Available online at www.sciencedirect.com

SCIENCE DIRECT®

Speech Communication 47 (2005) 424–435

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Lexical boosting of noise-band speech in open- and closed-set formats

J.D. Trout

Loyola University Chicago, Parmly Hearing Institute, 6525 North Sheridan Road, Chicago, IL 60626, United States

Received 20 March 2004; received in revised form 22 April 2005; accepted 26 April 2005

Abstract

Recognition of spoken words in noise and in quiet is more accurate for Lexically Easy words (high frequency words with few similar-sounding neighbors) than for Lexically Hard words (low frequency words with many similar-sounding neighbors). Using monosyllables, the present set of two experiments extends this finding to a perceptually interesting class of stimuli and test formats. In both open- and closed-sets, listeners attempted to identify amplitude-modulated and bandpass-filtered words [Shannon, R., Zeng, F., Kamath, V., Wgonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270, 303–304]—noise-band speech—shown to simulate the performance of cochlear implant (CI) patients using the same number of frequency channels. The words were synthesized from a database that controls for Lexical Difficulty, Talker Identity and Talker Gender. Word recognition was significantly more accurate for Easy words in both the open- and the closed-set experiments. These results indicate that, even when spoken word recognition is challenged by noise-band speech, the Easy–Hard effect survives the perceptually uncertain conditions of word variability. Consequences for models of spoken word recognition are explored.

Keywords: Spoken word recognition; Lexical boosting; Noise-band speech; Monosyllables; Open- and closed-set task

1. Introduction

Previous research on spoken word recognition has demonstrated that some words are "easy" to recognize and others "hard," depending upon a number of lexical properties and the structural relations among them; specifically, a word's lexical frequency, familiarity, neighborhood density, and neighborhood frequency have a potent influence on its intelligibility across a variety of listener populations and listening conditions (Luce, 1986; Torretta, 1995; Bradlow and Pisoni, 1999; Luce and Pisoni, 1998; Dirks et al., 2001; Dorman et al., 2000b; Sommers et al., 1997). According to nearly all models of spoken word recognition

E-mail address: jtrout@luc.edu

0167-6393/$ - see front matter
doi:10.1016/j.specom.2005.04.011

(Forster, 1976, 1979; Marslen-Wilson and Welch, 1978; Paap et al., 1982), lexical identification proceeds by discriminating among lexical items in memory that are activated by stimulus input, and as expected, easy words have more discriminable characteristics and are thus less confusable, than hard words.

A number of lexical characteristics combine to make words "Easy" or "Hard": frequency, familiarity, density, and neighborhood frequency. *Frequency* is the average number of times a word occurs in printed text per million words (Kucera and Francis, 1967). *Familiarity* (Nusbaum et al., 1984), a subjective rating, is the reported familiarity of a word (7.0 is "very familiar"). *Density* (Luce, 1986) is the number of "neighbors" a target word has that differ by one phoneme from it. *Neighborhood frequency* (Luce, 1986) is the mean frequency of words in the lexical neighborhood.

Forged by these four factors, easy words stand out, and hard words are buried, as can be seen in Fig. 1. These conditions determine the ultimate accessibility of those words.

The listener populations in which this Easy–Hard effect has been tested include normal listeners (Luce and Pisoni, 1998; Bradlow and Pisoni, 1999), cochlear implant (CI) patients (Dirks et al., 2001), and listeners with uncorrected hearing loss (Sommers et al., 1997). The listening conditions have included a variety of kinds of noise, and test formats. Luce and Pisoni (1998) used broadband noise to establish that Easy words were

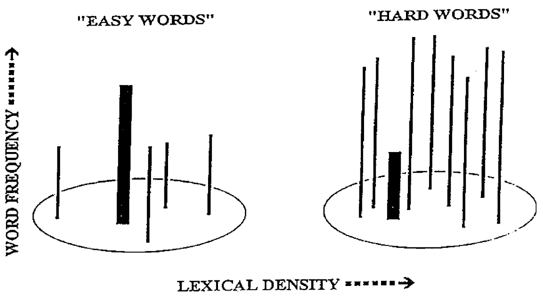

Fig. 1. The Neighborhood Activation Model (Luce, 1986): Word frequency, familiarity, and lexical density are characteristics of "lexical neighborhoods" that influence spoken word recognition (figure from Torretta (1995)).

more intelligible than Hard words under a variety of signal-to-noise ratios. Sommers et al. (1997) used Modified Rhyme Test words selected to meet certain requirements for lexical ease, masked by broadband noise at different signal-to-noise ratios presented to normal listeners, and in quiet to CI listeners in both open- and closed-set formats. In a series of interesting experiments, Dirks et al. (2001) replicated the Easy–Hard effect in an open-set format using a "speech-shaped noise" designed to follow the envelope of the entire amplitude spectrum for the duration of the natural speech that it is used to mask. Dirks et al. (2001) used a different database than the Easy–Hard Multi-Talker Speech Database used here (and separate groups of young normal hearing listeners and elderly listeners with sensorineural hearing loss), examined the separate contributions of neighborhood structure, such as neighborhood density, and did so without a closed-set task.

