



Published in final edited form as:

Lang Linguist Compass. 2009 January 1; 3(1): 236–264. doi:10.1111/j.1749-818X.2008.00112.x.

Speaking and Hearing Clearly: Talker and Listener Factors in Speaking Style Changes

Rajka Smiljanić* and Ann R. Bradlow
Northwestern University

Abstract

This article provides an overview of the research concerning the nature of the distinct, listener-oriented speaking style called ‘clear speech’ and its effect on intelligibility for various listener populations. We review major findings that identify talker, listener and signal characteristics that contribute to the characteristically high intelligibility of clear speech. Understanding the interplay of these factors sheds light on the interaction between higher level cognitive and lower-level sensory and perceptual factors that affect language processing. Clear speech research is, thus, relevant for both its theoretical insights and practical applications. Throughout the review, we highlight open questions and promising future directions.

Introduction

In everyday communication, the goal of talkers is to communicate their messages in a manner that is intelligible to listeners. When they are aware of a speech perception difficulty on the part of the listener due to background noise, a hearing impairment, or a different native language, talkers will naturally and spontaneously modify their speech in order to accommodate the listener. They will likely adopt a distinct speaking style called ‘clear speech’. In an attempt to make themselves more intelligible, talkers will typically speak more slowly, more loudly, and will articulate sounds in a more ‘exaggerated’ manner. Clear speech modifications are aimed at providing the listener with more salient acoustic cues in the speech signal that may enhance their ability to access and comprehend the message.

Conversational-to-clear speech articulatory changes present a within-talker variation along the hypo- to hyper-articulation continuum reflecting the trade-off between clarity of speech (listener-oriented output) and economy of effort (talker-oriented output) (H&H Theory, Lindblom 1990). In this respect, clear speech is related to other goal-oriented speaking styles, such as infant- and computer-directed speech and Lombard speech, in which talkers ‘adjust’ their output ‘on-line’ to meet the demands of their target audience or the communicative situation (Junqua 1993; Kuhl et al. 1997; Skowronski and Harris 2005). Clear speech also shares characteristics with variation due to prosodic strengthening and fast-to-slow rate modifications (Miller and Dexter 1988; Miller et al. 1986; de Jong 1995; Cho 2002, 2005; Hirata 2004; Cole et al. 2007). These adjustments due to speaking style, rate and prosodic/discourse structure can all be viewed as involving hyper-articulation changes, such as reduction of target undershoot (or alternatively, retargeting of articulatory gestures) and enhancement of phonemic contrasts. Clear speech is unique among these goal-oriented speaking styles in that it is specifically aimed at enhancing intelligibility for adult interlocutors with perceptual difficulties. Clear speech changes may, thus, involve different articulatory scaling mechanism

*Correspondence address: Rajka Smiljanić, Department of Linguistics, University of Texas at Austin, Calhoun Hall 407, 1 University Station B5100, Austin, TX 78712-0198, USA. rajka@austin.utexas.edu.

from, for instance, changes due to speaking rate pressures. Similarly, increased affective prosody aimed at capturing children's attention and increased vocal effort aimed at overcoming speaking in noise may be absent (or less present compared to child-directed and Lombard speech) from clear speech.¹

The goal of clear speech research is to identify, on the one hand, the consistent salient articulatory-acoustic features that characterize clear speech productions across talker groups and, on the other, which of these conversational-to-clear speech modifications impact intelligibility for various listener groups. Through a better understanding of talker-, listener- and signal-related factors that contribute to high intelligibility, clear speech research aims to shed light on the interaction between higher level cognitive and lower-level sensory and perceptual factors that affect language processing. Ultimately, clear speech research can inform us about underlying mechanisms of (1) speech production plasticity that allow talkers to adapt their output to the immediate communicative setting, and (2) speech perception processing that allow listeners to take advantage of these articulatory-acoustic modifications in accessing and segmenting the speech signal, accessing lexical items and their meaning and arriving at the right syntactic and prosodic structure of the intended utterance. Importantly, clear speech research has practical implications for developing speech-enhancement algorithms to improve quality of speech communication for populations with listening difficulties (e.g., hearing-impaired, hearing aid users, and non-native listeners), for clear speech training (e.g., for clinicians, family members of patients with hearing loss) and for speech technology applications (e.g., talker selection for training speech recognition). Clear speech research is, thus, relevant for both its theoretical insights and practical applications.

In this article, we review significant findings over the past few decades regarding principles that guide clear speech production and intelligibility. We build this examination of clear speech research on an excellent review by Uchanski (2005). While we include in the present article a discussion of some seminal articles discussed in Uchanski (2005), our main goal is to highlight more recent studies and to indicate promising new directions in clear speech research. Substantial progress has been made in identifying a wide range of acoustic/articulatory adjustments that accompany conversational-to-clear speech transformation and in demonstrating a clear speech intelligibility benefit for various listener groups. However, as will become evident, not all of the clear speech research goals have been achieved to date. We start by discussing various clear speech articulatory-acoustic modifications and signal distortions examined across studies along with the effect of these adjustments on speech intelligibility. We expand this review by including cross-language comparison of clear speech production and perception for various new populations, such as non-native talkers and listeners, as well as clinical populations, and the use of new materials in clear speech research. These new developments are aimed at exploring the connection between lower-level auditory-perceptual factors and higher level linguistic and cognitive factors in clear speech production and perception with the goal of better understanding intelligibility variability and language processing in real-life communicative situations. Throughout the review, we highlight open questions and promising new lines of research.

¹To keep the scope of this review manageable and focused, we refer the readers to the work cited above to find out more details about the phonetic outcomes for each of these distinct speaking styles and for the rate- and prosodic-related changes.

Talker-, Listener- and Signal-dependent Effects in Clear Speech

BACKGROUND: RESEARCH METHODOLOGY²

Before we review some of the important clear speech findings, we briefly clarify the use of terminology in this article and discuss most common methodological approaches across studies.³ In line with the Uchanski (2005) review and other work that predominantly uses conversational speech to contrast it with the clear speaking style, we use these same two terms throughout the article. It is important to note that both conversational and clear speech terms refer to read laboratory speech elicited by specific instructions given to talkers rather than to the spontaneous speech occurring in a more natural setting. The instructions most typically involve asking talkers to read the same set of materials twice: once in conversational and once in clear speaking style (Picheny et al. 1986; Schum 1996; Krause and Braida 2002; Ferguson and Kewley-Port 2002; Ferguson 2004; Smiljanić and Bradlow 2005). Clear speaking instructions include a version of the following: ‘read the materials as if you were talking to someone who is hearing impaired, not a native speaker of your language’ or ‘speak clearly and precisely’, etc. In most cases, talkers were allowed to interpret the task in a way that they saw appropriate. Most of the talkers were randomly selected and this resulted in a wide variability of clear speech strategies and intelligibility gains (see below for discussion). In Krause and Braida (2002), talkers with experience in public speaking were chosen for the task and they were further trained to produce clear speech. In this study, care was taken to monitor speakers’ speaking rate, such that clear speech sentences were elicited with a metronome at normal/ conversational rate. Recent work indicates that speech output may differ in fact depending on the type of instructions, that is, talking to a native vs. a non-native listener, child vs. adult vs. computer and in quiet vs. in noise may not result in equivalent acoustic-phonetic changes (Wassink et al. 2006; Smith 2007; Uther et al. 2007). Furthermore, new methods for eliciting reduced, citation and hyperarticulated speech in the laboratory are also being investigated (Harnsberger et al. 2008). Clarifying how the nature of the instructions and the task affect clear speech production presents an important new direction in clear speech research.

Number of talkers, listeners and signal characteristics varied across studies. For instance, Payton et al. (1994) and Ferguson and Kewley-Port (2002) had only one talker, Bradlow et al. (2003) had one male and one female talker, while Ferguson (2004) included 41 talkers. Similarly, the number of listeners varied from two hearing-impaired listeners in Payton et al. (1994) to 60 in Schum (1996) (types of perceptual deficits are discussed in more detail below). Perception tasks were most commonly sentence/word/syllable/sound-in-noise listening tests where subjects hear the stimuli and write down or repeat what they heard. Picheny et al. (1985) presented their listeners with stimuli in quiet while other presentations included reverberation, speech-shaped or wide-band noise, multitalker babble, etc. at various signal-to-noise ratios (Payton et al. 1994; Krause 2001; Ferguson and Kewley-Port 2002; Bradlow and Bent 2002). In addition, to varying presentation conditions, a few studies used speech cue enhancement manipulations in order to investigate possible connections between specific acoustic cues and intelligibility (Picheny et al. 1989; Stollman et al. 1994; Hazan and Simpson 1998, 2000; Liu and Zeng 2006). These studies strive to limit the number of acoustic cues that co-vary in conversational-to-clear speech adjustments and to establish more directly the consequences for perception. As will become evident below, the role of the type of masking and of individual acoustic-phonetic cues on intelligibility is still poorly understood and presents a crucial step toward better understanding of the nature of clear speech and toward applying this knowledge in assistive listening devices.

²Here, we are providing a summary of methods used most typically across studies rather than giving a comprehensive survey of all methods. This should be kept in mind when comparing results across studies.

