

Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors

Ann R. Bradlow

Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, Illinois 60208

David B. Pisoni

Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, Indiana 47405

(Received 3 August 1998; revised 26 March 1999; accepted 17 June 1999)

In order to gain insight into the interplay between the talker-, listener-, and item-related factors that influence speech perception, a large multi-talker database of digitally recorded spoken words was developed, and was then submitted to intelligibility tests with multiple listeners. Ten talkers produced two lists of words at three speaking rates. One list contained lexically “easy” words (words with few phonetically similar sounding “neighbors” with which they could be confused), and the other list contained lexically “hard” words (words with many phonetically similar sounding “neighbors”). An analysis of the intelligibility data obtained with native speakers of English (experiment 1) showed a strong effect of lexical similarity. Easy words had higher intelligibility scores than hard words. A strong effect of speaking rate was also found whereby slow and medium rate words had higher intelligibility scores than fast rate words. Finally, a relationship was also observed between the various stimulus factors whereby the perceptual difficulties imposed by one factor, such as a hard word spoken at a fast rate, could be overcome by the advantage gained through the listener’s experience and familiarity with the speech of a particular talker. In experiment 2, the investigation was extended to another listener population, namely, non-native listeners. Results showed that the ability to take advantage of surface phonetic information, such as a consistent talker across items, is a perceptual skill that transfers easily from first to second language perception. However, non-native listeners had particular difficulty with lexically hard words even when familiarity with the items was controlled, suggesting that non-native word recognition may be compromised when fine phonetic discrimination at the segmental level is required. Taken together, the results of this study provide insight into the signal-dependent and signal-independent factors that influence spoken language processing in native and non-native listeners. © 1999 Acoustical Society of America. [S0001-4966(99)02410-8]

PACS numbers: 43.71.Es, 43.71.Hw [JMH]

INTRODUCTION

Speech perception and spoken word recognition accuracy depend on a wide range of talker-, listener-, and utterance-related characteristics, all of which can vary across communicative situations. A large and continuously growing body of work has provided us with important new information regarding the way in which talkers modify their speech production and articulation depending on a variety of linguistic and paralinguistic factors. For example, Lindblom (1990) showed how speakers vary their output along a continuum of hyper- and hypo-speech, using hyper-speech to assist a listener under “difficult” listening conditions, and hypo-speech when the talker believes less articulatory precision can be tolerated by the listener. A similar idea has been investigated over the past decade or so in a series of studies that examined the acoustic-phonetic factors that differentiate a “conversational” style of speech from a “clear” style of speech, such as one might use when addressing a person with a hearing loss (Picheny *et al.*, 1985, 1986, 1989; Uchanski *et al.*, 1996). Similarly, under the “Lombard effect,” talkers increase their vocal effort when talking in a noisy environment (Hanley and Steer, 1949; Draegert, 1951; Lane and Tranel, 1971), and adults adopt a hyper-articulated style of

speech when addressing infants (Fernald and Simon, 1984; Fernald *et al.*, 1989; Grieser and Kuhl, 1988; Kuhl *et al.*, 1997). These studies, and many others, have provided a great deal of new information about the way in which individual talkers modify and adjust their articulatory patterns to accommodate situational demands. However, aside from establishing that the “clear” speech style does indeed provide an intelligibility advantage over “conversational” speech (Picheny *et al.*, 1985), considerably less attention has been paid to the direct perceptual consequences, from the listener’s point of view, of different styles of speech (see Summers *et al.*, 1988; Lively *et al.*, 1993). Important questions that remain to be answered are: (1) Which of the clear speech transformations are most effective in aiding speech communication? And (2), how do listeners tune their performance according to communicative and situational demands? In order to develop a more complete understanding of the interplay between the talker-, listener-, and item-related factors that influence speech production and perception, we need to look at how the speech signal varies across a range of conditions, as well as how these variables affect listener performance.

With this overall goal in mind, recent work in our laboratory has focused on some of the factors that contribute to

variability in speech perception at the word and sentence levels. Our general approach stems from a basic view of speech communication as a highly adaptive process on the parts of both the talker and the listener. In carrying out our research, we believe that the use of large multi-talker multi-listener speech databases is essential for gaining a deeper understanding of the stimulus variability that is inherent in real-world speech production and perception.

To date, several factors have been shown to directly influence overall speech intelligibility by native listeners of American English. First, the degree of variability in the stimulus materials has been shown to have a major impact on the listener's speech recognition accuracy. For example, word recognition accuracies decrease and response times increase when listeners are presented with spoken word lists that incorporate a high-degree of stimulus variability due to the presence of multiple talkers and multiple speaking rates, relative to spoken word lists in which such stimulus variability is minimized (Mullennix *et al.*, 1989; Sommers *et al.*, 1994). Second, familiarity on the part of the listener with the talker's voice and articulatory characteristics enhances word recognition accuracy under difficult listening conditions. For example, Nygaard *et al.* (1994) showed recently that listeners were more accurate at identifying novel words in noise when the words were spoken by a talker who they had been trained to identify than when the same words were spoken by a novel talker (see also Nygaard and Pisoni, 1998). Third, the lexical characteristics of the particular words in a stimulus set exert a strong influence on overall intelligibility. Several recent studies have shown that lexically "easy" words (i.e., words with few phonetically similar "neighbors" with which they could be confused) are recognized better than lexically "hard" words (i.e., highly confusable words with many phonetically similar neighbors) (Pisoni *et al.*, 1985; Luce, 1986; Luce *et al.*, 1990; Luce and Pisoni, 1998). Finally, in a first attempt at identifying the talker-specific acoustic-phonetic characteristics that correlate with inter-talker intelligibility differences, Bradlow *et al.* (1996) showed that talkers who exhibited a high-degree of "articulatory precision" in their speech generally had higher overall speech intelligibility scores than talkers who tended to produce more "reduced" speech (see also Wright, 1997). Taken together, these recent studies demonstrate that a range of talker-, listener-, and item-related factors affect the observed variability in overall speech intelligibility.

The present study extends this line of research by investigating the *combined* effects of various talker-, listener-, and item-related characteristics on isolated word recognition. The rationale of this study was that, in order to develop a comprehensive understanding of variability in speech production and perception, we need to directly investigate the ways in which multiple sources of variability operate in combination. Specifically, we hypothesized that perceptual difficulties introduced by one factor might be attenuated or amplified by the presence of another factor. For example, we expected that a relatively high degree of phonetic reduction introduced by a fast speaking rate might be tolerated when a listener becomes familiar with the speech of a particular talker. Conversely, we expected that hard word recognition would be

especially difficult for non-native listeners when there is a mismatch between the native and target language phoneme inventories. In order to test these predictions, we conducted two experiments, each of which examined spoken word recognition under conditions that manipulated talker-, listener-, and item-related factors both separately and in combination.

In experiment 1, we used a large database of digital speech recordings to assess the effects of speaking rate, lexical discrimination, and listener-talker adaptation on isolated word intelligibility. By directly examining the separate and combined effects of these characteristics on native-language speech intelligibility, we hoped to gain insight into perceptual processes that underlie native language word recognition. Specifically, we wanted to investigate the separate and combined effects of "signal-dependent" factors, such as speaking rate, and "signal-independent" factors, such as knowledge of the sound-based structure of the lexicon (Lindblom, 1990). Furthermore, the availability of this carefully constructed, multi-talker, multi-listener database provided us with a set of digital speech recordings along with normed intelligibility scores that could then be used in experiments that directly investigate spoken word recognition in a variety of special populations like non-native listeners or listeners with hearing impairments. Accordingly, in experiment 2 we used the same materials as in experiment 1 to investigate stimulus variability and spoken word recognition by non-native listeners. We wanted to see how non-native listeners cope with stimulus variability, and which demographic and linguistic variables correlate with non-native speech intelligibility.

