Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners

Ann R. Bradlow[a] and Jennifer A. Alexander[b]
Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, Illinois 60208

(Received 25 July 2006; revised 11 January 2007; accepted 17 January 2007)

Previous research has shown that speech recognition differences between native and proficient non-native listeners emerge under suboptimal conditions. Current evidence has suggested that the key deficit that underlies this disproportionate effect of unfavorable listening conditions for non-native listeners is their less effective use of compensatory information at higher levels of processing to recover from information loss at the phoneme identification level. The present study investigated whether this non-native disadvantage could be overcome if enhancements at various levels of processing were presented in combination. Native and non-native listeners were presented with English sentences in which the final word varied in predictability and which were produced in either plain or clear speech. Results showed that, relative to the low-predictability-plain-speech baseline condition, non-native listener final word recognition improved only when both semantic and acoustic enhancements were available (high-predictability-clear-speech). In contrast, the native listeners benefited from each source of enhancement separately and in combination. These results suggests that native and non-native listeners apply similar strategies for speech-in-noise perception: The crucial difference is in the signal clarity required for contextual information to be effective, rather than in an inability of non-native listeners to take advantage of this contextual information per se. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2642103]

PACS number(s): 43.71.Es, 43.71.Hw, 43.71.Gv [PEI] Pages: 2339–2349

I. INTRODUCTION

Speech recognition differences between native and highly proficient non-native listeners tend to emerge under suboptimal listening conditions. For example, several studies have shown that in the presence of background noise or reverberation, non-native listeners who had attained a very high level of proficiency in English were less accurate at speech recognition than native listeners. In contrast, under more favorable listening conditions, speech recognition accuracy by these highly proficient non-native listeners was comparable to that of native listeners (e.g., Nábelek and Donahue, 1984; Takata and Nábelek, 1990; Mayo *et al.*, 1997; Meador *et al.*, 2000; Rogers *et al.*, 2006). Thus, the general pattern of experimental findings, and the common experience of non-native listeners, is that the detrimental effects of environmental signal distortion are greater for non-native than native language speech perception.

One possible explanation for this phenomenon is that the primary source of the sharper decline for non-native than native listener speech-in-noise perception is at the segmental level. According to this explanation, the masking effects of noise on the acoustic cues to phoneme identification are more detrimental for non-native listeners than for native listeners, presumably due to their reduced experience with the full range of cues for any given phoneme. Specifically, native listeners are likely to have developed highly effective and efficient strategies for compensating for the masking effects of background noise and reverberation by focusing their attention on segmental cues that are less vulnerable to noise-related distortion [e.g., see Parikh and Loizou (2005), and Jiang *et al.* (2006) for some examples of noise-induced cue-weighting shifts in English]. In contrast, based on learned patterns of perception from the native language, non-native listeners may attend primarily to cues that, while relatively reliable in the non-native target language in quiet, are largely obscured by background noise.

An alternative explanation for the extra difficulty of non-native listeners under suboptimal listening conditions is that noise and/or reverberation have detrimental effects at all levels of processing with the result that overall non-native levels of performance on sentence- or word-in-noise recognition tasks reflect cumulative effects of noise throughout the processing system. For example, in addition to the dramatic effects of noise on acoustic cues to segment identity, the presence of background noise can dramatically affect listener access to prosodic boundary cues, such as silent pauses or sudden rises and falls in fundamental frequency. Native listeners may be able to compensate for this loss of phonetic information for phrase and discourse structure by drawing on their highly practiced sentence processing mechanisms. In contrast, due to their relatively poorly developed syntactic, semantic, and pragmatic processing skills in the target language, non-native listeners cannot readily draw on higher-level linguistic structural and contextual information in order to recover from losses at the perceptual level.

In a recent direct test of these alternatives, Cutler *et al.* (2004) examined native and non-native listener phoneme perception in the context of meaningless CV and VC syllables. These syllable-sized stimuli ensured that the perception of phoneme-level information was isolated from lexical-

[a]Electronic mail: abradlow@northwestern.edu
[b]Electronic mail: jenalex@northwestern.edu

or sentence-level information. Cutler *et al.* found similar declines in English phoneme identification with a decrease in signal-to-noise ratio for native and for highly proficient Dutch-speaking non-native listeners. This finding indicates that disproportionate difficulties with speech-in-noise perception for non-native listeners such as these probably do not stem from disproportionate effects of noise at the level of phoneme identification where word- and sentence-level factors are irrelevant. Instead, Cutler *et al.* (2004) suggest that the special troubles of non-native speech-in-noise perception result from the compounding of difficulties at lower levels of processing with limitations at higher levels.

Though limited to highly proficient Dutch listeners, the finding reported in Cutler *et al.* (2004) dovetails perfectly with a finding reported in Mayo *et al.* (1997). These authors found that highly proficient late bilinguals (listeners who had acquired the target language after puberty) benefited less from sentence- level contextual information for word recognition than either native listeners or highly proficient early bilinguals (who had acquired the target language as an infant or toddler). Mayo *et al.* (1997) examined sentence-final word recognition across various signal-to-noise ratios and across conditions in which the target word was either highly predictable or not at all predictable from the preceding sentence context. The native and early bilingual listeners tolerated significantly greater noise levels for high than for low predictability sentences. In contrast, the late bilinguals showed no difference in noise tolerance levels across high and low predictability sentences. Furthermore, the slopes of the psychometric functions (accuracy across signal-to-noise ratios) for the native listeners and the early bilinguals were significantly greater for the high predictability sentences than for the low predictability sentences; that is, there was a dramatic effect of noise on final word recognition accuracy in exactly the sentences where context mattered. In contrast, the late bilinguals showed approximately parallel psychometric functions for the high and low predictability sentences. Thus, while the native and early bilingual listeners showed a strong benefit from contextual information at the sentence level, relatively late (yet highly proficient) non-native listeners showed essentially no benefit for final word recognition in high predictability sentences as compared with final word recognition in low predictability sentences.

A remaining question is whether the observed non-native listener deficit in drawing on higher level contextual information to aid in speech recognition persists even under enhanced signal conditions. It is possible that the sentence-in-noise perception patterns across native and non-native listeners are qualitatively similar in showing reduced exploitation of contextual information under degraded signal conditions and significant benefits from context under more favorable listening conditions. Indeed, if extended over a wider range of noise levels in both directions, it is possible that the average high and low predictability psychometric functions presented by Mayo *et al.* (1997) for all listeners groups (native, earlier bilinguals, and later bilinguals) would diverge dramatically at some point on the low-noise end and converge at some point on the high-noise end of the noise level axis, indicating some degree of context dependency for all listeners. Thus, it is possible that non-native speech-in-noise perception is not necessarily doomed by limited abilities to exploit contextual information as a means of recovery from recognition losses at the levels of phoneme identification and lexical access. Instead, if given a signal of sufficient acoustic clarity, non-native listeners (both early and late learners) may exhibit a strong ability to exploit semantic-contextual information at the sentence level.

