	L	Conversational	She pointed at the horse
12	pig	pants	L	Conversational	She pointed at the pig
13	corn	cup	L	clear	Dad pointed at the corn
14	book	bus	L	clear	Mom looked at the bus
15	bed	bal	L	clear	We read about the bed

(*continued on next page*)

(*continued*)

ORDER 2

Trial #	*Left Image*	*Right Image*	*Predictability (H/L)*	*Style (clear/Conv/IDS)*	*Auditory Stimulus*
16	doll	shoe	L	IDS	He pointed at the shoe
17	chair	cheese	L	IDS	She talked about the chair
18	cat	car	L	IDS	We pointed at the cat
	filler flying bird				
19	fish	fork	H	Conversational	I went to the pond and caught a fish
20	horse	sock	H	Conversational	We put the shoe on after the sock
21	ball	bed	H	Conversational	We played catch with the ball
22	corn	cup	H	IDS	I drink juice out of a cup
23	doll	shoe	H	IDS	The girl played with her doll
24	book	bus	H	IDS	I like to read a book
25	fish	fork	H	clear	I eat spaghetti with a fork
26	cheese	chair	H	clear	Mice like to eat cheese
27	pig	pants	H	clear	I fell and ripped my pants
	filler bouncing ducks				
28	car	cat	L	Conversational	She pointed at the car
29	pig	pants	L	Conversational	She pointed at the pig
30	sock	horse	L	Conversational	She pointed at the horse
31	cup	corn	L	clear	He talked about the cup
32	book	bus	L	clear	Mom looked at the bus
33	ball	bed	L	clear	We read about the bed
34	chair	cheese	L	IDS	She talked about the chair
35	doll	shoe	L	IDS	He pointed at the shoe
36	cat	car	L	IDS	We pointed at the cat

ORDER 3

Trial #	*Left Image*	*Right Image*	*Predictability (H/L)*	*Style (Clear/Convl/IDS)*	*Auditory Stimulus*
1	ball	bed	H	IDS	We played catch with the ball
2	horse	sock	H	IDS	We put the shoe on after the sock
3	fish	fork	H	IDS	I went to the pond and caught a fish
4	book	bus	H	clear	I like to read a book
5	corn	cup	H	clear	I drink juice out of a cup
6	shoe	doll	H	clear	The girl played with her doll
7	cheese	chair	H	Conversational	Mice like to eat cheese
8	fish	fork	H	Conversational	I eat spaghetti with a fork
9	pants	pig	H	Conversational	I fell and ripped my pants
	filler swimming fish				
10	cat	car	L	IDS	She pointed at the car
11	sock	horse	L	IDS	She pointed at the horse
12	pig	pants	L	IDS	She pointed at the pig
13	corn	cup	L	Conversational	Dad pointed at the corn
14	book	bus	L	Conversational	Mom looked at the bus
15	bed	bal	L	Conversational	We read about the bed
16	doll	shoe	L	clear	He pointed at the shoe
17	chair	cheese	L	clear	She talked about the chair
18	cat	car	L	clear	We pointed at the cat
	filler flying bird				
19	fish	fork	H	IDS	I went to the pond and caught a fish
20	horse	sock	H	IDS	We put the shoe on after the sock
21	ball	bed	H	IDS	We played catch with the ball
22	corn	cup	H	clear	I drink juice out of a cup
23	doll	shoe	H	clear	The girl played with her doll
24	book	bus	H	clear	I like to read a book
25	fish	fork	H	Conversational	I eat spaghetti with a fork
26	cheese	chair	H	Conversational	Mice like to eat cheese
27	pig	pants	H	Conversational	I fell and ripped my pants

(*continued*)

ORDER 3					
Trial #	*Left Image*	*Right Image*	*Predictability (H/L)*	*Style (Clear/Convl/IDS)*	*Auditory Stimulus*
	filler bouncing ducks				
28	car	cat	L	IDS	She pointed at the car
29	pig	pants	L	IDS	She pointed at the pig
30	sock	horse	L	IDS	She pointed at the horse
31	cup	corn	L	Conversational	He talked about the cup
32	book	bus	L	Conversational	Mom looked at the bus
33	ball	bed	L	Conversational	We read about the bed
34	chair	cheese	L	clear	She talked about the chair
35	doll	shoe	L	clear	He pointed at the shoe
36	cat	car	L	clear	We pointed at the cat