The overall goal of these experiments was to describe in detail, and ultimately to provide a principled account of the relations between the various talker-, listener-, and item-related factors that influence spoken word recognition by both native and non-native listeners. While this was primarily an exploratory study, we believe that this type of fundamental knowledge about the way in which listeners compensate for multiple sources of variability in speech provides insight into the perceptual mechanisms that underlie spoken language processing.

I. EXPERIMENT 1

A. Method

1. The "easy" and "hard" word lists

An "easy" list and a "hard" list of words (75 items each) were constructed such that the two lists differed in terms of three lexical characteristics (Pisoni *et al.*, 1985; Luce, 1986; Luce *et al.*, 1990; Luce and Pisoni, 1998). First, using the word frequency counts provided by the Brown Corpus of printed text (Kucera and Frances, 1967), the words were selected such that the mean word frequency of the easy list was significantly higher than the mean frequency of the words in the hard list (309.7 versus 12.2. per million). Second, using an on-line version of Webster's Pocket Dictionary (20 000 entries) in conjunction with a custom-designed lexical search program, words were selected such that the mean neighborhood density (the number of phonetic "neighbors") of the easy list was lower than the mean neighborhood den-

TABLE I. Descriptive statistics for the ‘‘easy’’ and ‘‘hard’’ word lists. Familiarity and frequency are characteristics of the target word itself. Density is the number of lexical neighbors, and mean neighborhood frequency is the mean frequency of all of these neighbors.

Easy words	Familiarity	Frequency	Density	Mean neighborhood frequency
Mean	6.97	309.69	13.53	38.32
Median	7	106	14	33.3
Standard deviation	0.08	1127.65	4.42	21.87
Minimum	6.5	36	1	2.33
Maximum	7	9816	20	79.67
Range	0.5	9780	19	77.33
Hard words				
Mean	6.81	12.21	26.61	282.23
Median	6.92	3	26	216.48
Standard deviation	0.23	45.85	4.91	215.96
Minimum	6.25	1	11	74.85
Maximum	7	365	39	1066.59
Range	0.75	364	28	991.75

sity of the hard list (13.5 versus 26.6). In these neighborhood density counts, a neighbor of a given target word was defined as any word that differed from the target word by a one phoneme addition, substitution, or deletion in any position (Greenberg and Jenkins, 1964). For example, some of the neighbors of the word ‘‘cat’’ are ‘‘pat, cot, cap, scat, at.’’ Third, the two word lists were constructed such that the mean neighborhood frequency (the mean frequency of the neighbors) of the easy list was much lower than the mean neighborhood frequency of the hard list (38.3 versus 282.2 per million). The net result of these three lexical manipulations was that the easy list consisted of a set of words that occur frequently in the language, and have few phonetically similar, low-frequency neighbors with which they could be confused. In contrast, the hard list consisted of words with many neighbors that are high in frequency relative to the target word. Thus, easy words ‘‘stand out’’ from sparse neighborhoods; hard words are ‘‘swamped’’ by dense neighborhoods. Finally, in order to ensure that subjects would be familiar with all of the words in both lists, all words had been judged as highly familiar by normal-hearing adults, i.e., received a familiarity rating of 6.25 or higher on a 7-point scale where 1 indicated the lowest and 7 indicated the highest degree of familiarity (Nusbaum *et al.*, 1984). Table I provides descriptive statistics for the various lexical characteristics of the words in the two word lists. The items in the two lists of words are provided in the appendix.¹

2. Digital speech recordings

Ten talkers (five males and five females) were recorded producing both the easy and the hard word lists at three different speaking rates (fast, medium, and slow), giving a total of 4500 tokens (150 words \times 3 speaking rates \times 10 talkers). None of the talkers had any known speech or hearing impairments at the time of recording, and all were native speakers of General American English. The talkers were recruited from the Indiana University community and were paid for their participation. All talkers were told in advance that they would be asked to produce three word lists of 150 words each at three different speaking rates. Each individual

talker was allowed to regulate his/her own speaking rate, so long as the three rates were distinct. An analysis of the word durations for each talker at each of the three rates confirmed that each talker successfully produced the three lists with three distinct speaking rates. The mean durations were 809 ms (range 576–1030 ms), 525 ms (range 466–579 ms), and 328 ms (range 264–413 ms) for the slow, medium, and fast words, respectively, confirming that the talkers were successful at producing three distinct rates of speech.

All 150 words (75 easy plus 75 hard) were presented to the talkers in random order on a CRT monitor in a sound-attenuated booth (IAC model 401A). The stimuli were transduced with a Shure (SM98) microphone, and digitized online (16-bit analog-to-digital converter (DSC Model 240) at a 20-kHz sampling rate). The recordings were all live monitored by an experimenter for gross misarticulations and hesitations. Each individual digital file was then edited by hand to remove the silent portions at the beginning and end of each stimulus. The root-mean-square amplitude of each of the digital speech files was then equated. Finally, the files were converted to PC WAV format for presentation to listeners using a PC-based perceptual testing system (Hernandez, 1995).

3. Speech intelligibility tests

Speech intelligibility scores were collected from *independent* groups of ten normal-hearing listeners, each of whom transcribed the full set of 150 words from one talker at one speaking rate, for a total of 30 groups of 10 listeners (10 talkers \times 3 speaking rates). The listeners were all recruited from the Indiana University community and were paid for their participation. None of the listeners reported any prior history of a hearing or speech impairment at the time of testing. The words were presented to the listeners in random order over matched and calibrated Beyer DT-100 headphones via a PC-based perceptual testing system (Hernandez, 1995). The words were presented in the clear (no background noise was added) at a comfortable listening level (70 dB SPL). On each trial, the listeners heard the word and then typed in their response on a computer keyboard. Each

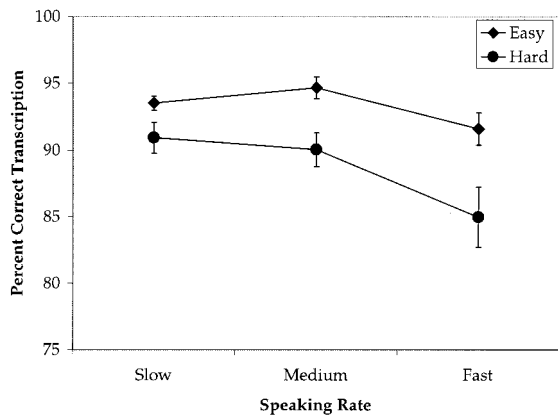


FIG. 1. Mean transcription accuracy scores across all talkers and listeners for the easy and hard words at the slow, medium, and fast speaking rates. The error bars represent the standard error of the mean.

listener received a different randomization of the 150 test words. In the data scoring, a word was counted as correct if all of the letters were present and in the correct order, if all the letters were present but not in the correct order (to allow for obvious typographical errors), or if the transcribed word was a homophone of the intended word.