Accordingly, the purpose of the present study was to investigate the ability of non-native listeners to benefit from a combination of semantic and acoustic enhancements for sentence-in-noise recognition. We tested the hypothesis that for all listeners, including native listeners as well as non-native listeners across the full range of proficiency levels in the target language, speech recognition accuracy is facilitated by the availability of higher-level semantic information to the extent that such information is well specified in the signal. What varies across listeners with different levels of proficiency and experience with the target language is the definition of "well specified." For native listeners with highly practiced skills in phonetic processing of the language, contextual information early on in an utterance may be sufficiently well perceived even in conditions with relatively high levels of signal distortion and/or masking to provide an effective means of recovering from perceptual losses. In contrast, for non-native listeners with less practice and experience with the sound structure of the target language, contextual information that occurs early in an utterance may only be a useful source of information for later-occurring portions of the utterance if it is sufficiently well-perceived. This, in turn, may require a relatively high degree of signal clarity. A specific prediction of this hypothesis that we set out to test is that under conditions where the signal clarity is enhanced substantially due to a clear speaking style on the part of the talker, non-native listeners should exhibit a recognition advantage for words that are highly predictable from the preceding utterance in comparison with words that are not at all predictable from the preceding context. That is, we sought to extend the finding of Mayo *et al.* (1997) to the situation where greater contextual information is made available to non-native listeners by means of signal enhancement through naturally produced clear speech. In contrast to Mayo *et al.* (1997) (and due primarily to practical limitations regarding the available pool of study participants) the present study did not systematically vary non-native listener proficiency. The focus of this study was instead on identifying the conditions under which a group of non-native listeners with varying levels of target language proficiency could make use of contextual information for the processing of an incoming spoken word.

The overall design of this experiment involved the manipulation of two independent factors: semantic cues and acoustic cues. Following numerous previous studies (e.g. Kalikow *et al.*, 1977; Mayo *et al.*, 1997; Fallon *et al.*, 2002 and many others) the semantic-contextual cues were manipulated by using English sentences in which the final word was highly predicted by the preceding words ("high probability sentences") and sentences in which the final word could not be predicted on the basis of the preceding words ("low probability sentences"). The acoustic-phonetic cues were manipulated by including English sentence recordings in both plain and clear speaking styles (e.g., Picheny *et al.*, 1985; Ferguson and Kewley-Port, 2002; Bradlow and Bent, 2002; Uchanski, 2005; Smiljanic and Bradlow, 2005). Both native and non-native English listeners responded to all four types of sentences (high probability clear speech, high probability plain speech, low probability clear speech, and low probability plain speech), allowing us to assess the separate and combined effects of semantic and acoustic cues to speech-in-noise perception by native and non-native listeners.

In contrast to the terminology used in previous work (e.g., Picheny *et al.*, 1985, 1986, 1989; Payton *et al.*, 1994; Uchanski *et al.*, 1996; Ferguson and Kewley-Port, 2002; Bradlow and Bent, 2002; Uchanski, 2005; Smiljanic and Bradlow, 2005, and several others), in this report we have adopted the term "plain speech" instead of "conversational speech." We propose this change in terminology in order to better reflect the fact that this mode of speech production is distinct from truly conversational speech as presented in, for example, the Buckeye Corpus of Conversational Speech (http://buckeyecorpus.osu.edu/), in which talkers were recorded conversing freely with an interviewer in a small seminar room. For our purposes, the key distinction between "plain" and "clear" speech is with respect to intelligibility. In view of the fact that the plain speech samples are read from a script (i.e., not spontaneous responses to a topic or interviewer's question) and are recorded in a relatively formal laboratory setting (i.e., in a sound-attenuated booth rather than in a more relaxed conversational setting), and since their primary purpose is to serve as a baseline from which to measure the intelligibility advantage of clear speech, we refer to them in the present paper as plain rather than conversational speech recordings.

II. METHOD

A. Materials

Since none of the previously published sets of high and low predictability sentences were designed for use with the population of non-native listeners of interest in this study (i.e., non-native listeners with considerably less experience with spoken English than the participants in the study of Mayo *et al.*, 1997), we followed the general procedures outlined in Fallon *et al.* (2002) to develop a new set of sentences. First, a list of approximately 300 high probability sentences was compiled by combining the sentence lists published in Kalikow *et al.* (1977), Bench and Bamford (1979), Munro and Derwing (1995), and Fallon *et al.* (2002) with some original sentences of our own that followed that same general pattern of those in Kalikow *et al.* (1977) and Fallon *et al.* (2002). The sentences were printed with the final (target) word of each sentence replaced by a dotted line. The sentences were randomly arranged into two lists ("List A" and "List B," respectively); each of the two lists contained exactly half of the sentences.

In order to assess the predictability of the final word in these candidate sentences for our target subject population, 24 non-native English speakers evaluated the sentences in Lists A and B. These non-native English speakers were recruited from the group of participants in the Northwestern University International Summer Institute (ISI) 2004 (as described in the following, this program provides intensive English language instruction and general acculturation for incoming international graduate students). Each participant was given a printout of either List A or List B and was asked to make his or her best guess as to the identity of the final word of each sentence. They were required to work individually and were not allowed to use dictionaries or any other reference resources. The task took between 20 and 50 min to complete. All of these subjects were paid for their participation.

Following this initial test, 113 of the "best" sentences, that is, those that yielded the most consistent responses, were combined into a third list, "List C." We then presented List C with the final (target) word of each sentence replaced by a dotted line to 14 native English speakers. These participants were recruited from the Northwestern University Department of Linguistics subject pool and received course credit for their participation. Their responses were tallied and, again, only those sentences that yielded the most consistent responses were included in a fourth list, "List D." List D, which contained 89 sentences, was then checked for predictability with the same procedure as described earlier by 9 non-native English speakers who were recruited from the Northwestern University English as a Second Language (ESL) program. Finally, J.A.A. created a low predictability match for each high predictability sentence using sentence frames modified slightly from those published in Fallon *et al.* (2002).

The final list (provided in the Appendix) consisted of 120 sentences (60 high and 60 low predictability). In this final set of sentences, 43% (26/60) of the high predictability sentences were completed by the final group of 9 non-native respondents with 100% consistency (i.e., 9/9 respondents filled in the same final words), 30% (18/60) were completed with 89% consistency (i.e., 8/9 respondents filled in the same final words), 23% (14/60) were completed with 78% consistency (i.e., 7/9 respondents filled in the same final words), and 3% (2/60) were completed with 67% consistency (i.e., 6/9 respondents filled in the same final words). For the test with native listeners, 83% of the words (50/60) were completed with 100% consistency (14/14 respondents), 13% (8/60) were completed with 93% consistency (13/14 respondents), and the remaining 3% (2/60) were completed with 86% consistency (12/14 respondents).