ORDER 4	(equal to order 1, with predictability reversed and half the target sides randomly mixed)				
Trial #	*Left Image*	*Right Image*	*Predictability (H/L)*	*Style (Clear/Conv/IDS)*	*Auditory Stimulus*
1	ball	bed	L	clear	Mom pointed at the ball
2	horse	sock	L	clear	Mom looked at the sock
3	fish	fork	L	clear	Mom talked about the fish
4	book	bus	L	Conversational	Dad looked at the book
5	corn	cup	L	Conversational	He talked about the cup
6	shoe	doll	L	Conversational	We read about the doll
7	cheese	chair	L	IDS	He looked at the cheese
8	fish	fork	L	IDS	Dad read about the fork
9	pants	pig	L	IDS	He read about the pants
	filler swimming fish				
10	car	cat	H	clear	We drove to the store in our car
11	horse	sock	H	clear	I learned how to ride a horse
12	pants	pig	H	clear	The farmer fed the pig
13	cup	corn	H	IDS	Dad pointed at the corn
14	bus	book	H	IDS	Dad rides to work on the bus
15	bal	bed	H	IDS	I fell asleep on my bed
16	shoe	doll	H	Conversational	I know how to tie a shoe
17	cheese	chair	H	Conversational	I sat down on the chair
18	car	cat	H	Conversational	The dog chased the cat
	filler flying bird				
19	fish	fork	L	clear	Mom talked about the fish
20	horse	sock	L	clear	Mom looked at the sock
21	ball	bed	L	clear	Mom pointed at the ball
22	corn	cup	L	Conversational	He talked about the cup
23	doll	shoe	L	Conversational	We read about the doll
24	book	bus	L	Conversational	Dad looked at the book
25	fish	fork	L	IDS	Dad read about the fork
26	cheese	chair	L	IDS	He looked at the cheese
27	pig	pants	L	IDS	He read about the pants
	filler bouncing ducks				
28	cat	car	H	clear	We drove to the store in our car
29	pants	pig	H	clear	The farmer fed the pig
30	horse	sock	H	clear	I learned how to ride a horse
31	corn	cup	H	IDS	I drink juice out of a cup
32	bus	book	H	IDS	Dad rides to work on the bus
33	bed	ball	H	IDS	I fell asleep on my bed
34	cheese	chair	H	Conversational	I sat down on the chair
35	shoe	doll	H	Conversational	I know how to tie a shoe
36	car	cat	H	Conversational	The dog chased the cat

Order 5 was equal to Order 2 except with the Predictability (H/L) of each sentence reversed, and half the target sides randomly mixed; Order 6 was equal to Order 3 except with the Probabilities (H/L) of each sentence reversed, and half the target sides randomly mixed.

References

Adriaans, F., & Swingley, D. (2017). Prosodic exaggeration within infant-directed speech: Consequences for vowel learnability. *The Journal of the Acoustical Society of America, 141*(5), 3070–3078.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419–439.

Altmann, G. T. (2011). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica, 137*(2), 190–200.

Assmann, P., & Summerfield, Q. (2004). The perception of speech under adverse conditions. In *Speech processing in the auditory system* (pp. 231–308). New York: Springer.

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

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*, 614–636.

Bard, E. G., Shillcock, R. C., & Altmann, G. T. M. (1988). The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Perception and Psychophysics, 44*(395), 408.

Bard, E. G., Sotillo, C., Kelly, M. L., & Aylett, M. P. (2001). Taking the hit: Leaving some lexical competition to be resolved post-lexically. *Language and Cognitive Processes, 16*(731), 737.

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

Ben-David, B. M., Chambers, C. G., Daneman, M., Pichora-Fuller, M. K., Reingold, E. M., & Schneider, B. A. (2010). Effects of aging and noise on real-time spoken word recognition: Evidence from eye movements. *Journal of Speech, Language, and Hearing Research, 54*(1), 243–262.