These transcription scores provided a means of investigating the effects of speaking rate (fast versus medium versus slow) and lexical discrimination (easy versus hard) on isolated word intelligibility. Additionally, since each group of listeners transcribed the full set of 150 words by a single talker at a single rate in a single transcription session, we could also use these intelligibility data to investigate whether listeners adapted to talker-specific characteristics by comparing intelligibility scores from the beginning to scores from the end of the transcription session. We hypothesized that this kind of listener–talker “attunement” on the part of the listener, which occurs over the course of exposure to the speech of a particular talker, mediates the effects of lexical difficulty (easy versus hard) and speaking-rate (fast versus medium versus slow) such that some of the perceptual difficulty introduced by these stimulus factors could be overcome by listener–talker adaptation.

B. Results

Figure 1 shows the overall percent correct transcription scores across all talkers and listeners for the easy and hard word lists at each of the three speaking rates. As expected based on earlier investigations of the effects of these lexical characteristics on speech perception (Pisoni *et al.*, 1985; Luce, 1986; Luce *et al.*, 1990; Luce and Pisoni, 1998), the easy word lists were consistently transcribed more accurately than the hard word lists. As shown in Table II, the higher transcription accuracy for the easy list relative to the hard list held true for most of the talkers at all three speaking rates. The exceptions were for talkers 2 and 7 at the slow rate and for talker 7 at the medium rate, where there was a very small advantage for the hard word list. Thus, the word identification advantage for easy words over hard words is a highly robust effect that generalizes across multiple talkers and speaking rates. The critical difference between easy and hard

TABLE II. Mean intelligibility scores across all ten listeners for the easy and hard word lists by each talker at each speaking rate.

Talker	Easy			Hard		
	Slow	Medium	Fast	Slow	Medium	Fast
1	91.07	92.40	86.13	82.67	81.20	72.27
2	94.40	95.47	94.27	94.80	94.40	89.33
3	94.67	94.00	94.93	88.93	89.60	92.53
4	92.40	96.00	88.27	88.67	87.20	78.00
5	94.00	94.40	86.27	89.47	91.33	75.47
6	92.93	93.87	91.87	92.80	90.40	89.73
7	90.67	89.20	89.47	91.07	90.26	87.87
8	94.93	96.27	92.93	93.60	88.40	89.47
9	95.07	96.67	95.73	92.40	92.13	84.40
10	95.07	98.40	96.27	94.93	95.46	90.67
Mean	93.52	94.67	91.61	90.93	90.04	84.97

words is that hard words require the listener to discriminate between a large set of competitors. In other words, in order to recognize a hard word correctly, the listener must make fine phonetic discriminations between words at the segmental level. The fact that this lexical competition effect is observed even under highly favorable listening conditions suggests that the ability to make fine phonetic discriminations is a skill that is prone to disruption, and as such is likely to be affected even more when conditions are less than favorable such as in the case of non-native listeners, noisy listening environments, or a hearing impairment.

Figure 1 also shows a substantial decline in transcription accuracy for the fast rate relative to the medium and slow rates for both the easy and the hard word lists; however, there was no intelligibility advantage for the slow rate over the medium rate. This pattern of results was somewhat surprising in view of the fact that, on average, the slow words were about 54% longer in duration than the medium words (see also Torretta, 1995). Thus, it appears that isolated word intelligibility is not enhanced by slowing the speaking rate. However, the absence of any difference may have been due to a ceiling effect for word intelligibility in quiet listening conditions.

These initial observations were all confirmed by a repeated-measures ANOVA (nested design) on the arcsine transformed data (Studebaker, 1985) with both speaking rate (fast, medium, slow) and lexical discrimination (easy, hard) as within subject variables, and the intelligibility scores for each talker in each condition averaged across all ten listeners as the dependent variable (see Table II). There was a main effect of speaking rate [$F(2,18) = 11.127$, $p < 0.001$], and a main effect of lexical discrimination [$F(1,18) = 28.494$, $p < 0.001$]. There was also a significant speaking rate by lexical discrimination interaction [$F(2,18) = 5.862$, $p = 0.011$], due to the increasing intelligibility difference between easy and hard words as the speaking rate increases. An examination of the paired contrasts showed a significant difference (at the $p < 0.005$ level) between the fast and medium rates for both the easy and the hard words. There was no difference between the medium and slow rates for the hard words, whereas for the easy words there was a small but significant ($p = 0.038$) advantage for the medium rate over the slow rate. Furthermore, at all three rates, the easy versus hard

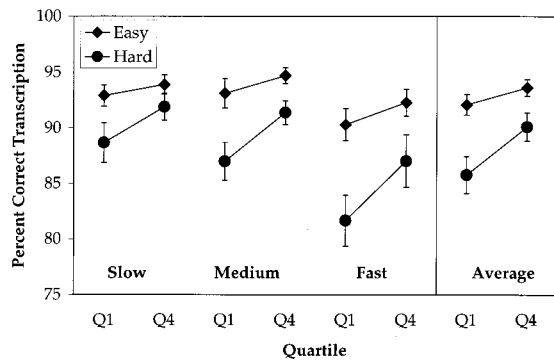


FIG. 2. Mean transcription accuracy scores across all talkers and listeners for the easy and hard words in the first and fourth quartiles at the slow, medium, and fast speaking rates, and averaged across all three speaking rates. The error bars represent the standard error of the mean.

difference was significant at the $p < 0.005$ level.

The words in the easy and hard lists in this database were selected so that the effect of lexical difficulty could be assessed across the lists. In other words, the easy-hard difference across lists is largely categorical, rather than gradient. However, as shown in Table I, there is some degree of intralist variability in lexical difficulty.² Thus, we were able to perform correlational analyses on the various lexical characteristics and word intelligibility across the entire set of 150 words. Results showed a significant negative correlation between neighborhood density and intelligibility at all three speaking rates (slow: $r = -0.213$, $p < 0.01$; medium: $r = -0.356$, $p < 0.0001$; fast: $r = -0.360$, $p < 0.0001$). Furthermore, using a measure of target word “prominence,” which we defined as mean neighborhood frequency minus target word frequency, we found a trend towards a negative correlation between prominence and intelligibility at the medium and fast speaking rates (medium: $r = -0.143$, $p = 0.08$; fast: $r = -0.155$, $p = 0.06$). These results provide additional support for the fundamental assumptions of the neighborhood activation model of spoken word recognition, specifically, the assumption that spoken words are recognized relationally in the context of other phonetically similar words in the mental lexicon (Luce and Pisoni, 1998).

The final step in our analysis of these intelligibility data was to investigate whether isolated word intelligibility improves as the listener becomes accustomed to the talker’s voice. In particular, we wondered whether hard words that were presented later in a transcription session would be more accurately transcribed than hard words presented earlier in the session. We were interested in whether listener-talker adaptation might compensate for the processing difficulties introduced by the lexical discrimination factor.

Figure 2 shows the percent correct transcription scores for the easy and hard words in the first quartile (Q1) and fourth quartile (Q4) of the transcription sessions at each of the three speaking rates as well as across all three rates. In each case, the first and fourth quartiles were taken as the first and last 38 words presented to the listeners, respectively. Because each listener received a different randomization of the 150 words, differences due to particular items were controlled for over the entire group of listeners. As shown in Fig.

TABLE III. Mean intelligibility scores for each speaking rate in the first and fourth quartile.