One monolingual female talker of American English (aged 30 years, with no known speech or hearing impairment) was recorded producing the full set of 120 sentences in both clear and plain speaking styles (for a total of 240 recorded sentences). The complete set of 120 sentences was read first in the plain speaking style followed by a second recording in the clear speaking style. At the time of the plain style recording the talker was not aware of the fact that a clear speaking style condition would follow. The plain speaking style was elicited by instructing the talker to "read the sentences as if you are talking to someone familiar with your voice and speech patterns." The clear speaking style was elicited by instructing her to "read the sentences very clearly, as if you are talking to a listener with a hearing loss or to a non-native speaker learning your language."

The high and low predictability sentences were mixed together and printed in random order on sheets of paper in groups of 21 with 2 filler sentences at the top and bottom of each page, giving a total of 25 sentences per page (21 target sentences +4 fillers) over a total of 6 pages. The filler sentences were presented to help the talker avoid "list intonation" around the page boundaries. The recording session was conducted in a double-walled, sound-attenuated booth. The talker wore a head-mounted microphone (AKG C420 Headset Cardioid Condenser) and the speech was recorded directly onto flash card using a Marantz PMD670 Professional Solid-State digital recorder (22.050 kHz sampling rate, 16 bit amplitude resolution).

The recorded sentences were segmented into individual sentence-length files which were subsequently equated for rms amplitude across the whole sentence duration. Each file was then digitally mixed with speech-shaped noise at a signal-to-noise ratio of +2 dB for presentation to the non-native test subjects and at a signal-to-noise ratio of −2 dB for presentation to the native listener control subjects. Each of the final stimulus files consisted of a 400 ms silent leader, followed by 500 ms of noise, followed by the speech-plus-noise file, and ending with a 500 ms noise-only tail. The noise in the 500 ms, noise-only header and tail was always at the same level as the noise in the speech-plus-noise portion of the stimulus file. These signal-to-noise ratios (+2 dB for non-natives and −2 dB for natives) were determined on the basis of our experience with prior speech-in-noise experiments with comparable (but different) sentence materials and listener groups, as well as some (rather limited) pilot testing with the current set of stimuli (all sentence types, i.e., high and low context in plain and clear speech) and a small number of subjects (6 native and 3 non-native listeners).

We selected these signal-to-noise ratios with the goal of eliciting comparable levels of speech recognition accuracy for the native and non-native listeners. The advantage of this approach is that it ensures that all listeners are operating at approximately the same effective level of speech recognition accuracy and therefore group differences in performance relative to the baseline condition (i.e., low context sentences in plain speech) are not subject to confounding influences of starting level differences. There is precedent for adopting this approach in, for example, the literature on speech perception by elderly listeners (e.g., Sommers, 1996, 1997). Here we adopt this approach with the understanding that, while the effect on speech recognition of higher noise levels for native listeners may not be qualitatively identical to the effect of perceptual mistuning of the non-native listeners to the target language, the overall equivalence of performance levels in the baseline condition due to different signal-to-noise ratios facilitated valid comparisons of improvement with contextual and/or phonetic enhancement.

B. Participants

Ninety-three adults participated in this study. Of these participants, 57 were native speakers and 36 were non-native speakers of American English. All of the native English speakers were recruited from the Northwestern University Department of Linguistics subject pool and received course credit for their participation. Of the 57 native speakers of English, 21 were excluded from the final analyses due to experimenter error ($n=8$), a bilingual language background ($n=11$) or a reported speech or hearing impairment ($n=2$). The remaining 36 native English speaking participants (22 females and 14 males) ranged in age from 17 to 30 years.

The non-native listeners were recruited from ISI 2005 (Northwestern's International Summer Institute) and received payment for their participation. The participants in this program had all been accepted into a graduate program at Northwestern University and had therefore demonstrated a relatively high level of proficiency with English communication (as measured by a minimum score of 560 on the pencil-and-paper TOEFL examination or 220 on the computer-based version of the test). Based on the accepting departments' subjective experiences with the spoken English skills of previous students from the students' home countries, the participants had been nominated for participation in this summer program, which provides one month of intensive English language training as well as a general introduction to life in the United States.

Table I provides additional details regarding the non-native participants in this study. As shown in Table I, the non-native participants came from various native language backgrounds with the breakdown as follows: Mandarin Chinese ($n=23$), Italian ($n=3$), Korean ($n=2$), Tamil ($n=2$), and 1 each of French, German, Gujarati, Hindi, Kikuyi, and Telugu. They ranged in age from 21 to 32 years, and had 9–23 years of English language study. At the time of testing, the majority of non-native participants had between 1 and 4 weeks of experience living in an English speaking environment. Three non-native listeners had several more weeks (6, 11, and 35 weeks) and 2 non-natives had 2–3 years worth of experience living in the United States prior to testing. As part of the ISI program orientation, participants were divided into eight groups based roughly on proficiency level as determined by the Speaking Proficiency English Assessment Kit (SPEAK) test. (As a reference point, note that Northwestern University requires a SPEAK score of 50 for a non-native English speaking student to be appointed as a teaching assistant). As shown in Table I and explained further in the following, we attempted to distribute the number of non-native listeners from each proficiency group evenly across four experimental conditions (A–D).

It should be noted that the group of non-native listeners was not balanced in terms of native language background. The group was dominated by native speakers of one language: 23 of the 36 non-native listeners, or 64%, were Mandarin speakers. Moreover, the remaining 13 non-native listeners came from vastly different native language backgrounds, making it impossible to conduct any meaningful comparisons across listeners with different native languages. This distribution of native languages is typical of the Northwestern ISI program participants and is a fairly accurate reflection of the distribution of international graduate students across the university. Since the task in the present

TABLE I. Some measures of the English language experience of the non-native listener participants.