³We could not include all the details about methods and measurements for each study discussed in order to keep this review manageable. Readers are referred to individual papers for more details.

The clear speech intelligibility benefit has been investigated for various materials including meaningful sentences (Gagne et al. 1995; Schum 1996; Bradlow and Bent 2002; Ferguson and Kewley-Port 2002; Bradlow et al. 2003; Bradlow and Alexander 2007), nonsense sentences (Picheny et al. 1985; Payton et al. 1994; Uchanski et al. 1996; Krause 2001; Krause and Braida 2002; Smiljanić and Bradlow 2005), words (Gagne et al. 1994; Uchanski et al. 1996) and syllables/vowels (Ferguson and Kewley-Port 2002; Gagne et al. 2002). The use of these varied materials has been well-justified depending on the goals of the studies. Semantically anomalous sentences are used in order to minimize the effect of context on word and sentence production and recognition. Similar logic pertains to the use of the word materials. The use of syllables in clear speech research is in part guided by the desire to limit the number of potentially important clear speech acoustic effects (which increase with longer stretches of speech) and in this way to more directly assess the effects of a smaller number of identifiable speaking style changes on intelligibility. A careful consideration and a development of new sets of materials have also been determined by the listener group (e.g., non-native listeners or children) and by the specific acoustic-phonetic measurements performed (Smiljanić and Bradlow 2005, 2008a; Bradlow and Alexander 2007).

Finally, while it would be impossible to include a discussion of all the measurements taken in clear speech studies, we can identify two broad categories of measurement: global measurements and segmental measurements. The first includes speaking rate, pause frequency and duration, fundamental frequency average and range, long-term spectra (distribution of spectral energy over the course of the utterance) and temporal envelope modulations (intensity envelope spectra). The second include vowel formant changes [vowel steady state and/or transitions of first (F1) and second (F2) formant frequencies], vowel space (area between vowel categories as defined by $F1 \times F2$ dimensions), segment duration, consonant-vowel ratio, voice onset time (VOT), short-term spectra (spectrum of the signal around a particular point in time), sound insertion, stop consonant burst elimination, etc. (Picheny et al. 1986, 1989; Matthies et al. 2001; Perkell et al. 2002; Ferguson and Kewley-Port 2002, ²⁰⁰⁷; Bradlow et al. 2003; Krause and Braida 2004; Liu et al. 2004; Smiljanić and Bradlow 2005, 2007).⁴ Next, we move to discuss some of the clear speech findings.

WHO BENEFITS FROM CLEAR SPEECH?

Numerous studies on English clear speech have established that clear speech enhances intelligibility for various listener populations. In a seminal article, Picheny et al. (1985) provided evidence of a significant clear speech benefit for five hearing-impaired subjects listening to nonsense sentences. The average clear speech intelligibility increase was 17 percentage points independent of talker, listener, presentation level and frequency-gain characteristics. Payton et al. (1994) extended these results to conditions of listening in noise, reverberation and combined noise and reverberation. Like Picheny et al. (1985), they found that clear speech was more intelligible for both normal-hearing and hearing-impaired listeners in every environment (average clear speech gain was 20 and 26 percentage points for normal-hearing and hearing-impaired listeners, respectively). Since then, numerous studies have replicated these results and expanded them to other listener populations: adults with normal or impaired hearing (Uchanski et al. 1996; Ferguson and Kewley-Port 2002; Krause and Braida 2002; Ferguson 2004; Liu et al. 2004; Smiljanić and Bradlow 2005; Maniwa et al. 2008), elderly adults (Schum 1996; Helfer 1998), native and non-native listeners (Bradlow and Bent 2002; Bradlow and Alexander 2007; Smiljanić and Bradlow 2007, submitted) and children with and without learning impairments (Bradlow et al. 2003). A clear speech benefit was also found for audio-only and audio-visual modalities for both younger and older listeners (Gagne

⁴Some of the same measurements may be done differently in different studies. It is important to keep this in mind when comparing results across studies. Readers are referred to the cited studies for more details.

et al. 1994, 1995, 2002; Helfer 1997, 1998). Although the magnitude of the clear speech intelligibility advantage varied across talkers, listener groups, presentation levels, frequency-gain characteristics and materials, the cumulative findings of these studies showed that the clear speech effect is very reliable and robust (up to 26 percentage points in Payton et al. 1994), demonstrating that clear speech can be used to improve performance for many listener groups in various communicative situations.

A notable exception to the above studies is the data in Ferguson and Kewley-Port (2002). In their study, conversational and clear vowels were recorded by one male talker in a consonant-vowel-consonant context and mixed with 12-talker babble. The stimuli were presented to young normal hearing adults and elderly hearing impaired listeners. While they found that listeners with normal hearing enjoyed a 15 percentage point clear speech advantage for vowels presented in noise, elderly listeners with hearing impairment listening to the vowels in noise without amplification showed no clear speech benefit. As hearing-impaired listeners could not take advantage of the clear speech strategies produced by the talker in the study while normal hearing subjects benefited substantially, these results revealed that hearing loss affects the way listeners utilize and process available acoustic cues to identify vowels. These findings showed that the relationship between clear speech enhancement strategies and intelligibility improvement is not straightforward and that, furthermore, some articulatory-acoustic enhancements have different consequences (possibly including detrimental) for listener groups with different speech perception deficits. Consistent with these results, Maniwa et al. (2008) have recently shown no clear speech benefit for non-sibilant fricatives by listeners with simulated sloping hearing loss. A closer examination of how clear speech modifications interact with various types of perceptual deficits in different listener populations is an essential line of current and future research.⁵

WHAT MAKES CLEAR SPEECH CLEAR?

Along with investigating the effects of speaking style on intelligibility, clear speech research has strived to identify the acoustic-phonetic features that characterize clear speech productions and contribute to its high intelligibility. The accumulated results of acoustic analyses show that naturally produced clear speech typically involves a wide range of acoustic/articulatory adjustments, including a decrease in speaking rate (longer segments as well as longer and more frequent pauses), wider dynamic pitch range, greater sound-pressure levels, more salient stop releases, greater rms intensity of the non-silent portions (i.e. release burst, frication, and/or aspiration) of obstruent consonants, increased energy in 1000–3000 Hz range of long-term spectra and higher-voice intensity (Picheny et al. 1986, 1989; Matthies et al. 2001; Perkell et al. 2002; Ferguson and Kewley-Port 2002; Bradlow et al. 2003; Krause and Braida 2004; Liu et al. 2004; Smiljanić and Bradlow 2005). One feature that has consistently been shown to accompany naturally produced English clear speech is vowel space expansion (Picheny et al. 1986; Moon and Lindblom 1994; Ferguson and Kewley-Port 2002, 2007; Bradlow et al. 2003; Krause and Braida 2004). Vowel hyper-articulation, associated with clear speech, increases $F1 \times F2$ distance among vowel categories and produces less target undershoot, that is, expected formant targets are ‘better’ approximated, thus, making vowel categories more distinct and presumably less perceptually confusable. Alternatively, clear speech vowel space expansion can be viewed as re-targeting of articulatory gestures (cf. Moon and Lindblom 1994). Vowel space expansion has also been associated with an intelligibility advantage on the basis of inter-talker differences in overall conversational intelligibility. Talkers who are naturally more intelligible tend to produce more expanded vowel spaces (Byrd 1994; Bond and Moore 1994; Moon and Lindblom 1994; Bradlow et al. 1996; Hazan and Markham 2004).

⁵Other listener populations (non-native and clinical) are discussed below.

Moon and Lindblom (1994) examined formant frequencies of English front vowels in the context of /w_l/in conversational and clear speech. They found smaller target undershoot effects and more rapid formant transitions, suggesting faster articulatory gestures, in clear speech compared to conversational speech. This effect was independent of speaking rate differences in the two speaking styles. They argued that clear speech tokens may involve an active reorganization of phonetic gestures to overcome coarticulatory and undershoot effects, presumably due to the output (listener-) oriented nature of clear speech. In contrast, Krause and Braida (2004) found no significant vowel space expansion for vowels produced in clear speech at conversational/normal speaking rates. The vowel space for clear speech at slow speaking rates was expanded relative to conversation speech, but the effect was not found for vowels that were shorter due to the normal/conversational speaking rate. Sentence-in-noise listening tests demonstrated that talkers were able to produce clear speech at a variety of speaking rates as evident in the clear speech intelligibility advantage for both slow and normal/conversational clear speech over conversational speech. Two global level acoustic properties that characterized clear speech produced at normal/conversational speaking rates that were likely linked to the increase in intelligibility were identified: increased energy in the 1000–3000 Hz range of long-term spectra and increased modulation depth of low frequency changes in the intensity envelope. Krause and Braida (2004) concluded that the changes in steady-state vowel formant locations were not likely to be contributing to high intelligibility of clear speech at normal/conversational speaking rates. Their results suggested that vowel space expansion is not a necessary clear speech feature but rather that it may be a concomitant result of slowing down and allowing talkers more time to reach more extreme vowel targets. However, the somewhat smaller intelligibility gain of clear speech at normal/conversational speaking rate compared to clear speech at slow speaking rate (14% vs. 18%) suggests that expanded vowel space may contribute to intelligibility.⁶ It is, therefore, possible that acoustic strengthening at the segmental level (having more ‘extreme’ formant values) can be beneficial to listeners through making the underlying (intended) phoneme categories more transparent in the acoustic signal. Although it may not be crucial and although a direct link between an expanded vowel space (each acoustic cue that characterizes it) and an intelligibility increase is not fully established, vowel hyperarticulation does seem to be an enhancement strategy available to talkers (under optimal production conditions) that may contribute to the overall magnitude of the clear speech intelligibility benefit.