	First quartile	Fourth quartile
slow	90.80	92.90
medium	90.05	93.04
fast	85.98	89.67

2, hard words presented in the last quartile were generally more accurately transcribed than hard words presented in the first quartile across all three speaking rates. In contrast, there was no noticeable difference between easy words presented in the first and fourth quartiles at all three speaking rates, a finding that may be due to a “ceiling” effect for easy words.

An ANOVA on the arcsine transformed data (Studebaker, 1985) for the intelligibility scores averaged across all three rates showed the expected main effect of lexical category [$F(1,9) = 27.826$, $p < 0.005$]. There was also a main effect of quartile [$F(1,9) = 22.648$, $p = 0.001$], indicating that the Q4 intelligibility scores were significantly higher than the Q1 intelligibility scores. Furthermore, there was a significant quartile by lexical category interaction [$F(1,9) = 8.344$, $p = 0.018$], due to the greater Q4-Q1 difference for the hard words than for the easy words. Interestingly, a pairwise comparison showed a nonsignificant difference between the easy words in the first quartile and the hard words in the fourth quartile. Separate ANOVA’s on the arcsine transformed data for each speaking rate showed that for all three rates there was a main effect of quartile, such that the Q4 intelligibility scores were consistently higher than the Q1 intelligibility scores [slow: $F(1,9) = 9.298$, $p = 0.014$; medium: $F(1,9) = 12.166$, $p < 0.007$; fast: $F(1,9) = 19.322$, $p < 0.002$]. There was also a main effect of lexical discrimination, such that easy words had higher intelligibility scores than hard words [slow: $F(1,9) = 7.301$, $p = 0.024$; medium: $F(1,9) = 19.937$, $p < 0.002$; fast: $F(1,9) = 22.538$, $p < 0.001$]. Furthermore, there was a tendency towards a quartile by lexical category interaction for the medium and fast rates [slow: $F(1,9) = 1.270$, $p = 0.289$; medium: $F(1,9) = 5.074$, $p = 0.051$; fast: $F(1,9) = 3.857$, $p = 0.081$].

These data on the time-course of word recognition indicate that as the listener becomes accustomed to the talker’s voice and specific articulatory patterns, the intelligibility difficulty introduced by the lexical characteristics of hard words can be overcome to a large extent. Furthermore, a comparison of the first and fourth quartile intelligibility scores across the three speaking rates (see Table III) showed that the intelligibility of fast rate words in the fourth quartile (mean = 89.67%) approached the intelligibility scores for the slow and medium rate words in the first quartile (means = 90.80% and 90.05%, respectively). In other words, the listener’s experience with the talker’s speech tended to compensate for the intelligibility difficulty introduced by the fast speaking rate. In general, this pattern of results suggests that listener-talker adaptation and attunement are important factors in speech perception that combine with other talker- and item-related factors, such as speaking rate and lexical

discrimination, in determining the overall intelligibility of normal speech by normal listeners.

C. Summary and discussion

The primary goal of this initial experiment was to examine the combined effects of various talker-, item-, and listener-related factors on spoken word recognition by native listeners by using a carefully constructed multi-talker, multi-listener speech database. Results showed that overall word intelligibility was adversely affected by lexical discriminability: easy words had higher overall intelligibility than hard words. This effect of lexical discrimination was a listener-related factor that results from knowledge on the part of the listener regarding the sound-based structure of the lexicon of the language. We also observed a decline in overall intelligibility for the fast speaking rate: slow and medium rate words both had higher overall intelligibility scores than fast rate words. This speaking rate effect was a signal-related factor that presumably results from acoustic-phonetic adjustments on the part of the talker when he or she is required to consciously adjust speaking rate. We also observed a relationship between the various factors whereby the difficulties imposed by one factor, such as a fast speaking rate or an inherently difficult lexical item, could be overcome by the advantage gained through the listener's experience with the speech of a particular talker. Taken together, these data demonstrate that speech intelligibility is subject to a multitude of highly dynamic variables that have their basis in specific talker-, item-, and listener-related factors. These findings underscore the view of speech communication as an adaptive process from both the talker's and the listener's points of view. In the next experiment, we extended our investigation of factors affecting recognition of spoken words to another listener population, non-native listeners of English.

II. EXPERIMENT 2

Spoken word recognition by non-native speakers depends on a wide range of skills including novel contrast categorization, the adoption of non-native processing strategies, and vocabulary development in the target language. Current research on non-native speech perception has been dominated by the study of the first of these skills, namely, non-native phoneme perception [e.g., see Strange (1995) and references therein]. The bulk of this research has focused on understanding the effects of the first language phoneme inventory on the ability to discriminate and identify second language phonemic contrasts. The findings have led to the development of several models that account for the different degrees of difficulty associated with the perception and production of different non-native contrasts (Best, 1995; Flege, 1995), and has provided researchers with important information about the effects of linguistic background on speech sound perception and categorization. However, we still do not know to what extent the perception of larger linguistic units by non-native listeners depends on fine-grained phoneme discrimination and identification. Is accurate phoneme categorization a necessary prerequisite for accurate word recognition by non-native listeners? Or, does novel phonemic

contrast perception arise from the ability to recognize word-sized units that contrast minimally with each other in the target language?

A similar issue is central to the study of first language acquisition in children. Current research in infant speech perception and early word learning has suggested that the system of meaningful contrasts develops only after infants have developed the skills to perceive and extract words-sized units from the speech stream. As Jusczyk (1997) notes,

“...it is unlikely that filling in a phonetic inventory is the primary force that drives infants' acquisition of the sound structure of their native language. Rather, the acquisition of phonemic categories and phonemic distinctions falls out of learning to segment and recognize words in the fluent speech of one's native language” (p. 109).

While adult second-language acquisition differs in many respects from infant first-language acquisition, it is likely that the need to recognize words is the primary force behind both processes. According to this point of view, sensitivity to non-native phonemic contrasts develops in response to the addition of new lexical items that reflect the specific contrast in question. While the adult second-language learner has the advantage of mature analytic skills that can aid the perception of phonological features at the segmental level, it is likely that novel phoneme perception can function in a linguistically meaningful manner only once the contrast in question signals a known lexical contrast. In other words, acquiring knowledge of the sound-based structure of the target language lexicon is just as important in non-native speech perception as gaining experience with the structure of the target language phoneme inventory. In order to fully understand non-native speech perception, we need to investigate recognition of word-sized units by non-native listeners using stimulus materials that are well controlled in terms of the sound-based structure of the target language. Accordingly, in experiment 2 we used the stimuli from the multi-talker database developed in experiment 1 to investigate spoken word recognition by non-native listeners.

In particular, we wanted to determine whether non-native listeners of English show the same effect of lexical discriminability as native listeners. Specifically, do non-native listeners have greater difficulty with “hard” words than with “easy” words? This outcome would suggest that non-native listeners develop lexicons of their second language using the same sound-based organizational principles as native listeners. We also wanted to know how non-native listeners perform under conditions of high stimulus variability due to a change in talker across items in a spoken word list. Previous research has shown that native listener word recognition is more accurate when surface characteristics, such as talker-related characteristics, remain consistent across items in a list (Mullennix *et al.*, 1989; Sommers *et al.*, 1994). Furthermore, as we found in experiment 1 above, native listeners show evidence of adaptation and tuning to these talker-related characteristics especially under conditions where word recognition is more difficult (i.e., lexically “hard” words). Thus, as a step towards gaining further in-

sight into the factors affecting recognition of spoken words, we wanted to see how non-native listeners cope with talker variability across items in a list.