Presentation condition	SPEAK[a] test score	Native language	Age at test	Time in USA (weeks)	English study (years)
A	51.6	Tamil	23	4	18
A	49.2	Mandarin	22	2	12
A	48	Mandarin	24	2	12
A	46	Mandarin	22	157	17
A	42.5	Mandarin	22	1	12
A	40	Mandarin	21	4	11
A	39.5	Mandarin	22	4	10
A	37.9	Mandarin	25	4	12
A	37.0	Mandarin	22	1	10
A	36.7	Mandarin	23	3	12
B	59.1	German	28	35	22
B	56.2	Tamil	21	6	17
B	47.7	Korean	27	4	16
B	46.5	Mandarin	23	1	13
B	41.7	Mandarin	22	2	10
B	40.4	Mandarin	26	2	13
B	38.6	Mandarin	30	4	17
B	37.5	Mandarin	25	4	13
B	30	Mandarin	22	4	10
C	52.0	French	23	116	12
C	50.4	Hindi	23	3.5	18
C	42.9	Mandarin	22	2	10
C	40.5	Mandarin	28	3	16
C	40	Mandarin	22	1	10
C	39.75	Mandarin	22	4	10
C	N/A	Italian	22	2	16
D	56.7	Gujarati	23	11	19
D	49.6	Kikuyu	26	4	23
D	43.7	Mandarin	21	2	10
D	41.8	Italian	24	2	16
D	41.2	Mandarin	24	2	9
D	40	Telugu	31	2	9
D	39.3	Mandarin	22	2	14
D	36.7	Mandarin	21	4	9
D	32.5	Korean	32	3	19
D	N/A	Italian	26	2	16

SPEAK=Speaking English Assessment Kit.

study (recognition of words in simple English sentences) requires processing over many levels of representation rather than focusing on specific phonetic contrasts, we assumed that our data would reveal general patterns of non-native language speech recognition rather than specific patterns of interactions between structural features of the target language and the listeners' native languages. Nevertheless, we acknowledge this limitation of our data.

C. Procedure

Both native and non-native subjects were tested in groups of one to three. The data collection session began with a language background questionnaire which probed the subjects' language learning experiences (both native and foreign languages) as well as their self-reported performance on standardized tests of English language proficiency (non-native subjects only). For the sentence-in-noise recognition

TABLE II. Distribution of sentences across the four presentation conditions. Sentence numbers correspond to the listing in the Appendix .

Condition	High context plain style	Low context plain style	High context clear style	Low context clear style
A	1–15	31–45	16–30	46–60
B	16–30	46–60	1–15	31–45
C	31–45	1–15	46–60	16–30
D	46–60	16–30	31–45	1–15

test, each participant was seated in front of a computer monitor in a sound-attenuated booth. The audio files were played via the computer sound card over headphones (Sennheiser HD580) at a comfortable listening level, which was set by the experimenter before the start of the experiment. The participant's task was to listen to each sentence stimulus and write just the final word on specially prepared answer sheets. After each trial, the participant pressed a button on a response box to trigger the start of the next trial. Each trial was presented only once, but subjects could take as long as they needed to record their responses.

In order to familiarize subjects with the task, the test session began with 8 practice items (these items were not included in the subsequent test). After completion of these practice items the experimenter checked that the subject understood the task and was ready to begin the test. Each subject responded to a total of 60 sentences, which included a randomly ordered presentation of 15 high context sentences in plain speech, 15 low context sentences in plain speech, 15 high context sentences in clear speech, and 15 low context sentences in clear speech. Over the course of the entire test session, each subject heard each of the 60 target words only once. In order to guard against any inherent differences in either keyword or sentence intelligibility across the four style-context conditions, four test conditions were compiled such that the 60 target words/sentences were evenly distributed across the conditions (see Table II). For the high probability sentences, the response consistency rates from both the native and non-native respondents in the testing during the sentence development phase were evenly distributed across the four sublists shown in Table II: The four average consistency rates for the native respondents ranged from 96% to 100%; for the non-native respondents consistency rates ranged from 85% to 92%. As shown in Table I, these four presentation conditions were evenly distributed within the non-native participants' proficiency level groups, and equal numbers of native subjects participated in each of the four conditions. The average SPEAK score for participants in conditions A, B, C, and D were 42.9, 44.2, 44.3, and 42.4, respectively, which are very close to the overall average SPEAK score of 43.4.

Each subject received a final word recognition accuracy score out of 15 for each of the four word conditions (high context plain speech, low context plain speech, high context clear speech, and low context clear speech). Cases of responses containing alternate or incorrect spellings were accepted as correct when the scorer felt sure that the participant had correctly identified the target word but simply failed to spell it as expected. The scores were converted to percent

TABLE III. Final word durations in milliseconds for each of the four sentence types.

	High context plain style	Low context plain style	High context clear style	Low context clear style
Mean	452	459	594	613
Median	450	461	596	614
Std. Dev.	62	58	69	74
Std. Error	8	8	9	10
Minimum	242	305	424	461
Maximum	610	572	835	812

correct scores, and then converted to rationalized arcsine transform units (RAU) (Studebaker, 1985) for the statistical analyses. Scores on this scale range from −23 RAU (corresponding to 0% correct) to +123 RAU (corresponding to 100% correct).

III. RESULTS

A. Durational analyses of the sentence stimuli

Before turning to the final word recognition accuracy results, we examined the durations of all final words in the full set of 240 sentence stimuli. A decrease in speaking rate is a well-documented feature of clear speech (e.g., Picheny *et al.*, 1986, 1989; Payton *et al.*, 1994; Uchanski *et al.*, 1996; Bradlow *et al.*, 2003; Krause and Braida, 2002, 2004; Smiljanic and Bradlow, 2005); we therefore expected a highly significant increase in final word duration for clear versus plain speech. Moreover, several studies have indicated a general tendency toward some degree of hyperarticulation for words in low predictability contexts relative to words in high predictability contexts (e.g., Lieberman, 1963; Jurafsky *et al.*, 2001; Wright, 2002; Aylett and Turk, 2004; Munson and Solomon, 2004; Munson, 2007; Scarborough, 2006). Since the purpose of this study was to examine the interaction of higher-level semantic cues and lower-level acoustic cues to final word recognition, it was necessary to determine the extent to which the words in our high versus low predictability sentence contexts differed at the acoustic level too. Here we focus exclusively on word duration as an indicator of hyperarticulation, although it should be noted that the effects of style and predictability are likely to be reflected along multiple acoustic dimensions.

Table III shows some descriptive statistics for final word duration in each of the four sentence types. A two-factor repeated-measures analysis of variance (ANOVA) showed highly significant main effects of style [plain versus clear, $F(1,59)=420.66$, $p<0.0001$] and context [high versus low predictability, $F(1,59)=9.15$, $p=0.004$]. The two-way interaction was not significant. This general pattern of increased duration for clear versus plain speech is typical of this speaking style shift and simply indicates that, as expected, the talker in this study produced a large plain versus clear speaking style difference. As shown in Table III, the increase in word duration from plain to clear speech was 154 ms (33.6%) and 142 ms (31.4%) in the low and high predictability contexts, respectively. The small but reliable decrease in duration for high versus low predictability final words is also consistent with previous work indicating some degree of probability-dependent hypo-articulation (e.g., Scarborough, 2006) such that a given word in a highly predictable context is generally reduced relative to its occurrence in a less predictable context. In summary, this comparison of final word durations across the four sentence types established that the talker produced a large and consistent increase in word duration for clear relative to plain speech, and a small but consistent decrease in duration for words in high versus low predictability contexts.