The present research employs amplitude-modulated and bandpass-filtered noise, or "noise-band" speech. Noise-band speech has special characteristics. It is designed to simulate the noisy and stratified output of a cochlear implant; portions of this output retain the perceptual cues of voicing and manner (Van Tasell et al., 1987; Shannon et al., 1995) by simulating the excitation of specific regions of the basilar membrane; large portions of the spectrum are eliminated, and those that remain are band-limited noise, modulated at the mean amplitude of that region of the spectrum. By comparing simulation data to performance of cochlear implant patients, recent research has established that noise-band "speech" simulates performance using a cochlear implant (Dorman and Loizou, 1998; Dorman et al., 1997; Fishman et al., 1997). Unlike sinewave speech (Remez et al., 1981), which specifically synthesizes spectral features associated with formant structure, the frequency bands of noise-band speech can be arbitrarily distributed, and so are not selected specifically to track such perceptually important features as formants. Finally, the natural speech masked by broadband noise still retains information about natural formant features whose excursions move continuously across the range of the spectrum; by contrast, the bands in noise-band speech are stratified, cutting uniformly through the spectrum of natural speech. Thus, the generation of noise-band speech is insensitive to the continuous movement of formants and their trajectories, eliminating an important source of recognition available in speech in quiet or masked by broadband noise.

Noise-band speech has been widely presented to normal-hearing subjects, but only once using words selected for their Easy and Hard characteristics. In a study using a design, word set, and training regimen different from the one used here, Dorman et al. (2000a) found an Easy–Hard effect for multisyllable noise-band words presented in isolation to normal-hearing children, and natural multisyllabic words presented to children with cochlear implants. At the same time, Dorman et al. (2000a) did not perform a closed-set test, and so could not trace the processing advantages of limiting response alternatives. Although a number of different populations (CI patients, deaf subjects, and normal hearing subjects) have been presented with Easy vs. Hard words, and recognition performance on noise-band speech have been tested in both open- and closed-set formats, the unique combination of factors in the present experiments provide new information about features of the lexicon by examining its reaction to specific processing demands: Easy–Hard, noise-band monosyllables, presented to normal adult listeners in open- and closed-set test formats. It is important to note that, by combining these factors we may undertake a novel investigation of the mechanisms that underlie spoken word recognition. First, it is a central question in speech perception which aspects of the spectrum get encoded, and which features can enhance recognition processes. Noise-band speech is well-adapted to address this question; it consists of isolated bandwidths distributed across the frequency spectrum and consequently would provide insight into the extent to which linguistic representations can be evoked by the linguistic information preserved in these bands. Second, this experimental arrangement allows us to examine how the chronic activation and general accessibility of Easy words compensates for specific limitations on signal access imposed by amplitude-modulation and bandpass-filtering. The simultaneous manipulation of test format allows an additional means of examining the influence of distinct lexical processing demands.

The first goal of the present experiments, then, is to investigate the extent to which the Easy–Hard effect survives the severe and calculated obstacles to signal access posed by "noise-band" monosyllables—when presented to normal listeners. Monosyllables have unique and interesting lexical characteristics that fortunately limit the sources of information that could explain performance on noise-band tokens. Monosyllables as a class have shorter duration, and so are challenging in recognition paradigms, but they also lack the intra-lexical prosodic cues of alternating rhythm that might otherwise assist recognition, and thus complicate the interpretation of recognition performance. If the Easy words are more intelligible than Hard words, then we may conclude that the Easy word advantage is not restricted to words whose structure is fully present, even if masked, in the stimulus. However, if there is no difference in intelligibility, we may conclude that the Easy word advantage is insufficient to compensate for the signal access obstacle presented by amplitude-modulated and bandpass-filtered words. In either case, the results would be of interest because they would provide information about the ability of the mechanisms normally responsible for the intelligibility advantages of Easy words to still process that lexical information once subjected to noise-band synthesis. An intelligibility advantage for Easy words in this context would make at least two important contributions to speech research, one theoretical and one clinical. An Easy–Hard effect promises to illuminate the theoretical issue of the organization of words in the lexicon, and how these sound patterns are accessed from the lexicon by the modulation information carried by five-channel noise-band speech. Together with the finding that a developed lexicon places pediatric cochlear implant patients at a distinct advantage (Pisoni et al., 2000; Pisoni, 2005), any differential processing features of Easy and Hard words might be used to formulate practical, clinical strategies or training materials to improve performance for particular listener populations.