Furthermore, in this experiment we assessed both spoken word recognition and written word familiarity. This comparison across these two modalities in adult second-language learners allowed us to look at non-native aural proficiency and non-native lexical development independently of each other. This independent measurement of non-native spoken word recognition and lexical development was particularly important because these two abilities might be confounded in non-native listeners. We know that spoken words are recognized by native listeners in the context of other words and that words requiring fine phonetic discrimination are more difficult to recognize (Luce and Pisoni, 1998). Thus, we might expect that non-native listeners will have particular difficulty with hard words since we know that fine phonetic discrimination of foreign language phonemes is particularly difficult for non-native listeners. However, lexically hard words are defined as words of lower frequency in the language, thus we might expect non-native listeners to be less familiar with hard words than easy words and therefore less likely to recognize them correctly. Thus, in order to understand the interaction of phonetic and lexical effects on non-native word recognition independently of word familiarity, we need to obtain independent measures of spoken word recognition and knowledge of the lexicon of the target language. Accordingly, we obtained both measures in experiment 2.

A. Method

1. Subjects

Two groups of subjects participated in this experiment. The first group, the experimental group, included 20 non-native listeners of English who were recruited from the Indiana University community. They ranged in age from 21 to 33 years, and had studied English for 2 to 18 years. The group included 8 males and 12 females. They came from diverse language backgrounds, with the breakdown as follows: 6 Koreans, 4 Chinese, 3 Russians, 2 Japanese, 2 Spanish, 1 Bengali, 1 Nepali, and 1 Dani. The second group, the control group, included 20 native English listeners. They were also recruited from the Indiana University community, and ranged in age from 20 to 42 years. This group included 6 males and 14 females. All subjects were paid for their participation. None reported any known speech or hearing impairment at the time of testing.

2. Stimuli and procedures

All subjects performed two separate tasks. The first task was a spoken word recognition task in which subjects heard a word over headphones and typed what they heard into a computer keyboard. The stimuli for this task came from the multi-talker database of words that was described in experiment 1 above. Only words from the medium rate set were used in this experiment. Two separate lists of words were compiled. The first list consisted of 78 items produced by a single female talker whose mean intelligibility score for the medium rate words was closest to the average intelligibility

score across all ten talkers. Within this “single-talker” list, half of the words ($n=39$) came from the easy list and half of the words ($n=39$) came from the hard list. The second list consisted of 72 items, half of which were easy ($n=36$) and half of which were hard ($n=36$). The items in this “multiple talker” list were produced by the nine remaining talkers, four females and five males, with each talker producing four of the easy words and four of the hard words. There was no overlap between the items in the two lists. The single- and multiple-talker lists were presented to the listeners binaurally over matched and calibrated (DT-100 Beyer) headphones at a comfortable level (70 dB SPL). The order of list presentation (single-talker versus multiple-talker) was counterbalanced across listeners. Within each list, the words were presented in random order and the listeners were instructed to type the word they heard into the keyboard. Each word was presented only once with no possibility of repetition. However, the experiment was self-paced, allowing the listeners to correct spelling errors or make best guesses when entering their responses on the computer keyboard.

The second task was a word familiarity rating task in which subjects rated their familiarity with a list of English words. In this task, subjects responded to 300 words that were presented in standard American English orthography on a computer monitor. Subjects entered their response by pushing a button on a custom-made 7-button box after the word appeared on the screen. Subjects were instructed to use a 7-point scale where 1 indicated “I have never seen this word,” 4 indicated “I have seen this word but don’t know its meaning,” and 7 indicated “I know this word.” Of the 300 words used in this task, 150 came from the “easy” and “hard” lists used in experiment 1 and in the spoken word recognition task of experiment 2. The remaining 150 words were a subset of words that were taken from a longer list of words that had been used in a previous familiarity rating task with native listeners (Lewellen *et al.*, 1993). Of these, 50 received low ratings, 50 received medium ratings, and 50 received high ratings from the native listeners in this earlier study.

Taken together, the list of 300 words used in the present experiment included all of the words used in the spoken word recognition task plus a set of words known to cover a wide range of familiarity ratings from native listeners. Thus, this list provided us with a measure of the receptive vocabulary size of our non-native subjects relative to native subjects. Furthermore, these familiarity rating data allowed us to assess the extent to which non-native spoken word recognition depends on familiarity with the target word. All subjects performed the familiarity rating task after having completed the spoken word recognition task.

B. Results

1. Spoken word recognition

Figure 3 shows the overall percent correct transcription scores for the easy and hard words for the control subjects (left panel) and for the non-native subjects (middle panel) in the single-talker and multiple-talker conditions, respectively. As expected, the control subjects displayed higher overall

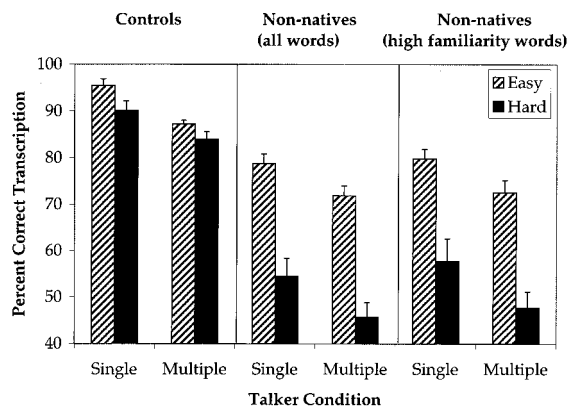


FIG. 3. Mean transcription accuracy scores for the easy and hard words in the single and multiple talker conditions for the control subjects (left panel), the non-native subjects (middle panel), and only the items of high familiarity to the non-native subjects (right panel). The error bars represent the standard error of the mean.

word recognition scores than the non-native listeners. The overall mean and standard deviation for the control subjects were 89.22% and 6.83%, respectively. For the non-native subjects, the mean and standard deviation were 62.73% and 12.24%, respectively. However, both subject groups showed similar patterns of results across the single- and multiple-talker conditions, as well as across the easy and hard words. For both groups, the overall percent correct recognition rate in the multiple-talker condition was lower than in the single-talker condition, indicating that both groups were able to take advantage of the consistent talker information in the single-talker condition. The difference between word recognition accuracy scores in the single- and multiple-talker conditions was 7.2% for the control subjects and 7.9% for the non-native subjects. Additionally, both groups showed higher recognition accuracy scores for the easy than for the hard words. However, there was a strong interaction between subject group and lexical category. Whereas the control subjects showed a difference of 4.3% between easy and hard words, the non-native subjects showed a much larger difference of 25.2%, and this difference was present for both the single- and multiple-talker conditions. The pattern of results displayed in Fig. 3 was confirmed by a three-factor ANOVA on the arcsine transformed data (Studebaker, 1985) with group (non-native, control), talker (single, multiple) and lexical category (easy, hard) as factors. This analysis showed main effects of group [$F(1,38) = 113.234, p < 0.001$], talker [$F(1,38) = 48.085, p < 0.001$], and lexical category [$F(1,38) = 127.146, p < 0.001$]. There was also a significant lexical category \times group interaction [$F(1,38) = 38.861, p < 0.001$]. None of the other interactions was significant.