B. Final word recognition accuracy

Figure 1 and Table IV show the final word recognition accuracy scores for the native and non-native listeners (left

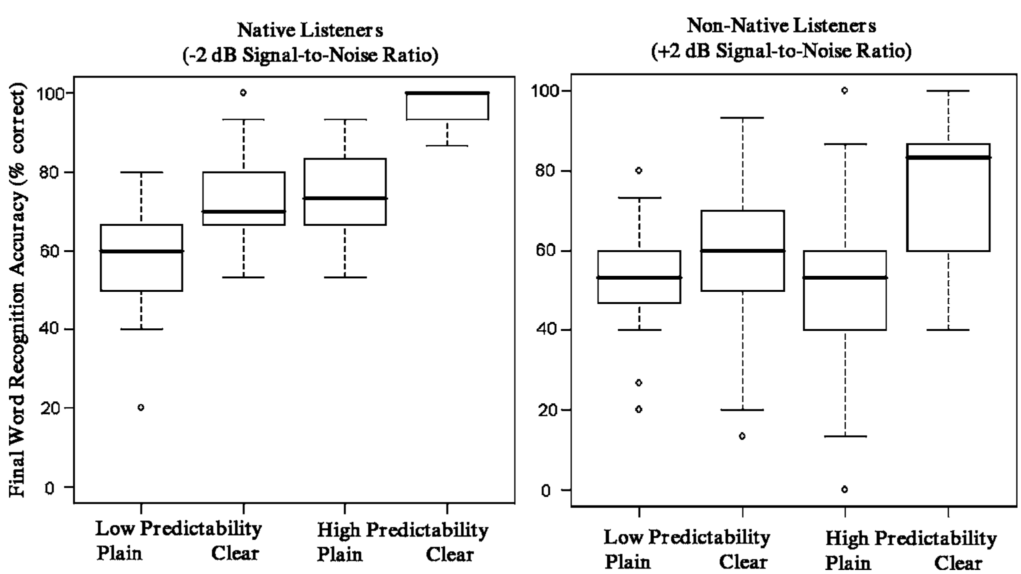

FIG. 1. Final word recognition accuracy (in % correct) for the native (left panel) and non-native listeners (right panel) in both styles of speech (plain and clear) and in both high and low predictability contexts. Whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range shown in the boxes.

TABLE IV. Final word recognition accuracy in percent correct for each of the four sentence types. [Note that all statistics reported in the text were conducted on transformed data along the RAU (rationalized arcsine units) scale of the percent correct values reported here].

Listener	Native				Non-native			
Style	Plain		Clear		Plain		Clear	
Context	High	Low	High	Low	High	Low	High	Low
Mean	77	58	98	73	50	51	77	60
Median	73	60	100	70	53	53	83	60
Std. Dev.	10	12	4	11	18	13	17	18
Std. error	1.7	1.9	0.65	1.9	3.1	2.1	2.8	3.0
Minimum	53	20	87	53	0	20	40	13
Maximum	93	80	100	100	100	80	100	93

panel and right panel, respectively) in both speaking styles and with both types of sentences. Recall that the native and non-native listeners were presented with the sentences mixed with noise at a −2 and +2 dB signal-to-noise ratio, respectively. This difference in signal-to-noise ratio was introduced into the procedural design as an attempt to equate the levels of performance across the native and non-native listener groups in the "baseline" condition of low predictability sentences in plain speech. As shown in Fig. 1 and Table IV, the average native and non-native listener final word recognition accuracy scores in the plain speech low predictability context were 58% correct and 51% correct, respectively. While the difference between these scores is significant [unpaired $t(70)=2.2$, $p=0.03$], we have assumed for the purposes of the subsequent analyses that they are close enough to indicate that the two groups of listeners were operating in approximately the same range of accuracy scores and that any observed differences in the ability to take advantage of semantic-contextual cues or to benefit from clear speech enhancements were due to differences in listener-related, native versus non-native language processing strategies rather than in task-related factors that may have differed across the two listener groups.

The overall pattern of word recognition accuracies showed that, relative to the baseline condition of low predictability sentence contexts in plain speech, non-native listener final word recognition accuracy improved by 9 percentage points (from 51% to 60% correct) when presented with clear speech. In contrast, the same non-native listeners showed no benefit for final word recognition from low to high predictability sentence contexts (51% to 50% correct) in the plain speech style. However, they showed a substantial improvement of 26 percentage points from the baseline condition (low predictability sentences in plain speech) when both top-down, semantic-contextual cues and bottom-up, acoustic-phonetic enhancements were available (from 51% correct to 77% correct).

The native listeners in this study benefited from both sources of enhancement whether presented singly or in combination, relative to the baseline condition of low predictability sentence contexts in plain speech. Native listener final word recognition accuracy improved by 15 percentage points (from 58% to 73% correct) when presented with clear speech, by 19 percentage points (from 58% to 77% correct) when presented with high predictability sentences, and by 40 percentage points (from 58% to 98% correct) when presented with high predictability sentences in clear speech.

Separate two-way ANOVAs with style and context as within-subjects factors for the native and non-native listener groups were conducted on the RAU transformed data. For the native listeners, both main effects and the two-way style-context interaction were highly significant at the $p < 0.0001$ level [style: $F(1,70)=161.59$, context: $F(1,70)=223.144$, style-context: $F(1,70)=34.04$]. Post hoc comparisons showed significant differences (at the $p < 0.0001$ level) for all of the pair-wise comparisons except for low context clear speech versus high context plain speech. This pattern of results indicates that the native listeners derived a significant final word recognition benefit from both semantic-contextual information and acoustic-phonetic enhancements and that these two sources of intelligibility enhancement worked together and were mutually enhancing in the high predictability, clear speech condition.

For the non-native listeners, both main effects and the two-way style-context interaction were highly significant at the $p < 0.0001$ level [style: $F(1,70)=60.22$, context: $F(1,70)=27.35$, style-context: $F(1,70)=19.48$]. Post hoc comparisons showed significant differences at the $p < 0.0001$ level for all of the pair-wise comparisons except for two cases: The difference between plain and clear speech in the low context condition was significant at the $p < 0.001$ level and there was no difference between the high and low context conditions in plain speech. Thus, for these non-native listeners, final word recognition accuracy generally improved from plain to clear speaking styles in both high and low predictability sentence contexts; however, these non-native listeners only benefited from a highly predictive context in the clear speaking style. The lack of a context effect in plain speech is consistent with the finding of Mayo *et al.* (1997) that highly proficient, late bilinguals benefited less from contextual information than native listeners and highly proficient, early bilinguals. In the present study, we extended this finding by demonstrating that non-native speech recognition can be improved by a combination of semantic-contextual and acoustic-phonetic enhancements as shown by the boost in performance in the high predictability clear speech condition relative to the other three conditions.