Benders, T. (2013). Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behavior and Development, 36*(4), 847–862.

Boersma, P., & Weenink, D. (2012). *Praat: Doing phonetics by computer. [Computer software]*. Amsterdam: Universiteit van Amsterdam.

Bouwer, S. M., Mitterer, H., & Huettig, F. (2013). Discourse context and the recognition of reduced and canonical spoken words. *Applied Psycholinguistics, 34*, 519–539.

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.

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.

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*, 80–97.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20*(3–4), 255–272.

Brouwer, S. M., & Bradlow, A. R. (2015). The temporal dynamics of spoken word recognition in adverse listening conditions. *Journal of Psycholinguistic Research, 45*(5), 1151–1160.

Buckler, H., Goy, H., & Johnson, E. K. (2018). What infant-directed speech tells us about the development of compensation for assimilation. *Journal of Phonetics, 66*, 45–62.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*(1), 84–107.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61*(5), 1584–1595.

Cooper, R. P., & Aslin, R. N. (1994). Developmental differences in infant attention to the spectral properties of infant-directed speech. *Child Development, 65*(6), 1663–1677.

Cristià, A. (2010). Phonetic enhancement of sibilants in infant-directed speech. *The Journal of the Acoustical Society of America, 128*(1), 424–434.

Cristia, A. (2013). Input to language: The phonetics and perception of infant-directed speech. *Language and Linguistics Compass, 7*(3), 157–170.

Cristia, A., & Seidl, A. (2014). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language, 41*(4), 913–934.

Dahan, D. (2010). The time course of interpretation in speech comprehension. *Current Directions in Psychological Science, 19*(2), 121–126.

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods, 2*(3), 292.

Eaves, B. S., Jr, Feldman, N. H., Griffiths, T. L., & Shafto, P. (2016). Infant-directed speech is consistent with teaching. *Psychological Review, 123*(6), 758.

Englund, K. T. (2005). Voice onset time in infant directed speech over the first six months. *First Language, 25*(2), 219–234.

Fallon, M., Trehub, S. E., & Schneider, B. A. (2002). Children's use of semantic cues in degraded listening environments. *The Journal of the Acoustical Society of America, 111*(5), 2242–2249.

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.

Fernald, A. (1984). The perceptual and affective salience of mothers' speech to infants. *The Origins and Growth of Communication*, 5–29.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, 1497–1510.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27*(2), 209.

Fernald, A., Pinto, J. P., Swingley, D., Weinbergy, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science, 9*(3), 228–231.

Fish, M. S., García-Sierra, A., Ramírez-Esparza, N., & Kuhl, P. K. (2017). Infant-directed speech in English and Spanish: Assessments of monolingual and bilingual caregiver VOT. *Journal of Phonetics, 63*, 19–34.

Francis, A. L. (2010). Improved segregation of simultaneous talkers differentially affects perceptual and cognitive capacity demands for recognizing speech in competing speech. *Attention, Perception, & Psychophysics, 72*(2), 501–516.

Francis, A. L., & Nusbaum, H. C. (2009). Effects of intelligibility on working memory demand for speech perception. *Attention, Perception, & Psychophysics, 71*(6), 1360–1374.

Friederici, A. D., Steinhauer, K., & Frisch, S. (1999). Lexical integration: Sequential effects of syntactic and semantic information. *Memory & Cognition, 27*(3), 438–453.

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology, 45*(2), 220–266.

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.

Godoy, E., Koutsogiannaki, M., & Stylianou, Y. (2014). Approaching speech intelligibility enhancement with inspiration from Lombard and clear speaking styles. *Computer Speech & Language, 28*(2), 629–647.

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby) talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science, 24*(5), 339–344.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language, 14*(1), 23–45.

Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy, 18*(5), 797–824.

Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology, 24*(1), 14.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Attention, Perception, & Psychophysics, 28*(4), 267–283.

Harnsberger, J. D., Wright, R., & Pisoni, D. B. (2008). A new method for eliciting three speaking styles in the laboratory. *Speech Communication, 50*(4), 323–336.

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.