The significant difference in word recognition performance between the single- and multiple-talker conditions for both groups of subjects suggests that the ability to take advantage of consistent surface information about a particular talker's voice is a skill that transfers easily from first to second language. Conversely, this result suggests that the processing difficulty introduced by a high degree of variability in the stimulus set due to a change in talker from item to item

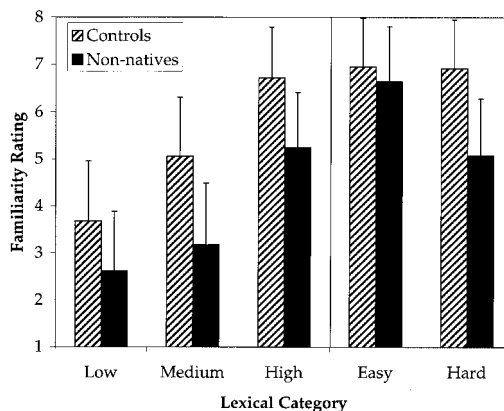


FIG. 4. Mean familiarity ratings for the control and non-native subjects on words of previously determined low, medium, and high familiarity (left panel), and the easy and hard words used in the present word recognition tests (right panel). The error bars represent the standard error of the mean.

is not particularly acute for non-native listeners. Rather, all listeners, regardless of language background, respond similarly to indexical, surface-level variability. The highly significant easy-hard word difference for the non-native listeners suggests that these listeners are developing an English language lexicon with the same sound-based structure as the native English listener lexicon. However, the fact that the non-native listeners showed much lower scores for hard words relative to the control subjects suggests that they have much greater difficulty when fine phonetic discrimination at the segmental level is required by the task. Nevertheless, these non-native subjects appear to be recognizing spoken words relationally in the context of other words they know, although at somewhat lower levels of accuracy relative to native speakers.

2. Familiarity ratings

Figure 4 shows the mean familiarity ratings given by the control and non-native subjects in response to the five word lists used in this task. The words in the low, medium, and high lists shown in the left panel were classified into these three categories based on earlier ratings from a large number of native listeners (Lewellen *et al.*, 1993). The easy and hard lists shown in the right panel contained the same easy and hard words that were presented to the subjects in the spoken word recognition task. For the low, medium, and high lists, the non-native listeners gave substantially lower familiarity ratings than the control subjects. However, both groups showed the expected pattern of increasing familiarity ratings from the low to the medium to the high word lists, suggesting that this task is indeed a valid measure of word familiarity in non-native listeners (see Lewellen *et al.*, 1993).

Of greater interest are the results of the familiarity rating task with the easy and hard words lists. These words were originally selected so that native listeners would be highly familiar with all the test words. This native listener familiarity is indicated in Fig. 4 by the high mean ratings for the control subjects (striped bars) for both the easy (mean rating = 6.9) and hard (mean rating = 6.9) words lists. In contrast, the non-native listeners (black bars) had a high mean

TABLE IV. Correlations between spoken word recognition accuracy, word familiarity ratings, and demographic variables. Numbers in parentheses refer to the range for each variable.

	Word recognition		Familiarity ratings	
	Easy (60%–89%)	Hard (25%–74%)	Easy (3.95–7.00)	Hard (3.69–7.00)
Age of English study onset (4–23 yrs)	+0.09	–0.22	+0.04	–0.61 ^b
No. of years of English study (<1–18)	+0.11	+0.28	+0.17	+0.37
No. of years in English environment (<1–8)	+0.37	+0.45 ^a	+0.04	+0.12

^a $p < 0.05$.

^b $p < 0.005$.

familiarity rating for the easy words (mean rating=6.6), but their ratings for the hard words were much lower (mean rating=5.1). Thus, the pattern of familiarity ratings parallels the pattern of word recognition scores for the non-native subjects, suggesting that part of their difficulty in recognizing hard words may stem from a lack of familiarity with the words themselves rather than from a difficulty with fine phonetic discrimination.

In order to assess the relationship between word familiarity and spoken word recognition performance in the non-native listeners, we reanalyzed the non-native spoken word recognition data by limiting our analysis to only those words that received a familiarity rating of 6 or higher. In this manner, both the non-native and the control subjects' scores reflect word recognition accuracy for words that are judged to be highly familiar to the listeners. The right panel of Fig. 3 shows the non-native subjects' mean word recognition accuracy scores in the single and multiple talker conditions only for the easy and hard words that received a familiarity rating of 6 or higher. On average, across all 20 non-native subjects, 105 of the original 150 words (70%) were included in this analysis. This includes an average of 54/75 (72%) of the easy words and 51/75 (68%) of the hard words.

As shown in Fig. 3, the general pattern of results for the non-natives that we observed for all words (middle panel) is present even when we remove the confounding factor of word familiarity (Fig. 3, right panel). A three-factor ANOVA [on the arcsine transformed data (Studebaker, 1985) for only the high familiarity non-native word recognition scores] with group (non-native, control), talker (single, multiple), and lexical category (easy, hard) as factors showed main effects of group [$F(1,38) = 97.340$, $p < 0.001$], talker [$F(1,38) = 38.760$, $p < 0.001$], and lexical category [$F(1,38) = 72.944$, $p < 0.001$]. There was also a significant lexical category \times group interaction [$F(1,38) = 20.139$, $p < 0.001$]. None of the other interactions were significant.

Thus, while non-native word recognition accuracy may be affected by familiarity with the lexical items, even when we controlled for familiarity, we observed a strong easy–hard lexical effect for these listeners. This pattern suggests that non-native listeners develop second-language mental lexicons that follow the same sound-based structure as the first-language mental lexicon, and that the fine phonetic discrimi-

nation required for accurate recognition of hard words is especially difficult for these listeners.

3. Correlational analyses

In order to further investigate the factors that underlie non-native listener responses to spoken words, we performed a series of correlational analyses between the mean spoken word recognition accuracy scores for each of the 20 non-native subjects and various demographic factors that we obtained from subjects at the start of the data collection sessions. We also performed a similar set of correlational analyses between these demographic variables and the mean familiarity rating score for each of the non-native subjects. In each case, we performed separate correlations for the easy word scores and the hard word scores. Table IV shows the results of these correlational analyses for the variables of greatest interest. For each variable, the numbers in parentheses represent the range of scores across all 20 subjects.

For all dependent variables, none of the correlations with the easy words were significant. This may be because the ranges of word recognition and familiarity rating scores for the easy words were more restricted than for the hard words. There was little or no variance in these measures for the easy words because of ceiling effects in performance. However, for the hard words several interesting correlations emerged. The data showed no correlation between age of onset of English study and hard word recognition; however, number of years in an English environment was significantly positively correlated with hard word recognition scores ($r = +0.45$). In contrast, there was no correlation between hard word familiarity and number of years in an English environment; however, age of onset of English study was significantly negatively correlated with hard word familiarity ($r = -0.61$). Number of years of formal English study was not significantly correlated with either hard word recognition or hard word familiarity. These correlations with the demographic variables suggest that spoken word recognition is an essentially aural skill that requires exposure to spoken language, whereas written vocabulary development is most aided by an early onset of formal second-language study.