Another finding that consistently occurs across clear speech studies with multiple talkers has been large variability in the implementation of clear speech strategies and in the resulting clear speech benefit (Gagne et al. 1994; Schum 1996; Perkell et al. 2002; Ferguson 2004; Krause and Braida 2004; Ferguson and Kewley-Port 2007). In an articulatory study exploring the interplay between the principles of economy of effort and clarity, Perkell et al. (2002) found that three out of seven talkers produced clear speech with articulator movements that had larger distances and durations compared to conversational speech. These talkers, furthermore, used increased effort (peak articulator speed) in the clear condition, suggesting that clear speech is indeed produced with greater effort than conversational speech. In contrast, three other talkers used only vowel duration increase in their clear speech while one talker mainly used intensity (loudness) increase without indications of increased articulatory effort (peak speed). Similar variability has been found for the magnitude of vowel space expansion in clear speech. Smiljanić and Bradlow (2005), for instance, observed that individual talkers employed varied articulatory strategies as indicated in the vowel formant movements to achieve vowel space expansion in clear speech (e.g., large/a/lowering vs. no/a/lowering; u/backing vs. no/u/

⁶Similar to the speaking rate independent target undershoot results of Moon and Lindblom (1994), Munson and Solomon (2004) reported that lexically ‘hard’ words (less frequent words with more phonological neighbors) can be shorter and still have more expanded vowel space compared to ‘easy’ words (more frequent with fewer neighbors), suggesting that slowing down cannot be the only determinant of articulatory targets.

backing) in both English and Croatian. There was also variability in the intelligibility gain that these talkers provided. A limited number of talkers prevented the authors to explore whether there was a direct connection between these different articulatory strategies and intelligibility benefit.

In a more detailed exploration of talker variability, Ferguson (2004) collected conversational and clear speech from 41 talkers, thus, creating a large database that allows assessment of the relationship between acoustic-phonetic variation across talkers and variability in intelligibility. Ferguson (2004) showed that for 41 talkers, clear speech vowel intelligibility for normal hearing listeners in noise varied widely from -12 to 33 percentage points. The size of the clear speech benefit was unrelated to talker age or experience in interacting with hearing impaired listeners. Female talkers tended to produce more intelligible clear speech compared to male talkers despite having similar conversational intelligibility scores. Similar variability in clear speech intelligibility across talkers was found for words and sentences as well (Gagne et al. 1994; Schum 1996), suggesting that intelligibility variability is characteristic of individual talker's productions rather than of vowel intelligibility only.

Recently, Ferguson and Kewley-Port (2007) investigated acoustic changes for vowels in clear speech for 6 'big benefit' and 6 'no benefit' talkers from their database of 41 talkers, that is, they focused their acoustic analyses on talkers who provided the largest and smallest clear speech intelligibility benefit for normal hearing listeners (as reported in Ferguson 2004). Two clear speech features, vowel space expansion and vowel lengthening, were associated with increased intelligibility. While both groups of talkers produced longer and more extreme vowels and raised F2 for front vowels, the increases for the 'big benefit' group were significantly greater compared to 'no benefit' group. Furthermore, the 'big benefit' talkers expanded their vowel space along the F1 dimension, a strategy that was absent in the 'no benefit' group. Interestingly, both groups showed an equivalent increase in vowel formant movement (dynamic features) in clear speech compared with conversational speech. Based on the acoustic comparisons between the 'big benefit' and 'no benefit' groups, it appears that vowel lengthening and vowel expansion along both F1 and F2 dimensions contributed significantly to increased intelligibility, while enhanced vowel formant dynamic features did not. The difference in the acoustic characteristics of vowels produced by the two talker groups also suggested that a certain 'threshold' of change (be it duration or spectral) needs to be achieved for it to affect intelligibility.

A direct relationship between some of the acoustic-phonetic features examined and intelligibility is, however, still rather tenuous given that in both 'big benefit' and 'no benefit' groups, there were some talkers who exhibited features characterizing the 'opposite' group (e.g., small vowel space expansion in a 'big benefit' group and large vowel lengthening in 'no benefit group'). It remains to be determined to what extent the observed acoustic changes extend to other talkers and whether the presence/absence of these acoustic cues can be correlated with increased/lowered intelligibility. A large talker database, such as Ferguson (2004), allows for the relationship between clear (and conversational) speech features and intelligibility to be examined more closely. Another important contribution of such a database is its potential to be used with various listening populations and to look at correlations of various acoustic-phonetic modifications and intelligibility benefit across these groups. In general, these studies reveal that the exact articulatory-acoustic cues that contribute to the clear speech advantage remain rather elusive. Research focus, thus, remains, on finding the salient acoustic-phonetic clear speech features and establishing their impact on intelligibility.

SPEECH CUE ENHANCEMENT PROCEDURES

As indicated in the discussion of vowel space characteristics, a direct relationship between acoustic-phonetic variation due to changes in speaking style and variability in intelligibility

has been difficult to establish. Although numerous acoustic-phonetic features of conversational-to-clear style transformations have been identified, it is not well-understood yet if and how each of these modifications affects intelligibility. As an alternative to looking at naturally produced clear speech, which involves modifications along various articulatory/acoustic dimensions, all of which can affect intelligibility, several studies used signal-processing to assess the role of individual acoustic cues on intelligibility (Picheny et al. 1989; Stollman et al. 1994; Uchanski et al. 1996). One acoustic cue that typically accompanies the conversational-to-clear speaking style change, namely, a speaking rate decrease, has been investigated through signal-processing in several studies.⁷

Slowing down in clear speech involves both longer segments and longer and more frequent pauses. In studies of inter-talker variability in English conversational speech intelligibility, however, overall speaking rate showed either no correlation with intelligibility (Bradlow et al. 1996) or correlation with intelligibility for some but not all talkers (Hazan and Markham 2004). Uniform or non-uniform time-scaling modifications of the speech signal (by speeding up clear speech and slowing down conversational speech) have resulted in lowered intelligibility compared to non-processed speech for hearing impaired, elderly, normal hearing children and adults confirming that clear speech involves acoustic changes, other than speaking rate, that contribute to its high intelligibility. Furthermore, these studies showed that slowing down due to the changes in speaking style cannot be achieved by a simple segmental lengthening algorithm. Uchanski et al. (1996) also examined the role of pauses in highly intelligible speech. They showed that introducing additional pauses into conversational sentences and excising existing pauses from clear sentences reduced intelligibility scores. Furthermore, key words excised from clear and conversational sentences were as intelligible as when presented in sentence context. Changes in pause structure, therefore, did not necessarily increase key word intelligibility.

Recently, Liu and Zeng (2006) attempted to identify the importance of temporal information in clear speech perception. They employed time-scaling algorithms to digitally compress (clear sentences) or stretch (conversational sentences) the experimental stimuli. They also inserted silent intervals into the conversational sentences to cause the overall duration to be equal to that of the clear speech. Finally, they investigated a relative contribution of temporal envelope (slow frequency modulations) and fine structure (faster frequency modulations) cues for clear speech perception. To do this, the authors extracted the clear and conversational speech signals of both their temporal envelopes and their fine structure. They then created synthesized speech by matching the temporal envelope of one of the speaking styles with the opposite fine structure.

Consistent with previous studies, they found that clear speech was reliably more advantageous over conversational speech for normal-hearing listeners. In contrast to previous studies, they found that gap-inserted conversational speech (increasing the number of pauses) improved intelligibility scores over the original conversational speech, although not at all SNRs, suggesting that speaking rate does contribute to the clear speech advantage. Perceptual results further showed that the temporal envelope contributed more to the clear speech advantage at high SNRs (it is critical for speech recognition in quiet), while the temporal fine structure contributed more at low SNRs (it is critical for speech recognition in noise). Importantly, all processed speech stimuli produced lower intelligibility than the original speech (with the exception of gap-inserted conversational speech having higher intelligibility than the original conversational speech). The detrimental effect on intelligibility points to the introduction of digital signal-processing artifacts into processed speech and, as the authors caution, the need

⁷It is important to keep in mind that artificial rate manipulations may not be the same as natural rate variation.

to consider these as a confounding factor when assessing the role of speaking rate (and of any other acoustic cues) in clear speech perception.