Hazan, V., & Markham, D. (2004). Acoustic-phonetic correlates of talker intelligibility for adults and children. *The Journal of the Acoustical Society of America, 116*(5), 3108–3118.

Hintz, F., & Scharenborg, O. E. (2016). The effect of background noise on the activation of phonological and semantic information during spoken-word recognition. *Interspeech*, 2816–2820. San Francisco, CA.

Hollich, G. (2005). *Supercoder: A program for coding preferential looking version 1.5. [Computer software]*. West Lafayette: Purdue University.

Johnson, E. K., Lahey, M., Ernestus, M., & Cutler, A. (2013). A multimodal corpus of speech to infant and adult listeners. *The Journal of the Acoustical Society of America, 134*(6), EL534–EL540.

Johnson, E. K., McQueen, J. M., & Huettig, F. (2011). Toddlers' language-mediated visual search: They need not have the words for it. *The Quarterly Journal of Experimental Psychology, 64*(9), 1672–1682.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America, 61*(5), 1337–1351.

Kaplan, P. S., Goldstein, M. H., Huckeby, E. R., & Panneton Cooper, R. (1995). Habituation, sensitization, and infants' responses to motherese speech. *Developmental Psychobiology, 28*, 45–57.

Keerstock, S., & Smiljanic, R. (2018). Effects of intelligibility on within- and cross-modal sentence recognition memory for native and non-native listeners. *Journal of the Acoustical Society of America, 144*(5), 2871–2881.

Kemler-Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language, 16*(1), 55–68.

Knoll, M., Scharrer, L., & Costall, A. (2009). Are actresses better simulators than female students? The effects of simulation on prosodic modifications of infant-and foreigner-directed speech. *Speech Communication, 51*(3), 296–305.

Krause, J. C. (2001). *Properties of naturally produced clear speech at normal rates and implications for intelligibility enhancement* Doctoral Dissertation. Massachusetts: Institute of Technology.

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.

Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Developmental Science, 10*(1), 110–120.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., et al. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*(5326), 684–686.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-30. https://CRAN.R-project.org/package=lmerTest.
Lam, J., & Tjaden, K. (2013). Intelligibility of clear speech: Effect of instruction. *Journal of Speech, Language, and Hearing Research, 56*(5), 1429–1440.
Lam, J., Tjaden, K., & Wilding, G. (2012). Acoustics of clear speech: Effect of instruction. *Journal of Speech, Language, and Hearing Research, 55*(6), 1807–1821.
Liss, J. M., Spitzer, S., Caviness, J. N., Adler, C., & Edwards, B. (1998). Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *The Journal of the Acoustical Society of America, 104*(4), 2457–2466.
Liu, H. M., Kuhl, P. K., & Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science, 6*(3).
Liu, S., & Zeng, F. G. (2006). Temporal properties in clear speech perception. *The Journal of the Acoustical Society of America, 120*(1), 424–432.
Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*(1), 1.
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.
Marslen-Wilson, W. D. (1984). Function and process in spoken word recognition: A tutorial review. In *Attention and performance: Control of language processes* (pp. 125–150). Erlbaum.
Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human perception and performance, 15*(3), 576.
Martin, A., Igarashi, Y., Jincho, N., & Mazuka, R. (2016). Utterances in infant-directed speech are shorter, not slower. *Cognition, 156*, 52–59.
Martin, A., Utsugi, A., & Mazuka, R. (2014). The multidimensional nature of hyperspeech: Evidence from Japanese vowel devoicing. *Cognition, 132*(2), 216–228.
Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology, 59*(3), 203–243.
Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7–8), 953–978.
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.
McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1–86.
McCoy, F. N., McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., Wingfield, A., et al. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *The Quarterly Journal of Experimental Psychology Section A, 58*(1), 22–33.
McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review, 15*(6), 1064–1071.
McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence? *Cognition, 129*(2), 362–378.
McQueen, J. M. (2007). Eight questions about spoken-word recognition. *The Oxford Handbook of Psycholinguistics*, 37–53.
Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology, 41*(5), 329.
Morgan, S., & Ferguson, S. H. (2017). Judgments of emotion in clear and conversational speech by young adults with normal hearing and older adults with hearing impairment. *Journal of Speech Language and Hearing Research, 60*(8), 1–10.
Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *The Journal of the Acoustical Society of America, 87*(6), 2705–2715.
Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*(3), 189–234.
Orfanidou, E., Davis, M. H., Ford, M. A., & Marslen-Wilson, W. D. (2011). Perceptual and response components in repetition priming of spoken words and pseudowords. *The Quarterly Journal of Experimental Psychology, 64*(1), 96–121.
Piazza, E. A., Iordan, M. C., & Lew-Williams, C. (2017). Mothers consistently alter their unique vocal fingerprints when communicating with infants. *Current Biology, 27*(1–6).
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.
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.
Picheny, M. A., Durlach, N. I., & Braida, L. D. (1989). Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *Journal of Speech, Language, and Hearing Research, 32*(3), 600–603.
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.
R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
Rabbitt, P. M. (1968). Channel-capacity, intelligibility and immediate memory. *The Quarterly Journal of Experimental Psychology, 20*(3), 241–248.
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*(supp. 2), S99–S105.
RStudio Team (2016). *RStudio: Integrated development for R*. Boston, MA: RStudio Inc.. http://www.rstudio.com/.
Scarborough, R., Dmitrieva, O., Hall-Lew, L., Zhao, Y., & Brenier, J. (2007). An acoustic study of real and imagined foreigner-directed speech. *Journal of the Acoustical Society of America, 121*(5), 3044.
Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 433–456). New York, NY: Oxford University Press.
Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime 1.0. [Computer Software]*. Pittsburgh, PA: Psychological Software Tools.
Schreiner, M. S., & Mani, N. (2017). Listen up! Developmental differences in the impact of IDS on speech segmentation. *Cognition, 160*, 98–102.
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), 1677–1688.
Smiljanić, R., & Bradlow, A. R. (2008). Temporal organization of English clear and conversational speech. *The Journal of the Acoustical Society of America, 124*(5), 3171–3182.
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.
Smiljanic, R., & Gilbert, R. C. (2017). Acoustics of clear and noise-adapted speech in children, young, and older adults. *Journal of Speech, Language, and Hearing Research, 60*(11), 3081–3096.
Smiljanic, R., & Sladen, D. (2013). Acoustic and semantic enhancements for children with cochlear implants. *Journal of Speech, Language, and Hearing Research, 56*(4), 1085–1096.
Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review, 27*(4), 501–532.
Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *The Journal of the Acoustical Society of America, 128*(1), 389–400.
Sundberg, U., & Lacerda, F. (1999). Voice onset time in speech to infants and adults. *Phonetica, 56*(3–4), 186–199.
Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition, 76*(2), 147–166.
Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 364*(1536), 3617–3632.
Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29*(6), 557–580.
Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017). Phonetic enhancement of Mandarin vowels and tones: Infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America, 142*(2), 493–503.
Tjaden, K., Kain, A., & Lam, J. (2014). Hybridizing Conversationalersational and clear speech to investigate the source of increased intelligibility in speakers with Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 57*(4), 1191–1205.
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.
Van der Feest, S. V. H., Blanco, C. P., & Smiljanic, R. (November, 2016; in preparation). Effects of speaking style and semantic context on online word recognition in young children. Poster presented at the 41th Boston university conference on language development. Boston, MA
Van der Feest, S. V. H., & Johnson, E. K. (2016). Input-driven differences in toddlers' perception of a disappearing phonological contrast. *Language Acquisition, 23*(2), 89–111.
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.
Van Engen, K. J., Chandrasekaran, B., & Smiljanic, R. (2012). Effects of speech clarity on recognition memory for spoken sentences. *PLoS One, 7*(9) e43753.
Wang, Y., Seidl, A., & Cristia, A. (2016). 11 Acoustic characteristics of infant-directed speech as a function of prosodic typology. *Dimensions of Phonological Stress, 311*.
Welby, P. (2007). The role of early fundamental frequency rises and elbows in French word segmentation. *Speech Communication, 49*(1), 28–48.
Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology/Revue canadienne de psychologie, 43*(2), 230.
Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition, 103*(1), 147–162.