A major difference between the present study and the Mayo *et al.* (1997) study is that in the present study we did not directly compare performance across groups of non-native listeners with different levels of English proficiency or with different ages of English acquisition onset. Instead, the primary analyses of the present study treated all of the non-native listeners as members of a single, broadly-defined group. Nevertheless, in order to gain some insight into the role of proficiency in determining access to acoustic-phonetic and contextual enhancements for word recognition, we conducted some additional correlational analyses. As shown in Table I, 34 of the 36 non-native listeners reported SPEAK scores. The range of scores on this test was wide enough to permit a correlational analysis between English proficiency (as reflected by these scores) and final word recognition in the present speech-in-noise perception test. For these listeners, SPEAK score was positively correlated with the average final word recognition score, i.e., the overall score averaged across speaking styles and sentence predictability contexts (Pearson correlation $=0.722$, $p < 0.0001$) and with the average high-low predictability difference score, i.e., the size of the context effect averaged across both speaking styles (Pearson correlation $=0.346$, $p < 0.05$). In contrast, the size of the clear speech effect (i.e., the clear-plain speech difference score averaged across predictability contexts) was not significantly correlated with SPEAK score. This pattern of correlations indicates that, while the clear speech effect apparently did not depend strongly on overall proficiency within the range represented by these non-native listeners, the ability to take advantage of higher-level, semantic-contextual cues did improve with increasing general English language proficiency.

IV. GENERAL DISCUSSION

The overall purpose of the present study was to assess whether non-native listener speech-in-noise perception could be improved by the availability of both top-down, semantic-contextual cues and bottom-up, acoustic-phonetic enhancements. The data complement data from previous research by establishing that, if given sufficiently rich information by means of a clear speaking style on the part of the talker, non-native listeners can indeed enhance their word recognition accuracy by taking advantage of sentence-level contextual information. This finding is consistent with the hypothesis that native and non-native listeners are both able to use contextual information to facilitate word recognition provided that the contextual information is well specified in the signal. The data establish further that naturally produced clear speech is an effective means of enhancing access to signal-dependent information for both native and non-native listeners thereby allowing the beneficial effects of contextual information to reveal themselves. A noteworthy implication of this finding is that, while listeners may have to turn up the volume on their radios as they switch from listening in their native language to listening in a non-native language, they may also be able derive dramatic benefit (at all levels of processing) from any signal clarity enhancing device.

With regard to the benefit offered by clear speech, a bottom-up acoustic-phonetic enhancement, the present data are consistent with the finding reported in Bradlow and Bent (2002) that the non-native listener average clear speech benefit was substantially smaller in magnitude than the average native listener clear speech benefit. In the present study, the clear speech benefits in the low predictability context for the native and non-native listeners were 15 and 9 percentage points, respectively. (The data in the high predictability context did not provide for a valid comparison of the magnitude of the clear speech benefit across listener groups since the native listeners reached ceiling levels of performance). In the earlier study, we interpreted this diminished clear speech benefit for non-native listeners as reflecting their reduced experience with the sound structure of the target language relative to native listeners. We reasoned that non-native listeners may not be sensitive to some of the dimensions of contrast that native talkers spontaneously enhance in clear speech production due to their lack of extensive experience with the full range of acoustic-phonetic cues for many of the target language contrasts. Other work has shown that, in addition to language-general features such as a decreased speaking rate and an expanded pitch range, clear speech production involves the enhancement of the acoustic-phonetic distance between phonologically contrastive categories (e.g., Ferguson and Kewley-Port, 2002; Krause and Braida, 2004, Picheny *et al.*, 1986; Smiljanic and Bradlow, 2005, 2007). Therefore, reduced sensitivity to any or all of the language-specific acoustic-phonetic dimensions of contrast and clear speech enhancement would yield a diminished clear speech benefit for non-native listeners. This may appear somewhat surprising given that clear speech production was elicited in our studies by instructing the talkers to speak clearly for the sake of listeners with either a hearing impairment or from a different native language background. However, as discussed further in Bradlow and Bent (2002), the limits of clear speech as a means of enhancing non-native speech perception likely reflect the "mistuning" that characterizes spoken language communication between native and non-native speakers.

A limitation of the Bradlow and Bent (2002) study was that the materials were all meaningful sentences and the listener's task was to write down the full sentences, which were then scored on the basis of a keyword correct count. Thus, since target word predictability was not controlled in the materials of that study, the relatively small clear speech effect for the non-native listeners may have been due (partially or even entirely) to their reduced ability to take advantage of contextual information available in the sentences rather than in their reduced ability to take advantage of the acoustic-phonetic modifications of English clear speech. By directly manipulating final word predictability, the present study addressed this limitation and confirmed that non-native listeners derive a significant, though relatively small benefit from the acoustic-phonetic enhancements of clear speech independently of their reduced ability to take advantage of higher-level semantic-contextual information provided in a sentence. It is thus likely that this smaller non-native clear speech benefit is indeed due to reduced experience with sound structure of the target language.

Two recent studies provide some information regarding the interaction of acoustic- and semantic-level information during native listener spoken language processing that have some bearing on the comparison between native and non-native listeners in the present study. First, in a study of lexical access in English, Aydelott and Bates (2004) showed different reaction time patterns in a lexical decision task depending on whether the speech stimuli were left unaltered or were distorted (by low-pass filtering or time compression). Specifically, they examined lexical decision reaction times to target words presented in a "congruent semantic context" (i.e., in a high predictability sentence context such as "On a windy day it's fun to go out and fly a" for the target word "kite"), in a "neutral semantic context" (i.e., in a low predictability sentence context such as "Its name is" for the target word "kite"), or in an "incongruent semantic context" (i.e., in a sentence context such as "On a windy day it's fun to go out and fly a" for the target word "yert"). The key finding for our purposes was that, when presented with unaltered speech stimuli, the participants showed the expected pattern of increasing reaction times from the congruent to the neutral to the incongruent semantic contexts. In contrast, when presented with distorted speech stimuli, some of these reaction time differences due to semantic context differences were neutralized, indicating that variation in signal clarity can affect the operation of "normal" facilitatory and inhibitory processes at the semantic level. Similarly, in a study of English word segmentation from connected speech, Mattys *et al.* (2005) demonstrated that higher-level, knowledge-driven lexical information interacts with lower-level, signal-derived, sublexical information according to a hierarchical organization of cues with descending weight assignments from lexical to segmental to prosodic cues. Of particular interest with regard to the present study was the finding that these cue weightings were effectively reversed under conditions of signal distortion due to the presence of background (white) noise. Specifically, as signal quality decreased (due to decreasing signal-to-noise ratios), thereby rendering any available contextual information increasingly inaccessible, listeners were forced to rely more heavily on lower-level signal-derived information than on contextual information for word segmentation. Conversely, when presented with intact speech (no added noise) from which contextual information was easily accessible, listeners took advantage of this higher-level information and relied less on lower-level acoustic-phonetic word boundary cues.