In earlier studies, Krause (2001) and Krause and Braida (2002) assessed the role of speaking rate on intelligibility by eliciting clear speech at normal (i.e., conversational) speaking rates naturally, thereby avoiding some of the pitfalls of signal-processing techniques. They demonstrated that talkers indeed can produce clear speech at normal rates with training. Clear speech at normal/conversational speaking rates increased intelligibility for normal-hearing listeners with simulated hearing losses (in noise), albeit to a slightly smaller degree compared to clear speech at slow rates. These results suggested that speaking slowly is not entirely responsible for the clear speech intelligibility advantage, but rather that clear speech has some inherent acoustic properties independent of speaking rate (Krause and Braida 2004). However, a slight decrease in intelligibility advantage of clear speech produced at conversational rates, along with findings that an increase in the number of pauses has a beneficial effect on intelligibility (Liu and Zeng 2006), suggests that while not crucial, a decrease in speaking rate may be a contributing factor to the intelligibility of clear speech. Precisely in what way the changes in the temporal structure of the speech signal would help language processing is not clear. This remains an open question. Combined, all of these studies clearly indicate that a more detailed knowledge of the principles that guide clear speech production is necessary before the full potential of speech processing systems to improve speech perception can be achieved.

Clear Speech Production and Perception Across Languages

A substantial body of previous work, reviewed so far, has provided us with important insights into the characteristics and benefits of clear speech in English. However, production and perception of clear speech in any language other than English has received considerably less attention. It seems rather obvious that various conversational-to-clear speech modifications would be common to talkers of different languages. For instance, speaking more slowly, more loudly and in a more ‘exaggerated’ way probably characterizes speech of most talkers when attempting to make themselves more intelligible regardless of which language they are speaking. An interesting question that cross-language examination allows us to address is whether clear speech modifications are also driven by language-specific patterns of hyperarticulation. If so, clear speech articulatory adjustments could enlarge the acoustic ‘distance’ between the language-specific contrastive categories reflecting phonological and prosodic properties that vary across languages. A systematic investigation of cross-linguistic clear speech patterns can provide us with invaluable insight into the role various acoustic cues play in expressing and enhancing phonological contrasts and into the interaction between general auditory-perceptual and structural factors that play a role in increased intelligibility of clear speech.

Studies focusing on English have provided us with indirect evidence that phonological structure may affect clear speech production. For example, Uchanski (1988, 1992) found that tense vowels were lengthened more than lax vowels, which increased the duration contrast between tense and lax vowels in English clear speech. Ferguson and Kewley-Port (2002) found that speaking style affected F2 frequency changes differently for front and back vowels with front vowels having higher F2 and back vowels having lower F2 in clear speech compared to conversational speech. This asymmetrical effect of speaking style on F2 changes enhanced the spectral distinction between front and back vowels. At the prosodic/suprasegmental level, Cutler and Butterfield (1990) found that temporal/metrical structure of English was made more salient in clear speech. In contexts in which a preboundary syllable occurred before a word-initial weak syllable, preboundary lengthening was exaggerated in clear speech. These are precisely the contexts in which the language-specific stress group based parsing (i.e., the

expected trochaic foot) would fail (Cutler et al. 1986; Cutler and Otake 1994). The exaggerated word-final lengthening presumably aids the listener in positing the correct word boundary even though the word-initial weak syllable could otherwise be mistakenly parsed as a word-final syllable. These findings demonstrated that the dimensions of phonological contrasts at the segmental and suprasegmental levels shape hyperarticulation strategies implemented by talkers. Thus, it appears that clear speech is in part guided by a contrast enhancement principle that maximizes the distance between language-specific contrasting sound categories and makes prosodic structure more salient. Furthermore, these results suggested that the acoustic-phonetic features that accompany conversational-to-clear speech adjustments may vary across languages reflecting their underlying phonological structure.

Clear speech research in languages other than English has been sparse. In a series of studies, Gagne and colleagues (1994, 1995, 2002) investigated the intelligibility of CV and VCV syllables in conversational and clear speech in auditory, visual and audiovisual modalities in Canadian French. Their results showed an overall clear speech intelligibility advantage over conversational speech in all three modalities. While the main goal of their study was to assess the contribution of different communication channels on the clear speech benefit, they also demonstrated that speaking clearly was a beneficial strategy that increased intelligibility in a language other than English. They did not, however, investigate acoustic-phonetic changes that characterized clear speech transformations in Canadian French. In a cross-language study of English and Spanish, Bradlow (2002) examined vowel production and CV coarticulation in clear and conversational speech. A direct comparison of the two languages that differ in their vowel inventories (large in English and small in Spanish) showed similar amounts of vowel space expansion as well as the maintenance of coarticulatory patterns for high/i/and/u/vowels in clear speech in both languages. There were no accompanying intelligibility data that would show whether these clear speech enhancements produced intelligibility benefit and whether the intelligibility benefit was of similar magnitude in the two languages.

Smiljanić and Bradlow (2005, 2008a) conducted an in-depth comparison of clear speech production and perception in English and Croatian, two languages with structural differences between their phonologies (e.g., 10+ vowels in English and 5 in Croatian). Sentence-in-noise test results showed that spontaneously produced clear speech enhanced intelligibility equally for both English and Croatian listeners in their respective native languages. Acoustic analyses revealed that talkers of both languages enhanced the overall salience of the speech signal through a decrease in speaking rate and expansion of pitch range. Further comparisons showed that talkers expanded the $F1 \times F2$ vowel space in clear speech equally in both languages. In this way, the distances between the contrastive vowel categories were increased in both languages regardless of the sizes of their vowel inventories. The inventory-independent vowel space expansion in Spanish (Bradlow 2002) and Croatian and English (Smiljanić and Bradlow 2005) suggests that talkers hyperarticulate even when segments are unlikely to be perceptually confusable (there are few vowel categories in Croatian and Spanish, and they are fairly distinct). This further suggests that when talkers realize that the listener is having a perceptual difficulty, they hyperarticulate globally, that is, they enhance many aspects of their productions without regard for economy of effort (Lindblom 1990).

Smiljanić and Bradlow (2008a) further investigated the effect of language-specific phonological features on hyperarticulation patterns. Specifically, they explored vowel and stop consonant contrasts along the acoustic-phonetic dimension of duration across speaking styles in the two languages. The two languages differ in how they use the duration cue. Croatian has a lexical contrast between short and long vowels that primarily differ in duration, while English has tense and lax vowels that are mainly distinguished through qualitative/spectral differences, but that also have a secondary duration contrast. While both English and Croatian have voiceless and voiced stops, they differ in their phonetic implementation of the two phonological

categories: Croatian has pre-voiced and short lag stops while English has short and long lag stops. The coordination of onset of voicing and stop closure release gestures, thus, varies across the two languages. Furthermore, the same phonetic category of short lag is encoded in Croatian as a phonologically voiceless and in English as a phonologically voiced stop. The results of acoustic analyses revealed that the Croatian long versus short vowel duration difference was increased by a larger amount than the English tense versus lax duration distinction (through larger lengthening of long vowels compared with short vowels) reflecting the primacy of the role that duration plays in encoding the contrast in Croatian versus the secondary role of duration in the English tense-lax contrast. Furthermore, in Croatian, voicing of prevoiced stops was lengthened while in English, aspiration of long lag stops was lengthened more in clear speech. Larger asymmetrical vowel lengthening in Croatian, longer VOT lengthening of the voiceless category in English, and greater prevoicing lengthening of the voiced category in Croatian demonstrated that clear speech strategies reflect language-specific patterns of duration cue manipulation in clear speech.

Although the results of absolute measures revealed that the vowel length and voicing contrasts were enhanced through asymmetric lengthening of the two members of the contrastive category in a language-specific manner, the relative lengthening patterns indicated stability rather than enhancement of these segmental duration contrasts. The ratio of long to short vowel duration in Croatian and tense to lax vowel duration in English was maintained across speaking styles. Similarly, the ratio of voicing/aspiration to the whole stop duration was stable in both English and Croatian across speaking styles. Thus, proportional duration distance between the contrastive categories in both languages was found to be remarkably stable across the two speaking styles, further suggesting that listeners may judge whether a segment belongs to the 'short' or the 'long' category in relation to the duration of the surrounding segments. A number of studies that examined speaking rate effects on temporal dimensions of speech (VOT, short/long vowels, single/geminate stops) also found a relational invariance in the production of these contrasts across speaking rates (Kessinger and Blumstein 1998; Boucher 2002; Hirata 2004; Hirata and Whiton 2005; but see also de Jong 2001 for lack of relational invariance). Some of these studies (e.g., Boucher 2002) argue for the perceptual invariance as well as production invariance while others (Miller and Volaitis 1989; Volaitis and Miller 1992 and recently Nagao and de Jong 2007) show that perceptual rate normalization for voicing contrasts is not relationally invariant but that the perceptual boundaries shift with changes in speaking rate. Importantly, all of these studies show that listeners take into account fairly local timing relations when judging category affiliation that is also hypothesized for perception of clear speech vowel and stop duration contrasts. It is important to note that fast-to-slow speaking rate changes, and conversational-to-clear speech modifications may not involve the same articulatory scaling mechanisms and the similarities and differences between the results of the various studies should, therefore, be interpreted with caution. More research toward understanding the mechanism that guides hyperarticulation in speech production (as well as language-specific patterns of contrast enhancement) and its implications for speech perception is needed. Moreover, cross-linguistic clear speech research has important implications for developing reliable models of linguistic behavior.