C. Summary and discussion

In experiment 2 we investigated some of the characteristics of non-native spoken word recognition as they relate to known characteristics of native spoken word recognition. We found that spoken word recognition by non-native listeners displayed the same overall patterns as for native listeners. Specifically, both groups of listeners recognized words more accurately when all the test words were spoken by the same talker relative to a condition where the talker changed from item to item. This finding suggests that the ability to take advantage of consistent surface phonetic information, such as consistencies in the talker's voice and articulatory patterns, is a language-independent skill that transfers easily from first-language to second-language word recognition.

We also found that both groups of listeners were more accurate at recognizing words that were distinctive or easily discriminated in their lexical neighborhood than those that had many similar sounding neighbors with which they can easily be confused. However, this effect was much more prominent for the non-native listeners, suggesting that these listeners have particular difficulty in recognizing words that require perception of fine phonetic detail for lexical discrimination. This pattern of results was observed even when we controlled for word familiarity across the easy and hard word lists.

Additionally, we found a dissociation between word recognition accuracy and word familiarity ratings with each representing a different skill. Hard word recognition correlated positively with number of years immersed in an English language environment but not with total number of years of English study or age of English study onset, suggesting that hard word recognition may be a good index of non-native aural proficiency independently of vocabulary development. In contrast, hard word familiarity was correlated negatively with age of onset of English study but not with number of years in an English language environment or with total number of years of English study, suggesting that hard word familiarity may be a good index of non-native lexicon development independently of non-native language aural proficiency.

III. GENERAL DISCUSSION

Taken together, these two perceptual experiments demonstrate various characteristics of word recognition by native and non-native listeners. From a methodological point of view, our results show the utility of a large multi-talker multi-listener digital speech database for investigations into spoken language processing. An important aspect of the database that was developed in the present study was that it included a large number of stimulus items produced by a large number of talkers that were then submitted to intelligibility tests by a large number of listeners. This approach to speech database development—one that always includes both production and perception data—has proved particularly effective as a means of investigating the effects of variability in the speech signal from the points of view of both the talker and the listener. We believe that an important goal of research in spoken language processing is to understand

both the sources of variability in the speech signal, and the effects of this variability on the listener (Stevens, 1996). In order to achieve these goals, researchers will need to devise new ways of investigating the separate and combined effects of various sources of stimulus variability in speech. Our multi-talker multi-listener database approach has proved particularly useful in this regard.

From a theoretical standpoint, the findings of the present study point to several key features of spoken language processing. The data demonstrate that spoken word recognition accuracy depends on a combination of at least three types of factors: (1) signal-related characteristics, such as speaking rate, (2) lexical factors, such as knowledge of the sound-based structure of the mental lexicon, and (3) instance-specific factors, such as the listener's prior experience with the talker's voice and articulatory habits. All three factors combine to determine overall speech intelligibility.

Of particular theoretical interest in this study is the finding that listeners adapt to the demands of the communicative situation in much the same way as talkers do. Just as talkers adapt their speech patterns to match the demands of the communicative situation, so do listeners tune and adjust their speech perception mechanisms to take advantage of surface level or paralinguistic consistencies in the signal (see also Nygaard *et al.*, 1994; Kakehi, 1992; Nygaard and Pisoni, 1998 for similar findings). This finding raises the basic question of what listeners are learning over the course of exposure to the speech of a particular talker. Recently, Nygaard and Pisoni (1998) suggested two possible mechanisms that underlie this form of perceptual learning. One possibility is that the listener becomes more efficient at performing the operations that map the talker-specific phonetic implementations to their abstract phonemic representations. In other words, the listener becomes well practiced at the specific procedures required to normalize across the particular talker's idiosyncratic phonetic implementation characteristics, in order to arrive at the intended symbolic representation of the speech signal. This view assumes that the linguistic and indexical (i.e., talker-specific) information conveyed by the speech signal are orthogonal, and that the recovery of the linguistic content is aided by more efficient separation of the linguistic and indexical aspects of the speech signal. The other possibility considered by Nygaard and Pisoni (1998) is that the linguistic and indexical aspects of the signal are integral. According to this view, the talker-specific indexical information and the linguistic content of a signal are carried by the same kinds of time-varying acoustic characteristics. Thus, a high degree of sensitivity to the talker-specific indexical aspects implies an equally high degree of sensitivity to the linguistic aspects of the signal. Consequently, talker familiarity and enhanced word recognition performance necessarily go together. While the data from the present study do not support either one of these alternatives over the other, our results do demonstrate that this type of sensitivity to consistent surface characteristics across items in a list is a feature of spoken language processing that functions independently of whether the listener is perceiving his or her native language or a foreign language.

The present findings also demonstrate a strong effect of

fine-grained phonetic discrimination on word recognition. Word recognition accuracy was always compromised when fine phonetic discrimination was needed to recognize a word, as in the case of hard words spoken at a fast rate for native listeners (experiment 1) or hard words spoken at a medium rate for non-native listeners (experiment 2). A fast speaking rate results in reduced acoustic-phonetic cues in the signal. Similarly, non-native listeners have reduced sensitivity to crucial acoustic-phonetic cues due to their lack of experience with speech in the target language. Thus, when fine-grained acoustic-phonetic discrimination is reduced, whether due to signal-related factors or listener-related factors, word recognition accuracy is reduced accordingly.

This finding suggests that while listeners may be primarily motivated to recognize word-sized units (Jusczyk, 1997), their ability to access lexical items is limited by the degree of low-level acoustic-phonetic detail that is available from the signal. In other words, spoken language processing relies on both accurate phoneme categorization and knowledge of the sound structure of the target language (Luce and Pisoni, 1998). Any attempt to enhance speech intelligibility for non-native listeners or for native listeners under difficult listening conditions due to hearing loss or environmental noise should consider both the degree of acoustic-phonetic detail available in the signal and the phonological and lexical nature of the stimulus materials to be recognized. Depending on various factors, such as those explored in this study, more or less acoustic-phonetic reduction may be tolerated without significant loss of intelligibility.

Consistent with this view, we might predict that, since talkers are presumably also attuned to the sound-based structure of the mental lexicon, they will tend to hyperarticulate hard words. Wright (1997) tested this prediction by performing acoustic analyses of the materials in the same database that we used in the present study. He found that the vowels in the easy words were significantly more centralized (i.e., reduced) than the vowels in the hard words. Nevertheless, as demonstrated by the highly robust easy-hard effect observed in the present study, this hyperarticulation was not sufficient to overcome the effect of lexical difficulty on the part of the listener.

Thus, while both talkers and listeners apparently adapt and modify their performance to the demands of the communicative situation, the effectiveness of each adaptive strategy in enhancing intelligibility can only be judged in relation to other factors that are known to affect speech intelligibility. For this reason, it is critical that speech researchers investigate the separate and combined effects of a wide range of talker-, listener-, and item-related factors on spoken language processing.

ACKNOWLEDGMENTS

We are grateful to Gina Torretta for data collection and processing, and to Luis Hernandez for technical support. We are also grateful to Chris Darwin, James Hillenbrand, John Kingston, and Terrance Nearey for many insightful and helpful comments. This research was supported by NIDCD Training Grant No. DC-00012 and by NIDCD Research Grant No. DC-00111 to Indiana University. Earlier versions

of this work were presented in the Fall of 1997 at the 134th meeting of the Acoustical Society of America in San Diego, CA (2–6 December 1997), and at the International Symposium on Speech Perception by Non-Native Listeners in Boston, MA (19–21 November 1997).