While these studies differed from the present study in numerous ways, perhaps most notably by the fact that they examined the effects of acoustic distortion rather than acoustic enhancement, they both demonstrate that, even for native listeners, speech processing strategies that involve higher-level semantic-contextual information can be more or less effective depending on access to the speech signal at the perceptual level. This situation is, of course, highly analogous to the pattern of findings for the non-native listeners in the present study. Like the native listeners in the priming study of Aydelott and Bates (2004) and in the word segmentation study of Mattys *et al.* (2005), the non-native listeners in the present word recognition study were only able to make use of higher-level contextual cues when the lower-level acoustic-phonetic cues were sufficiently clear that the preceding contextual information was indeed easily accessible. When presented with a sufficiently clear signal, the native listeners in the previous priming and segmentation studies and the non-native listeners in the present study all showed processing strategies that involved taking advantage of any available higher-level contextual information.

It is important to note that, in the present study, the signal clarity required to make effective use of contextual information was quite large for the non-native listeners compared to the native listeners, and required both a more favorable signal-to-noise ratio (recall that the non-native and native listeners were presented with stimuli at +2 and −2 dB signal-to-noise ratios, respectively) and a clear speaking style. It remains for future research to determine whether the perceptual patterns indicated in the present study will be obtained with more systematic control over listener proficiency in the target language and across a wider range of signal-to-noise ratios.

When viewed in relation to the work with native listeners presented with degraded signals (Aydelott and Bates, 2004; Mattys *et al.*, 2005), the difference in word recognition patterns across the native and non-native listeners in the present study can be described as a difference in the signal clarity required for semantic-contextual information to be effective, rather than as an inability of non-native listeners to take advantage of contextual information. Thus, as suggested in Sec. I, non-native listener speech-in-noise perception is not necessarily doomed by limited recovery resources at higher levels of processing. Instead, higher level support for information extracted from the acoustic-phonetic level is available to non-native listeners (just as it is for native listeners) albeit in a less efficient mode of operation. An open issue that remains to be investigated further is with regard to the origin and nature of the mechanism that underlies this relative inefficiency of non-native listeners. In all likelihood, the source of this feature of non-native speech processing is multifaceted including factors related to experience-dependent "mistuning" to all levels of linguistic structure of the target language (ranging from the subsegmental, segmental, and prosodic levels of sound structure to the more abstract syntactic, semantic, and pragmatic levels) as well as factors relating to the cognitive demands of managing two (or more) languages. An important challenge for future research is to identify these factors and then to ultimately propose means of overcoming these deficits by either human or technological enhancements. As demonstrated in the present study, clear speech may be a promising avenue to follow toward this goal.

ACKNOWLEDGMENTS

This research was supported by Grant No. NIH-R01-DC005794 from NIH-NIDCD. The authors gratefully ac-

knowledge the assistance of Judy Song and Cynthia Clopper with subject running.

APPENDIX: HIGH AND LOW PREDICTABILITY SENTENCES

High predictability sentences

1. The meat from a pig is called pork.
2. For dessert he had apple pie.
3. Sugar tastes very sweet.
4. The color of a lemon is yellow.
5. My clock was wrong so I got to school late.
6. In spring, the plants are full of green leaves.
7. A bicycle has two wheels.
8. She made the bed with clean sheets.
9. The sport shirt has short sleeves.
10. He washed his hands with soap and water.
11. The child dropped the dish and it broke.
12. The bread was made from whole wheat.
13. The opposite of hot is cold.
14. A wristwatch is used to tell the time.
15. The war plane dropped a bomb.
16. She cut the cake with a knife.
17. A chair has four legs.
18. Cut the meat into small pieces.
19. The team was trained by their coach.
20. The lady wears earrings in her ears.
21. People wear shoes on their feet.
22. When sheep graze in a field, they eat grass.
23. A rose is a type of flower.
24. Football is a dangerous sport.
25. The heavy rains caused a flood.
26. Bob wore a watch on his wrist.
27. Monday is the first day of the week.
28. The pan that was just in the oven is very hot.
29. Rain falls from clouds in the sky.
30. The boy laughed because the joke was very funny.
31. To cool her drink, she added a few cubes of ice.
32. A quarter is worth twenty-five cents.
33. An orange is a type of fruit.
34. People wear scarves around their necks.
35. I wrote my name on a piece of paper.
36. For your birthday I baked a cake.
37. Birds build their nests in trees.
38. My parents, sister and I are a family.
39. The good boy is helping his mother and father.
40. People wear gloves on their hands.
41. A book tells a story.
42. A pigeon is a kind of bird.
43. The sick woman went to see a doctor.
44. The lady uses a hairbrush to brush her hair.
45. At breakfast he drank some orange juice.
46. Last night, they had beef for dinner.
47. A racecar can go very fast.
48. Many people like to start the day with a cup of coffee.
49. He brought the book to school from home.
50. I wear my hat on my head.
51. Red and green are colors.
52. The stars come out at night.
53. February has twenty-eight days.
54. The picture is hung high on the bedroom wall.
55. We heard the ticking of the clock.
56. She laid the meal on the table.
57. She looked at herself in her mirror.
58. Elephants are big animals.
59. After my bath, I dried off with a towel.
60. In the morning it gets light, and in the evening it gets dark.

Low predictability sentences

1. Dad looked at the pork.
2. Mom talked about the pie.
3. We think that it is sweet.
4. Mom thinks that it is yellow.
5. He thinks that it is late.
6. She talked about the leaves.
7. He read about the wheels.
8. Dad talked about the sheets.
9. He looked at the sleeves.
10. We talked about the water.
11. We heard that it broke.
12. Dad pointed at the wheat.
13. She thinks that it is cold.
14. This is her favorite time.
15. Dad talked about the bomb.
16. Mom read about the knife.
17. She looked at her legs.
18. There are many pieces.
19. We read about the coach.
20. She pointed at his ears.
21. Mom looked at her feet.
22. Dad pointed at the grass.
23. She read about the flower.
24. This is her favorite sport.
25. He read about the flood.
26. He looked at her wrist.
27. This is her favorite week.
28. Mom thinks that it is hot.
29. Dad read about the sky.
30. Dad thinks that it is funny.
31. He talked about the ice.
32. He pointed at the cents.
33. He pointed at the fruit.
34. She talked about their necks.
35. We talked about the paper.
36. This is her favorite cake.
37. He read about the trees.
38. We read about the family.
39. Mom pointed at his father.
40. She looked at her hands.
41. We looked at the story.
42. We pointed at the bird.
43. Mom talked about the doctor.
44. He pointed at his hair.
45. Mom looked at the juice.
46. He talked about the dinner.
47. She thinks that it is fast.