Clear Speech Production and Perception in Second Language

Research reported in the studies reviewed so far has focused mainly on understanding how speech communication is challenged by listener- or signal-related factors, such as hearing loss or introduction of noise or reverberation in the speech signal. Non-native listeners differ in many respects from the listener populations that have traditionally been studied in clear speech research. While native listeners may experience a problem in accessing the speech signal due to hearing loss or added noise, non-native listeners' problems arise, in part, due to the lack of extensive experience with the sound structure (e.g., phonological contrasts and language-

specific phonetic implementation of those contrasts) and other higher levels of linguistic structure (e.g., lexicon, syntax, and semantics) of the target language. Non-native listeners, thus, have qualitatively different problems in processing the speech signal from the native listeners who are experienced in accessing the linguistic code of the target language. Given their inexperience with the sound structure of the target language, non-native listeners should exhibit a smaller clear speech advantage compared to native listeners, as the clear speech strategies, in part, enhance language-specific phonological contrasts that may differ across their first (L1) and their second (L2) languages.

Bradlow and Bent (2002) investigated whether naturally produced clear speech is an effective intelligibility-enhancement strategy for non-native listeners. They tested the hypothesis that clear speech production is native-listener oriented. In the study, non-native and native listeners were presented with meaningful English sentences in conversational and clear speaking style at two signal-to-noise ratios (-4 and 8 dB). They found a large clear speech benefit for native listeners and a substantially smaller benefit for non-native listeners. They also found that non-native listeners were more adversely affected by increased noise than native listeners. The smaller clear speech benefit for non-native listeners compared to native listeners was taken to reflect the nature of the non-native listeners' speech perception deficit, that is, their impeded access to the linguistic code of the target language. This further suggested that clear speech is native listener oriented and that additional exposure to and experience in processing the target language could lead to a larger clear speech benefit for non-native listeners.

Probing the nature of the non-native perceptual difficulties beyond their inability to take full advantage of language-specific clear speech enhancement strategies, Bradlow and Alexander (2007) explored the interaction of semantic and acoustic-phonetic cues on intelligibility for native and non-native listeners. They manipulated the availability of semantic context (high vs. low predictability sentences) and of the acoustic-phonetic cues (conversational vs. clear speech). In this way, they tested whether word recognition accuracy was facilitated by the availability of the contextual semantic cues when sufficient signal clarity was available to both native and non-native listeners.

Sentence-in-noise test results revealed a different pattern in taking advantage of both contextual and acoustic-phonetic cues for native and non-native listeners. Native listeners benefited from both sources of enhancements (semantic and acoustic-phonetic) separately and combined. In contrast, non-native listener final word recognition improved only for high predictability sentences that were produced in clear speech, that is, when both enhancements were present. Non-native listeners required more signal clarity (acoustic-phonetic information) in order to take advantage of the contextual information. Their deficit seems to come, in part, from the less effective use of the higher-level information due to information loss at the phoneme identification level. These results suggest that the disproportionate non-native speech-in-noise perceptual difficulties arise from an accumulated effect of difficulties at lower level of processing with limitations at higher levels, rather than from the disproportioned effect of noise alone at the level of phoneme identification when contextual-semantic effects are not present (similar implications, although not for clear speech, were reported in Mayo et al. 1997 and Cutler et al. 2004, among others).

Recently, Smiljanić and Bradlow (submitted) examined how talker and listener native language background and speaking style influence speech communication. In a series of experiments, they explored whether native and non-native clear speech strategies provide similar intelligibility benefits for native and highly proficient non-native listener groups matched and mis-matched with talkers for their native language background. Specifically, they looked at the intelligibility of conversational and clear speech produced in English by American English (AE) and by relatively high-proficiency Croatian talkers for AE and high-proficiency Croatian

listeners. The sentence-in-noise listening tests revealed that all listeners benefited from clear speech. Furthermore, native AE clear speech strategies were equally beneficial for these Croatian non-native listeners and for the native AE listeners. The results also showed that both non-native and native listeners used in these experiments preferred native speech over non-native speech (even though non-native talkers and listeners shared the same L1 background). Interestingly, some non-native talkers produced clear speech that was as beneficial to the native listeners as that of native talkers revealing a complex interaction of fluency and inherent talker clarity on intelligibility which is still rather poorly understood.

Combined, these results demonstrated that naturally produced clear speech is a beneficial strategy that enhances intelligibility for both native and non-native listeners. Unlike less proficient listeners in Bradlow and Bent (2002) study, these highly proficient non-native listeners were able to take advantage of the more salient acoustic enhancement cues provided by native clear speech to process L2, demonstrating that increased experience in L2 leads to an increase in clear speech benefit. Compared to native clear speech, non-native clear speech was shown to increase intelligibility less for both native and non-native listeners even when they shared L1 background. This further supported the notion advanced in the cross-language studies of clear speech (Bradlow 1995; Smiljanić and Bradlow 2005, 2008a) as well as in the studies of vowel space expansion (Picheny et al. 1986; Moon and Lindblom 1994; Ferguson and Kewley-Port 2002, 2007; Krause and Braida 2004) that clear speech modifications involve enhancement of phonological contrasts in a language-specific way. For non-native talkers, then, the inability to enhance the phonological contrast of the target language in a 'language-appropriate' way would result in a smaller intelligibility gain when compared to native clear speech.

Such detailed explorations of non-native talkers' conversational and clear speech are essential for understanding how non-native speech differs from the native speech and in what way non-native clear speech modifications reflect inexperience in using L2 and/or the 'transfer' of clear speech strategies from their native L1. Detailed comparisons of L1 and L2 clear speech productions across different language pairs will provide us with an insight into L1 and L2 interaction in proficient L2 talkers as well as into the interaction of talker- and listener-related factors in speech intelligibility. How various levels of proficiency and accentedness and type and amount of L2 exposure affect clear speech production and intelligibility is a direction of future research. Understanding the effects of non-native talker- and listener-related factors on intelligibility has important implications for classroom interactions and for other everyday communicative situations involving non-native populations.

Clear Speech Production and Perception Beyond Syllables and Short Sentences

In more natural communicative situations, that is, outside of laboratory, talkers produce speech that is more complex and involves producing utterances that are longer than syllables, words or short sentences. Spontaneously produced clear speech may, therefore, show different patterns and degrees of articulatory-acoustic adjustments compared to clear speech produced in laboratory conditions. Some of the changes may involve modifications of suprasegmental and prosodic features that have not been the focus of most laboratory clear speech studies. Addressing some of these concerns, Smiljanić and Bradlow (2008b) investigated production and perception of longer and syntactically more complex paragraphs. The first goal of the study was to assess the ability of talkers to maintain clear speech modifications across longer and more complex stretches of speech and to see whether intelligibility varies across the duration of the paragraphs. The second goal was to explore how much clear speech modifications involve changes above the level of segment, that is, how phrasing and the durations of successive vowels and consonant intervals as well as their variability (i.e., durational properties

of successive segments in relation to each other rather than of individual vowels or VOTs) are affected by changes in speaking style.

The results of sentence-in-noise listening tests showed a consistent intelligibility advantage for the paragraphs produced in clear speech compared with conversational speech. Listeners were on average better in word recognition in clear speech paragraphs compared to conversational speech paragraphs. The data, thus, showed strong evidence that talkers successfully maintained clear speech modifications across longer stretches of speech. In order to gain a better understanding of intelligibility levels across paragraphs, the authors examined intelligibility scores for successive paragraph portions. Paragraph portion-by-portion results revealed that intelligibility levels were not uniform across paragraph-length utterances in either speaking style, that is, intelligibility scores increased and decreased during the duration of the paragraphs in a parallel manner for clear and conversational speaking styles. Clear speech intelligibility, however, was consistently higher throughout the paragraphs compared to conversational speech intelligibility. Variability in intelligibility scores across paragraph portions in both speaking styles suggests that context and utterance-specific characteristics, such as introduction of new topics, lexical characteristics and word predictability along with speech clarity, affected intelligibility significantly.

The results of acoustic analyses revealed that in addition to increased durations of segments, clear speech was characterized by insertion of short segments that were dropped or coarticulated with surrounding sounds in conversational speech and increased phrasing, that is, increase in the number of pauses and prosodic phrases. This work replicated and extended previous findings (Picheny et al. 1986; Krause and Braida 2002, 2004; Bradlow et al. 2003; Smiljanić and Bradlow 2005, 2008a) by looking at durational and prosodic characteristics of utterances of various lengths and complexity in clear speech. In addition to increasing the accuracy of phoneme identification, possibly through greater articulatory precision in longer segments, these modifications conspire to make syllable and word boundaries and their internal structure more salient, presumably facilitating segmentation, lexical access and word recognition.