The second question addressed by this set of experiments concerns the extent to which listeners can exploit the Easy word advantage under different processing conditions of open- and closed-set test presentation. Open-set auditory word recognition requires that the listener identifies the target word on the basis of auditory presentation alone which, at least in principle, results in a complete lexical analysis of features and a full search of the lexicon. Closed-set auditory word recognition requires that the listener identify the target word among those on a list of options, a procedure that can culminate in accurate recognition on the basis of an incomplete input analysis and a partial lexical search. If the Easy–Hard effect is present in the open-set format but not the closed, we may conclude that the full analysis prompted by the open-set task recruits word features, such as lexical neighborhood characteristics, available in the Easy words that allow a processing advantage to emerge. By contrast, word recognition occurs in the closed-set task before a complete lexical analysis is carried out. Sommers et al. (1997) demonstrated precisely this fact using broadband noise. The question remains whether, given the unique perceptual characteristics of 5-channel noise-band monosyllables, the Easy–Hard effect disappears in the closed-set task as well, and if not, whether the reduced perceptual analysis in the closed-set is detectable in other measures of processing. The closed-set task does not require an adequate assessment of all of the factors routinely involved in spoken word recognition. The options in a closed-set format are limited, and the correct answer appears among them. The closed-set format tends to reduce the frequency bias across the set of alternatives, because the set of alternatives among which the subject must discriminate is known in advance. If the set of stimuli is relatively unchallenging, the bias may get set at zero; otherwise, the effect is simply attenuated. Simply put, in a closed-set task lexically Hard words have been made more like Easy words by having provided the correct answer in the response alternatives.

In either case, the results would be practically and theoretically interesting. If the Easy–Hard effect is present in the open-set format but not the closed, this would extend Sommers et al.'s finding to a different, theoretically interesting class of stimuli, and thereby advance knowledge of the mechanisms affecting spoken language processing. Minimally, it would show that the Easy–Hard effect is moderated by task demands. If the Easy–Hard effect emerges on both test formats, it would show (by parity of argument) that noise-band tokens preserve the Neighborhood characteristics of natural speech that are exploited by recognition processes. Should there be an Easy–Hard effect in both open- and closed-set formats, the moderating influence of analysis and search processes may be reflected in different effect sizes.

Recognition performance on noise-band speech improves with exposure to sentence and syllable context, limited response alternatives, and practice (van Tasell et al., 1992). In the case of sentences, this advantage has an easy explanation. As Fishman et al. (1997) put the point: "Probably less spectral information (i.e., fewer electrodes) is required when top-down linguistic processing can 'fill in the gaps' in a poor peripheral representation of the signal." (p. 1210) Not surprisingly, then, listeners find recognition of monosyllables in an open-set task (with little practice) very challenging. In order to compensate for the reduced information provided by monosyllables, the listener must either be given more practice, more linguistic (e.g., sentence) context, more channels, or less talker variability. After 8–10 h of participant training, for example, Shannon et al. (1995) observed nearly 100% correct on both open-set sentence recognition and closed-set vowel and consonant recognition with only four bands. Dorman et al. (1997) observed performance of 80–100% on vowels, consonants and sentences with either five sinusoidal or noise band carriers. Loizou et al. (1999) observed more than 90% recognition of sentences in open-set with a 5-band noise processor. Predictably, then, open-set recognition rates for monosyllables should be the lowest among the conditions tested. This prediction for open-set recognition rates is supported by our findings and those reported by others. Friesen et al. (2001) observed about 55% correct for monosyllable words

(CNC) in an open-set task with normal-hearing listeners and a 4-band noise processor. Using monosyllables, Dorman et al. (1997) divided the signals into 6–20 bands, and the listener required at least eight channels in order to achieve at least 90% accuracy in quiet. Noise-band sentences, however, required only four channels of noise band outputs by normal-hearing listeners to achieve 90% accuracy. Loizou et al. (1999) established that difficult sentences produced by multiple talkers require more than five channels for comprehension. For multisyllables, Dorman et al. (2000a) provided 10 sentences and 20 words in practice trials, and listeners achieved about 32% correct for four-band and about 85% for six-band items. In the case of monosyllables, research on noise-band speech has just begun and we do not yet know how differences in training, test set size, database type, etc., will affect performance levels. But current evidence suggests that we can expect the lowest recognition rates for monosyllables in open-set tasks with little practice.