48. Mom pointed at the coffee.
49. She pointed at the home.
50. She pointed at her head.
51. Mom read about the colors.
52. This is her favorite night.
53. There are many days.
54. We pointed at the wall.
55. She looked at the clock.
56. Dad read about the table.
57. We looked at the mirror.
58. He pointed at the animals.
59. Dad looked at the towel.
60. Dad thinks that it is dark.

Aydelott, J., and Bates, E. (**2004**). "Effects of acoustic distortion and semantic context on lexical access," Lang. Cognit. Processes **19**, 29–56.

Aylett, M., and Turk, A. (**2004**). "The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech," Lang Speech **47**, 31–56.

Bench, J., and J. Bamford, eds. (**1979**). *Speech-Hearing Tests and the Spoken Language of Hearing-Impaired Children* (Academic, London).

Bradlow, A. R., and Bent, T. (**2002**). "The clear speech effect for non-native listeners," J. Acoust. Soc. Am. **112**, 272–284.

Bradlow, A. R., Kraus, N., and Hayes, E. (**2003**). "Speaking clearly for learning-impaired children: Sentence perception in noise," J. Speech Lang. Hear. Res. **46**, 80–97.

Cutler, A., Webber, A., Smits, R., and Cooper, N. (**2004**). "Patterns of English phoneme confusions by native and non-native listeners," J. Acoust. Soc. Am. **116**, 3668–3678.

Fallon, M., Trehub, S. E., and Schneider, B. A. (**2002**). "Children's use of semantic cues in degraded listening environments," J. Acoust. Soc. Am. **111**, 2242–2249.

Ferguson, S. H., and Kewley-Port, D. (**2002**). "Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **112**, 259–271.

Jiang, J., Chen, M., and Alwan, A. (**2006**). "On the perception of voicing in syllable-inital plosives in noise," J. Acoust. Soc. Am. **119**, 1092–1105.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. (**2001**). "Probabilistic relations between words: Evidence from reduction in lexical production," in *Frequency and the Emergence of Linguistics Structure*, edited by J. Bybee and P. Hopper (Benjamins, Amsterdam, The Netherlands), pp. 229–254.

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337–1351.

Krause, J. C., and Braida, L. D. (**2002**). "Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility," J. Acoust. Soc. Am. **112**, 2165–2172.

Krause, J. C., and Braida, L. D. (**2004**). "Acoustic properties of naturally produced clear speech at normal speaking rates," J. Acoust. Soc. Am. **115**, 362–378.

Lieberman, P. (**1963**). "Some effects of semantic and grammatical context on the production and perception of speech," Lang Speech **6**, 172–187.

Mattys, S. L., White, L., and Melhorn, J. F. (**2005**). "Integration of multiple speech segmentation cues: A hierarchical framework," J. Exp. Psychol. **134**, 477–500.

Mayo, L. H., Florentine, M., and Buus, S. (**1997**). "Age of second-language acquisition and perception of speech in noise," J. Speech Lang. Hear. Res. **40**, 686–693.

Meador, D., Flege, J. E., and MacKay, I. R. (**2000**). "Factors affecting the recognition of words in a second language," Bilingualism: Lang. Cognit. **3**, 55–67.

Munro, M., and Derwing, T. (**1995**). "Foreign accent, comprehensibility, and intelligibility in the speech of second language learners," Lang. Learn. **45**, 73–97.

Munson, B., and Solomon, N. P. (**2004**). "The effect of phonological neighborhood density on vowel articulation," J. Speech Lang. Hear. Res. **47**, 1048–1058.

Munson, B. (**2007**). "Lexical access, lexical representation and vowel production," *in Papers in Laboratory Phonology IX*, edited by J. Cole and J. I. Hualde (Mouton de Gruyter, Berlin).

Nábělek, A. K., and Donohue, A. M. (**1984**). "Perception of consonants in reverberation by native and non-native listeners," J. Acoust. Soc. Am. **75**, 632–634.

Parikh, G., and Loizou, P. (**2005**). "The influence of noise on vowel and consonant cues," J. Acoust. Soc. Am. **118**, 3874–3888.

Payton, K. L., Uchanski, R. M., and Braida, L. D. (**1994**). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," J. Acoust. Soc. Am. **95**, 1581–1592.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). "Speaking clearly for the hard of hearing. I. Intelligibility differences between clear and conversational speech," J. Acoust. Soc. Am. **28**, 96–103.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). "Speaking clearly for the hard of hearing. II. Acoustic characteristics of clear and conversational speech," J. Speech Hear. Res. **29**, 434–446.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1989**). "Speaking clearly for the hard of hearing. III. An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech," J. Speech Hear. Res. **32**, 600–603.

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., and Abrams, H. B. (**2006**). "Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing," Appl. Psycholinguist. **27**, 465–485.

Scarborough, R. (**2006**). "Lexical and contextual predictability: Confluent effects on the production of vowels," presentation at the Tenth Conference on Laboratory Phonology, Paris, France.

Smiljanic, R., and Bradlow, A. R. (**2005**). "Production and perception of clear speech in Croatian and English," J. Acoust. Soc. Am. **118**, 1677–1688.

Smiljanic, R., and Bradlow, A. R. (**2007**). "Stability of temporal contrasts across speaking styles in English and Croatian," J. Phonetics.

Sommers, M. S. (**1996**). "The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition," Psychol. Aging **11**, 333–341.

Sommers, M. S. (**1997**). "Stimulus variability and spoken word recognition. II. The effects of age and hearing impairment," J. Acoust. Soc. Am. **101**, 2279–2288.

Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Takata, Y., and Nábělek, A. K. (**1990**). "English consonant recognition in noise and in reverberation by Japanese and American listeners," J. Acoust. Soc. Am. **88**, 663–666.

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., and Durlach, N. I. (**1996**). "Speaking clearly for the hard of hearing, IV. Further studies of the role of speaking rate," J. Speech Hear. Res. **39**, 494–509.

Uchanski, R. M. (**2005**). "Clear speech," in *Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell, Cambridge, MA).

Wright, R. (**2002**). "Factors of lexical competition in vowel articulation," in *Papers in Laboratory Phonology VI*, edited by J. Local, R. Ogden, and R. Temple (Cambridge University Press, Cambridge), pp. 26–50.