The authors also examined duration variability in successive consonantal and vocalic intervals across utterances. They found that after normalizing for changes in speaking rates (coefficients of variability were calculated as standard deviations of the consonantal and vocalic intervals divided by the average durations of the same consonantal and vocalic intervals for each utterance), the extent of variability of vocalic and consonantal intervals remained unchanged across speaking styles. These results suggest that talkers adhere to the locked-in local timing relations when producing neighboring segments at any given speaking rate. Increased intelligibility of clear speech, therefore, may be attributed to prosodic structure enhancement (increased phrasing and enhanced segmentability) and stable global temporal properties. Finally, such stable global temporal properties may aid the listeners in interpreting the incoming speech signal across speaking styles/rates and in deriving accurate information about the intended message (lexical access, segment identification, prosodic structure, etc.).

This line of research points to various unresolved issues to be addressed in future work. It is important to connect laboratory-based research on variability in speech production and intelligibility to more naturalistic communicative settings where speech typically involves longer utterances with highly variable phrase types. Eliciting spontaneous conversational and clear speech will provide us with more data that varies along the continuum from least formal spontaneous to most formal hyperarticulated speech (H&H theory, Lindblom 1990), thus, reflecting complex communicative conditions in which listeners and talkers interact and adapt to each other. This will allow us to gain an understanding of how flexible and adaptable production and perception mechanism are and how they may fail to achieve the most optimal

performance in real-life situations. Moreover, investigating more complex speech provides a window into the factors that contribute to specifying articulatory targets ranging from the need for clarity to other levels of linguistic representation, such as lexicon, prosody, and syntax. Importantly, future work needs to explore to what extent the results from laboratory-based clear speech research extend to spontaneously produced conversational and clear speech.

Clinical Populations and Speech Technology

The focus of early clear speech research has largely been on examining whether clear speech is more intelligible compared to conversational speech, finding acoustic-phonetic features that characterize highly intelligible speech and, ultimately, on understanding whether these features can be incorporated into speech processors for impaired listeners. Most clear speech research, therefore, focused on adult, native English listeners with impaired or with normal hearing and healthy native adult talkers. Numerous acoustic-phonetic features have been identified and a robust clear speech intelligibility benefit has been reported for these listener populations across studies. In this section, we discuss studies that have extended their investigation to other listener and talker populations (with exclusion of non-natives that were described above) and to providing speech-enhancement algorithms to improve quality of speech in various communication channels. Investigating whether the clear speech benefit extends to other (clinical) populations, and whether cue-enhancement algorithms improve quality of communication for these populations, can potentially provide insight into the particular character of their underlying perceptual deficits. This may then suggest possible deficit-specific therapeutic strategies and speech-enhancement procedures.

Children differ from adult listeners in that their exposure to the spoken language is less extensive. Both children's production and perception are qualitatively different from adults' (Lee et al. 1999). Bradlow et al. (2003) explored clear speech perception by children with and without learning disabilities in order to see whether they would be able to utilize clear speech cues to the same extent as the more experienced adults. All children had normal peripheral hearing at the time of testing, thus, assuring that the potential difference in scores between the two groups was not caused by auditory differences. As predicted, in a sentence-in-noise recognition task, all children benefited less from clear speech compared to native adult listeners, presumably due to their less advanced stage of language development. Although children with learning problems exhibited lower intelligibility scores compared to children with no learning deficits, the clear speech effect was substantial for both groups (9% and 16–18%). A notable result from this study was that when presented with clear speech, children with learning problems performed at the same level as children with no learning problems when presented with conversational speech. This study demonstrated that speech perception for children with learning problems may be enhanced in everyday communication by employing a simple strategy of speaking clearly. Speaking clearly can, thus, be used effectively as an intelligibility-enhancing strategy in clinical settings and in educational settings, along with other hearing-assistive technologies and communication strategies, such as physical acoustic modifications of the classroom and hearing aids (Crandell and Smaldino 1999).

Liu et al. (2004) compared a clear speech perceptual benefit for normal hearing adults and adults with cochlear implants who were all post-lingually deafened and varied in age years of implant use. Using a slightly different approach in this study, conversational and clear speech intelligibility was measured as a function of signal-to-noise ratio in order to derive psychometric functions that would characterize the clear speech advantage over a range of signal-to-noise ratios. Speech perception thresholds were then calculated for each speaking style, that is, a signal-to-noise ratio at which a subject can perform at 50% correct level was derived. This measure determines the clear speech benefit independent of variability in conversational speech perception ability at a fixed signal-to-noise ratio. Both groups of

listeners derived a significant clear speech advantage, although listeners with cochlear implants needed somewhat better signal-to-noise ratios in order to perform at the same level as normal hearing adults. The results also indicated that clear speech provided the maximal benefit when conversational speech intelligibility was moderate. When conversational speech intelligibility scores were too high or too low, the benefit of clear speech provided was smaller, suggesting that the degree of clear speech enhancement is, in part, determined by the initial stage, that is, intelligibility level of conversational speech. Finally, the study demonstrated a high degree of individual variability in clear speech perception for the cochlear implant users. Some cochlear-implant users benefited maximally from clear speech modifications while others derived a relatively small clear speech benefit. This result highlights the need to not only determine which intelligibility enhancing strategies are successful for various listener populations, but also for individual listeners.

Looking at yet another listener population whose perception deficits differ from the ones described earlier in this article, Zeng and Liu (2006) compared performance between clear and conversational speech perception in participants with auditory neuropathy (AN). AN is characterized by the disruption in auditory nerve activity as reflected in distorted or absent auditory brainstem responses and with normal or nearly normal cochlear amplification function (for details concerning the nature of the deficit, see Zeng and Liu 2006 and references therein). Individuals with AN exhibit difficulties in temporal processing and speech understanding, particularly in noise. Participants who had been clinically diagnosed with AN listened to short sentences in conversational and clear speaking style produced by one female talker. The listening conditions included noise and quiet and the stimulation modes compared monoaural acoustic, diotic acoustic, electric stimulation via a cochlear implant and binaurally combined acoustic and electric stimulation.

Consistent with previous clear speech studies, the results showed a significant clear speech benefit in both noise and in quiet across all stimulation conditions. Individuals with AN who had cochlear implants performed better than the individuals without implants in both listening conditions, thus, providing evidence for the effectiveness of electric hearing in addressing speech perception difficulties in quiet and in noise for listeners with AN. The ability of participants with AN to take advantage of clear speech enhancements suggested that innovative hearing aids that can convert conversational speech into clear speech could benefit these individuals. These less invasive treatment strategies could be greatly improved by tailoring signal processing schemes to the specific needs of this and other clinical populations. Specifically, for AN individuals, linear amplification, expansion of temporal modulation and shift of low frequency signals to high frequency regions, all reflecting the nature of their specific perceptual deficits, would alleviate some of their perceptual problems. Note that some of these amplifications would employ similar enhancements to naturally produced clear speech. Combined, the results of the studies described in this section so far (as well as the studies looking at non-native speech perception) showed that different listener populations are sensitive to different clear speech features and that our understanding of the specific perceptual difficulties of each listener population will allow us to address their needs in a more meaningful way.

A number of the studies examined so far in this section explored the extent to which clear speech enhancements provide a perceptual advantage for various populations. The focus of these studies has been almost exclusively on non-impaired talkers. Shifting the focus to talkers with speech and perception difficulties reveals what acoustic-phonetic features are present/absent compared with non-impaired clear speech and how the difference affects intelligibility. Clear speech production can be examined as a therapeutic tool that may increase otherwise compromised intelligibility by certain talker groups. Goberman and Elmer (2005) looked at conversational and clear speech production in individuals with Parkinson disease (PD), a

neurological disorder that may affect non-speech movement as well as aspects of speech production (for details concerning the nature of the deficit, see Goberman and Elmer 2005 and references therein). PD speech is characterized as hypokinetic dysarthria with deficits in respiration, phonation and articulation, which results in decreased intelligibility. Speech associated with PD typically involves decreased speaking rate and increased number of pauses, increased mean fundamental frequency (F0) and decreased F0 variability, as well as some degree of vowel reduction. Participants with diagnosed PD were recorded producing syllables, a short passage and a short monologue in conversational and clear speaking styles during a visit to their neurologist.

Acoustic analyses of individual clear speech productions revealed that a majority of talkers produced slower speech without increasing the number of pauses (the results varied somewhat across the materials). The majority of talkers also increased their mean F0 and F0 variability for most target materials. Although the overall vowel space increase in clear speech was not significant, 7 out of 12 participants showed higher values, that is, more distinct vowel categories, in clear speech compared to conversational speech. All of these effects of speaking style on PD productions were in the same direction as the ones produced by non-impaired talkers albeit to a smaller degree. PD participants were, thus, able to partially compensate for the loss of prosody and vowel neutralization when given the direction to speak clearly. It is possible that some PD talkers were already compensating for their less intelligible speech by slowing the articulation rates and increasing the number of pauses in their conversational speech, which would account for their inability to change these articulatory parameters even further in clear speech. Although long-term retention of the intelligibility-enhancing articulatory behavior is unknown, clear speech holds promise as a therapeutic strategy for improving PD speech. Future work should address the benefit that listeners can derive from PD clear speech.