APPENDIX: WORD LISTS

Easy words			Hard words		
was	live	dog	ban	rum	pawn
down	move	vote	bead	sane	bun
work	food	league	bean	soak	gut
long	size	thick	bug	suck	lice
both	cause	page	bum	tan	mid
thought	wrong	hung	chat	wed	wick
does	chief	join	cheer	white	hurl
put	faith	shop	comb	whore	moat
give	pool	roof	cot	wrong	teat
young	deep	leg	den	con	hash
thing	firm	lose	dune	doom	hid
peace	serve	theme	fade	hick	hoot
god	reach	soil	fin	rut	mace
five	mouth	pull	goat	toot	main
gave	teeth	chain	knob	wad	moan
death	gas	curve	lad	bud	mum
shall	jack	path	mall	dame	rim
real	check	dirt	mat	lace	rout
south	king	vice	mitt	lame	wail
job	shape	rough	mole	pad	hum
love	learn	fool	pat	chore	sill
full	ship	noise	pet	cod	beak
wife	neck	wash	pup	hack	hag
voice	watch	balm	rat	kin	wade
girl	judge	fig	rhyme	kit	weed

¹The entire on-line version of Webster's Pocket Dictionary, which includes the lexical characteristics for all of the 20 000 entries in this dictionary, is available in spreadsheet format (Microsoft Excel) from the Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, IN 47405.

²There was also a small amount of overlap between the two lists on all three of the lexical characteristics (namely, frequency, density, and mean neighborhood frequency). In fact, one word ("wrong") appeared in both word lists. While this was a somewhat regrettable situation, removing it from the analysis did not alter the overall intelligibility scores of either list at any speaking rate in a significant way (less than 0.11% difference in all cases).

Best, C. T. (1995). "A direct-realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-language Speech Research*, edited by W. Strange (York, Timonium, MD), pp. 171–206.

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

Dræger, G. L. (1951). "Relationships between voice variables and speech intelligibility in high level noise," *Speech Monogr.* **18**, 272–278.

Fernald, A., and Simon, T. (1984). "Expanded intonation contours in mothers' speech to newborns," *Dev. Psychobiol.* **20**, 104–113.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., and Fukui, I. (1989). "A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants," *J. Child Lang.* **16**, 477–501.

Flege, J. E. (1995). "Second language speech learning: Theory, findings and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Timonium, MD), pp. 233–272.

- Greenberg, J. H., and Jenkins, J. J. (1964). "Studies in the psychological correlates of the sound system of American English," *Word* **20**, 157–177.
- Grieser, D., and Kuhl, P. K. (1988). "Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese," *Dev. Psychobiol.* **24**, 14–20.
- Hanley, T. D., and Steer, M. D. (1949). "Effect of level of distracting noise upon speaking rate, duration and intensity," *J. Speech Hear. Disord.* **14**, 363–368.
- Hernandez, L. R. (1995). "Current computer facilities in the Speech Research Laboratory," *Research on Spoken Language Processing, Progress Report 20*, Indiana University, Bloomington, IN, pp. 389–394.
- Jusczyk, P. W. (1997). *The Discovery of Spoken Language* (MIT, Cambridge, MA).
- Kakehi, K. (1992). "Adaptability to differences between talkers in Japanese monosyllabic perception," in *Speech Perception, Speech Production, and Linguistic Structure*, edited by Y. Tohkura, Y. Sagisaka, and E. Vatikiotis-Bateson (OHM, Tokyo), pp. 135–142.
- Kucera, F., and Francis, W. (1967). *Computational Analysis of Present Day American English* (Brown U. P., Providence, RI).
- Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Koshevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, E. I., and Lacerda, F. (1997). "Cross-language analysis of phonetic units in language addressed to infants," *Science* **277**, 684–686.
- Lane, H. L., and Tranel, B. (1971). "The Lombard sign and the role of hearing in speech," *J. Speech Hear. Res.* **14**, 677–709.
- Lindblom, B. (1990). "Explaining phonetic variation: A sketch of the H & H theory," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, Dordrecht), pp. 403–439.
- Lewellen, M. J., Goldinger, S. D., Pisoni, D. B., and Greene, B. G. (1993). "Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision, and semantic categorization," *J. Exp. Psychol.* **122**, 316–330.
- Lively, S. E., Pisoni, D. B., Summers, W. V., and Bernacki, R. H. (1993). "Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences," *J. Acoust. Soc. Am.* **93**, 2962–2973.
- Luce, P. A., and Pisoni, D. B. (1998). "Recognizing spoken words: The neighborhood activation model," *Ear Hear.* **19**, 1–36.
- Luce, P. A. (1986). "Neighborhoods of words in the mental lexicon," *Research on Speech Perception, Technical Report No. 6*, Indiana University, Bloomington, IN.
- Luce, P. A., Pisoni, D. B., and Goldinger, S. D. (1990). "Similarity neighborhoods of spoken words," in *Cognitive Models of Speech Processing: Psycholinguistics and Computational Perspectives*, edited by G. Altmann (MIT, Cambridge, MA), pp. 122–147.
- Mullennix, J. W., Pisoni, D. B., and Martin, C. S. (1989). "Some effects of talker variability on spoken word recognition," *J. Acoust. Soc. Am.* **85**, 365–378.
- Nusbaum, H. C., Pisoni, D. B., and Davis, C. K. (1984). "Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words," *Research in Speech Perception, Progress Report 10*, Indiana University, Bloomington, IN, pp. 357–376.
- Nygaard, L. C., and Pisoni, D. B. (1998). "Talker-specific learning in speech perception," *Percept. Psychophys.* **60**, 335–376.
- Nygaard, L. C., Sommers, M. C., and Pisoni, D. B. (1994). "Speech perception as a talker-contingent process," *Psychol. Sci.* **5**, 42–46.
- Picheny, M. A., Durlach, N. I., and Braidia, L. D. (1985). "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech," *J. Speech Hear. Res.* **28**, 96–103.
- Picheny, M. A., Durlach, N. I., and Braidia, 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 Braidia, L. D. (1989). "Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to difference in intelligibility between clear and conversational speech," *J. Speech Hear. Res.* **32**, 600–603.
- Pisoni, D. B., Nusbaum, H. C., Luce, P. A., and Slowiaczek, L. M. (1985). "Speech perception, word recognition and the structure of the lexicon," *Speech Commun.* **4**, 75–95.
- Sommers, M. S., Nygaard, L. C., and Pisoni, D. B. (1994). "Stimulus variability and spoken word recognition: I. Effects of variability in speaking rate and overall amplitude," *J. Acoust. Soc. Am.* **96**, 1314–1324.
- Stevens, K. N. (1996). "Understanding variability in speech: A requisite for advances in speech synthesis and recognition," *J. Acoust. Soc. Am.* **100**, 2634.
- Strange, W. (editor) (1995). *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (York, Timonium, MD).
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.
- Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (1988). "Effects of noise on speech production: Acoustic and perceptual analyses," *J. Acoust. Soc. Am.* **84**, 917–928.
- Torretta, G. M. (1995). "The 'easy-hard' word multi-talker speech database: An initial report," *Research on Spoken Language Processing, Progress Report 20*, Indiana University, Bloomington, IN, pp. 321–334.
- Uchanski, R. M., Choi, S., Braidia, 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.
- Wright, R. (1997). "Lexical competition and reduction in speech: A preliminary report," *Research on Spoken Language Processing, Progress Report 21*, Indiana University, Bloomington, IN, pp. 471–486.