In addition to studying the intelligibility of speech produced by people with speech and perception impairments, studying the attitudes of listeners toward the speech of people with communication impairment is important as well. Successful communication depends in part on the attitudes of communication partners and oftentimes the attitudes towards impaired people are negative, thus, disadvantaging their communication efforts. Hanson et al. (2004) investigated the effectiveness of supplementation strategies used by speakers with dysarthria. Nine individuals with traumatic brain injury who demonstrated moderate to severe dysarthria were instructed to produce short sentences (HINT; Nilsson et al. 1994) in four speaking conditions: habitual speech (corresponding to conversational speech used throughout the present paper), alphabet supplemented speech, topic supplemented speech, and, relevant to this review, clear speech. In alphabet and topic supplanted speech, talkers point to the first letter of each word or to a word representing the topic of the sentence as it is spoken. Speech intelligibility levels were established in a previous study (Beukelman et al. 2002 as reported in Hanson et al. 2004) and showed that all supplementation methods, including clear speech, increased intelligibility compared to habitual/conversational speech. The alphabet supplementation strategy provided the largest intelligibility benefit.

Hanson and colleagues then asked four groups of listeners (general public, speech-language pathologists, allied health professionals and family members) to rate supplemented speech according to the effectiveness and acceptability of the strategy. All three supplementation strategies were rated as more favorable than habitual/conversational speech by all listeners. Although alphabet supplementation was ranked as most effective-acceptable, surprisingly, there was a negative correlation between intelligibility scores and attitude rankings for alphabet supplementation strategy. This counterintuitive dissociation between intelligibility and attitudes towards alphabet supplemented speech suggests that listeners may not understand the benefits realized by certain supplementation strategies and may, therefore, cause talkers to

abandon the strategy. Interestingly, there was a rather consistent positive correlation between intelligibility scores and attitude rankings for clear speech.

Combined, these studies show that careful examination of clear and conversational speech perception and production by impaired listeners and talkers may increase our understanding of the deficits associated with various speech and perception impairments, as well as indicate the need to tailor therapeutic strategies differently for various groups. Going beyond providing us with insights into the nature of these deficits, these studies also emphasize that educating talkers and their communication partners is crucial in fostering consistent use of supplementation strategies and in allowing impaired talkers to continue to use natural speech as a primary means of communication. Clear speech as one such strategy appears to be very effective for various listener populations with perceptual problems and for impaired talkers whose speech is less intelligible. However, larger studies with these and other populations and long-term effects of clear speech use are needed.

Conclusions: Open Research Questions and Future Directions

In this review, we have attempted to bring to light some of the most important as well as most recent clear speech research findings. Throughout the article, we have highlighted research questions and approaches that continue to generate important insights into the interplay between conversational-to-clear speech modifications and intelligibility for various listener groups. We have also indicated open questions and lines of future work. While many significant advances have been made in our understanding of clear speech effects in production and perception, numerous other goals still remain to be achieved. One of the biggest challenges of clear speech continues to be the lack of a firm link between specific articulatory-acoustic modifications and intelligibility. It is, therefore, impossible to provide, in a way of concluding remarks, a comprehensive list of conversational-to-clear speech adjustments that have been shown definitively to increase intelligibility. More work toward finding the salient acoustic-phonetic clear speech features and establishing their impact on intelligibility is still very much at the center of the clear speech research agenda. A better understanding of individual talker differences and of the variability in clear speech strategies, as well as their consequences for listener groups with various perceptual deficits is also needed. Clear speech research, therefore, must be expanded to include more varied talker and listener groups.

Another important research area involves connecting clear speech production and perception with other levels of linguistic structure. While it is crucial to understand which articulatory changes make clear speech more intelligible, it is equally important to understand how the interaction of various levels of linguistic structure and cognitive functioning determine clear speech output and how clear speech changes affect speech processing at different levels, such as sound category formation, speech segmentation, lexical access, and syntactic and prosodic processing. Expanding clear speech research to include language processing at various levels of linguistic structure will help shed light on the underlying mechanisms of speech production and perception that allow talkers to adapt their output, and allow listeners to take advantage of the clear speech adjustments.

It is important to extend clear speech research toward more naturalistic communication settings and to find whether the clear speech insights obtained in the laboratory extend to everyday communication conditions. Finally, it is crucial to further stimulate translational research, that is, to connect the discoveries in the laboratory with the actual development of algorithms, signal-processing schemas and therapeutic strategies that enhance speech communication for listeners with various perception difficulties.

Biographies

Rajka Smiljanić is an Assistant Professor of Linguistics at the University of Texas at Austin. Her work is concentrated in the areas of experimental phonetics, cross-language speech production and perception, clear speech, intelligibility, bilingualism and prosody. She received a bachelor's degree in English and Russian Languages and Literatures at the University of Zagreb, Croatia, and her MA and PhD from the Linguistics Department at the University of Illinois Urbana-Champaign. After receiving her doctorate, she worked as a Research Associate and Lecturer in the Linguistics Department at Northwestern University. She also taught in the Communications Disorders and Sciences Department at Rush University.

Ann Bradlow is a Professor of Linguistics at Northwestern University. Her primary research areas are speech perception, acoustics phonetics and cross-language/second-language phonetics. She received her bachelor's degree in Linguistics from the University of Chicago, and her MA and PhD from the Department of Modern Languages and Linguistics at Cornell University. Prior to joining the faculty in the Northwestern University Linguistics Department, Professor Bradlow worked as a postdoctoral fellow in the Speech Research Laboratory (Psychology Department) at Indiana University, and in the Auditory Neuroscience Laboratory (Communication Sciences & Disorders Department) at Northwestern University.

Works Cited

- Beukelman DR, Fager S, Ullman C, Hanson E, Logemann J. The impact of speech supplementation and clear speech on the intelligibility and speaking rate of people with traumatic brain injury. *Journal of Medical Speech-Language Pathology* 2002;10(4):237–42.
- Bond ZS, Moore TJ. A note on the acoustic-phonetic characteristics of inadvertently clear speech. *Speech Communication* 1994;14:325–37.
- Boucher VJ. Timing relations in speech and the identification of voice-onset times: A stable perceptual boundary for voicing categories across speaking rates. *Perception and Psychophysics* 2002;64(1):121–30. [PubMed: 11916295]
- Bradlow AR. A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America* 1995;97:1916–24. [PubMed: 7699173]
- Bradlow, AR. Confluent talker- and listener-related forces in clear speech production. In: Gussenhoven, C.; Warner, N., editors. *Laboratory phonology*. Vol. 7. Berlin, Germany/New York, NY: Mouton de Gruyter; 2002. p. 241-73.
- Bradlow AR, Alexander J. Semantic-contextual and acoustic-phonetic enhancements for English sentence-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America* 2007;121(4):2339–49. [PubMed: 17471746]
- Bradlow AR, Bent T. The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America* 2002;112:272–84. [PubMed: 12141353]
- Bradlow AR, Kraus N, Hayes E. Speaking clearly for learning-impaired children: sentence perception in noise. *Journal of Speech, Language, and Hearing Research* 2003;46:80–97.
- Bradlow AR, Torretta GM, Pisoni DB. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication* 1996;20:255–72.
- Byrd D. Relations of sex and dialect to reduction. *Speech Communication* 1994;15:39–54.
- Cho, T. *The effects of prosody on articulation in English*. New York, NY: Routledge; 2002.
- Cho T. Prosodic strengthening and featural enhancement: evidence from acoustic and articulatory realizations of /a, i/ in English. *Journal of the Acoustical Society of America* 2005;117(6):3867–78. [PubMed: 16018489]
- Cole J, Kim H, Choi H, Hasegawa-Johnson M. Prosodic effects on acoustic cues to stop voicing and place of articulation: evidence from radio news speech. *Journal of Phonetics* 2007;35(2):180–209.
- Crandell CC, Smaldino JJ. Improving classroom acoustics: utilizing hearing-assistive technology and communication strategies in the educational setting. *The Volta Review* 1999;101(5):47–63.

- Cutler A, Butterfield S. Durational cues to word boundaries in clear speech. *Speech Communication* 1990;9:485–95.
- Cutler A, Otake T. Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language* 1994;33:824–44.
- Cutler A, Mehler J, Norris D, Segui J. The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language* 1986;25:385–400.
- Cutler A, Webber A, Smits R, Cooper N. Patterns of English phoneme confusion by native and non-native listeners. *Journal of the Acoustical Society of America* 2004;116:3668–78. [PubMed: 15658717]
- de Jong K. The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America* 1995;97:491–504. [PubMed: 7860828]
- de Jong KJ. Effects of syllable affiliation and consonant voicing on temporal adjustment in a repetitive speech production task. *Journal of Speech, Language, and Hearing Research* 2001;44:826–40.
- Ferguson SH. Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *Journal of the Acoustical Society of America* 2004;116:2365–73. [PubMed: 15532667]
- Ferguson SH, Kewley-Port D. Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America* 2002;112:259–71. [PubMed: 12141351]
- Ferguson SH, Kewley-Port D. Talker differences in clear and conversational speech: acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research* 2007;50:1241–55.
- Gagne J-P, Masterson V, Munhall KG, Bilida N, Querengesser C. Across talker variability in auditory, visual, and audiovisual speech intelligibility for conversational and clear speech. *The Journal of the Academy of Rehabilitative Audiology* 1994;27:135–58.
- Gagne JP, Querengesser C, Folkeard P, Munhall KG, Masterson V. Auditory, visual, and audiovisual speech intelligibility for sentence-length stimuli: an investigation of conversational and clear speech. *The Volta Review* 1995;97(1):33–51.
- Gagne JP, Rochette AJ, Charest M. Auditory, visual and audiovisual clear speech. *Speech Communication* 2002;37:213–30.
- Goberman AM, Elmer LW. Acoustic analysis of clear versus conversational speech in individuals with Parkinson disease. *Journal of Communication Disorders* 2005;38:215–30. [PubMed: 15748725]
- Hanson EK, Beukelman DR, Fager S, Ullman C. Listener attitudes toward speech supplementation strategies used by speakers with dysarthria. *Journal of Medical Speech-Language Pathology* 2004;12(4):161–6.
- Harnsberger J, Wright R, Pisoni D. A new method for eliciting three speaking styles in the laboratory. *Speech Communication* 2008;50:323–36. [PubMed: 19562041]
- Hazan V, Simpson A. The effect of cue-enhancement on the intelligibility of nonsense word and sentence materials presented in noise. *Speech Communication* 1998;24:211–26.
- Hazan V, Simpson A. The effect of cue-enhancement on consonant intelligibility in noise: speaker and listener effects. *Language and Speech* 2000;43:273–94. [PubMed: 11216296]
- Hazan V, Markham D. Acoustic-phonetic correlates of talker intelligibility for adults and children. *Journal of the Acoustical Society of America* 2004;116:3108–18. [PubMed: 15603156]
- Helfer KS. Auditory and auditory-visual recognition of clear and conversational speech by older adults. *Journal of the American Academy of Audiology* 1997;9(3):234–42. [PubMed: 9644622]
- Helfer KS. Auditory and auditory-visual recognition of clear and conversational speech. *Journal of Speech, Language, and Hearing Research* 1998;40(4):432–43.
- Hirata Y. Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics* 2004;32:565–89.
- Hirata Y, Whiton J. Effects of speaking rate on the single/geminate stop distinction in Japanese. *Journal of the Acoustical Society of America* 2005;118(3 Pt 1):1647–60. [PubMed: 16240824]
- Junqua JC. The Lombard reflex and its role on human listeners and automatic speech recognizers. *Journal of the Acoustical Society of America* 1993;93.1:510–24. [PubMed: 8423266]

- Kessinger RH, Blumstein SE. Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics* 1998;26:117–28.
- Krause JC. Properties of naturally produced clear speech at normal rates and implications for intelligibility enhancement. 2001 Unpublished Doctoral Dissertation.
- Krause JC, Braida LD. Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *Journal of the Acoustical Society of America* 2002;112:2165–72. [PubMed: 12430828]
- Krause JC, Braida LD. Acoustic properties of naturally produced clear speech at normal speaking rates. *Journal of the Acoustical Society of America* 2004;115(1):362–78. [PubMed: 14759028]
- Kuhl PK, Andruski JE, Chistovich L, Chistovich I, Kozhevnikova E, Sundberg U, Lacerda F. Cross language analysis of phonetic units in language addressed to infants. *Science* 1997;227:684–6. [PubMed: 9235890]
- Lee S, Potamianos A, Narayanan S. Acoustics of children's speech: developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America* 1999;105(3):1455–68. [PubMed: 10089598]
- Lindblom, B. Explaining phonetic variation: a sketch of the H&H theory. In: Hardcastle, WJ.; Marchal, A., editors. *Speech production and speech modeling*. Amsterdam, The Netherlands: Kluwer Academic; 1990. p. 403–39.
- Liu S, Zeng FG. Temporal properties in clear speech perception. *Journal of the Acoustical Society of America* 2006;120(1):424–32. [PubMed: 16875238]
- Liu S, Del Rio E, Bradlow AR, Zeng FG. Clear speech perception in acoustic and electrical hearing. *Journal of the Acoustical Society of America* 2004;116:2374–2383. [PubMed: 15532668]
- Maniwa K, Jongman A, Wade T. Perception of clear fricatives by normal-hearing and simulated hearing-impaired listeners. *Journal of the Acoustical Society of America* 2008;123(2):1114–25. [PubMed: 18247912]
- Matthies M, Perrier P, Perkell JS, Zandipour M. Variation in anticipatory coarticulation with changes in clarity and rate. *Journal of Speech, Language, and Hearing Research* 2001;44:340–53.
- Mayo LH, Florentine M, Buus S. Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research* 1997;40:686–93.
- Miller JL, Dexter ER. Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology. Human Perception and Performance* 1988;14(3):369–378. [PubMed: 2971767]
- Miller JL, Green KP, Reeves A. Speaking rate and segments: a look at the relation between speech production and speech perception for the voicing contrast. *Phonetica* 1986;43:106–115.
- Miller JL, Volaitis LE. Effects of speaking rate on the perceived internal structure of phonetic categories. *Perception and Psychophysics* 1989;46:505–12. [PubMed: 2587179]
- Moon SJ, Lindblom B. Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America* 1994;96:40–55.
- Munson B, Solomon NP. The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research* 2004;47:1048–58.
- Nagao K, de Jong KJ. Perceptual rate normalization in naturally produced rate-varied speech. *Journal of the Acoustical Society of America* 2007;120:2882–98. [PubMed: 17550187]
- Nilsson M, Soli S, Sullivan J. Development of hearing in noise test for the management of speech reception thresholds in quiet and noise. *Journal of the Acoustical Society of America* 1994;95:1085–99. [PubMed: 8132902]
- Payton KL, Uchanski RM, Braida LD. Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *Journal of the Acoustical Society of America* 1994;95:1581–92. [PubMed: 8176061]
- Perkell JS, Zandipour M, Matthies ML, Lane H. Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *Journal of the Acoustical Society of America* 2002;112(4):1627–41. [PubMed: 12398468]
- Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research* 1985;28:96–103. [PubMed: 3982003]

- Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing II: acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research* 1986;29:434–46. [PubMed: 3795886]
- Picheny MA, Durlach NI, Braida LD. 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 and Hearing Research* 1989;32:600–3. [PubMed: 2779204]
- Schum DJ. Intelligibility of clear and conversational speech of young and elderly talkers. *Journal of the American Academy of Audiology* 1996;7(3):212–8. [PubMed: 8780994]
- Skowronski MD, Harris JG. Applied principles of clear and Lombard speech for automated intelligibility enhancement in noisy environments. *Speech Communication* 2005;48:549–58.
- Smiljanić R, Bradlow AR. Production and perception of clear speech in Croatian and English. *Journal of the Acoustical Society of America* 2005;118(3 Pt 1):1677–88. [PubMed: 16240826]
- Smiljanić, R.; Bradlow, AR. Clear speech intelligibility: listener and talker effects. *Proceedings of the XVIth International Congress of Phonetic Sciences; Saarbrücken, Germany. 2007.*
- Smiljanić R, Bradlow AR. Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics* 2008a;36(1):91–113.
- Smiljanić R, Bradlow AR. Temporal organization of English clear and plain speech. *Journal of the Acoustical Society of America* 2008b;124(5):3171–82.
- Smiljanić R, Bradlow AR. Non-native clear speech effect I: Intelligibility and accentedness. *Journal of the Acoustical Society of America.* submitted.
- Smith, C. Prosodic accommodation by French speakers to a non-native interlocutor. *Proceedings of the XVIth International Congress of Phonetic Sciences; Saarbrücken, Germany. 2007.*
- Stollman MHP, Kapteyn TS, Sleeswijk BW. Effect of time-scale modification of speech on the speech recognition threshold in noise for hearing-impaired and language-impaired children. *Scandinavian Audiology* 1994;23:39–46. [PubMed: 8184282]
- Uchanski, RM. Spectral and temporal contributions to speech clarity for hearing impaired listeners. Department of Electrical Engineering and Computer Science. Massachusetts Institute of Technology; 1988. Unpublished Doctoral Dissertation
- Uchanski RM. Segment durations in conversational and clear speech. 1992 Unpublished manuscript.
- Uchanski, RM. Clear speech. In: Pisoni, DB.; Remez, R., editors. *The handbook of speech perception.* Malden, MA/Oxford, UK: Blackwell; 2005. p. 207-35.
- Uchanski RM, Choi SS, Braida LD, Reed CM, Durlach NI. Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech and Hearing Research* 1996;39:494–509. [PubMed: 8783129]
- Uther M, Knoll M, Burnham D. Do you speak E-N-G-L-I-S-H? A comparison of foreigner- and infant-directed speech. *Speech Communication* 2007;49:2–7.
- Volaitis LE, Miller JL. Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America* 1992;92(2 Pt 1):723–35. [PubMed: 1506527]
- Wassink A, Wright R, Franklin A. Intraspeaker variability in vowel production: an investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers. *Journal of Phonetics* 2006;35:363–79.
- Zeng FG, Liu S. Speech perception in individuals with auditory neuropathy. *Journal of Speech, Language, and Hearing Research* 2006;49:367–80.