It has long been known in speech perception research that information about amplitude envelope-modulated noise allows the listener to infer lexical information about the word from which the envelope was extracted (Schroeder, 1968; Samuel, 1987; Bashford et al., 1996; Trout and Poser, 1990). In order to systematically examine the separate influences of Difficulty, Talker, and Gender, 5-channel speech envelope noise versions of the medium-rate words in the Easy–Hard database were generated. The resulting stimulus materials were tokens of "noise-band speech," generated in accordance with the methods set out in (Shannon et al., 1995) and described below.

2. Experiment 1: An open-set recognition experiment

The survival of the Easy–Hard effect was investigated under the obstacles to signal access posed by 5-channel "noise-band speech"—amplitude-modulated and bandpass-filtered words—when presented to normal listeners in open-set test formats. The extent of the effect of lexical difficulty, talker, and gender, was evaluated.

2.1. Method

2.1.1. Subjects

Twenty introductory psychology students from Loyola University Chicago were given partial course credit to serve as participants in the experiment. All participants were native speakers of American English, and none reported any history of speech or hearing problems.

2.1.2. Speech material

Two sets of 296 noise-band words were created using the "Easy–Hard Multi-Talker Speech Database," produced by the Speech Research Laboratory at Indiana University. One hundred and forty-eight words were created from each of four talkers. Two of these talkers were male, and two were female.

2.1.3. Signal processing

We created a "noise-band database" of all of the 1500 medium rate words from the "Easy–Hard Multi-Talker Speech Database". The database recorded 10 talkers, five females and five males, uttering the same 150 words. Seventy-five of the words were Hard, and seventy-five were Easy. (For complete details regarding this database, see Torretta, 1995; Bradlow and Pisoni, 1999.) The signal processing scheme produced the noise-band speech as follows. Beginning with normal speech samples within a broad frequency region, such as those of the Easy–Hard Database, natural speech is filtered by five gamma-tone filters. In the present case, each had a Q of 10 centered at 200 Hz, 400 Hz, 800 Hz, 1600 Hz, and 3200 Hz.[1] The envelope of the filtered speech waveform at the output of each filter is extracted by rectifying the filtered waveform and low-pass filtering it at 100 Hz. The same five gamma-tone filters that were used to filter the speech waveform are then used to filter a broadband noise. Each of these filtered noises is then multiplied by the envelope that was extracted from the corresponding filtered speech waveform. In other words, the extracted 200-Hz speech envelope is multiplied by the 200-Hz filtered noise, the

[1] The Q-value of a filter is given by the center frequency of the filter divided by the bandwidth of the filter.

extracted 400-Hz envelope is multiplied by the 400-Hz filtered noise, and so on. Overall, then, this procedure excises the temporal envelopes from filtered speech and inserts band-limited noise modulated by the envelopes in place of the filtered speech. In this manner, the amount of spectral information was severely diminished but temporal and amplitude cues were preserved.

The resulting tokens have a number of unique characteristics. First, because the placement of the bands is not guided by the attempt to capture or track speech information known to be perceptually important (such as formants, or talker-specific information), any level of recognition accuracy will be in spite of the insensitivity of the specific band location to perceptually important speech information. In particular, each band remained in a fixed frequency range, while perceptually important formants in natural speech move in and out of those frequency regions (Van der Horst et al., 1999). Second, these stimuli have only five bands, three fewer than was required for monosyllable performance to match that for sentences (Dorman et al., 2000a).

The test set of Experiments 1 and 2 contained 148 word-types (see Table 1). (Word-types are words categorized independently of the occasion they are spoken or who produces them.)

Talker variability was reduced by presenting only two of the talkers from the original database to any given subject. With each word-type spoken only once by two different talkers (divided by a between-subjects factor of Gender of Talker), each participant heard a total of 296 tokens.

Table 1
The means of lexical characteristics of words in the present Test Set, based on the Easy–Hard Multi-Talker Speech Database

	Easy		Hard	
	Present	Original	Present	Original
Frequency	312.1	309.7	11.1	12.2
Familiarity	7.0	7.0	6.8	6.8
Density	13.5	13.5	26.8	26.6
Neighborhood frequency	37.8	38.8	285.0	282.2

The means of the test items are only slightly different from those in the original database, in which the term 'wrong' occurred as both an Easy and a Hard word. It was deleted from the noise-band database used here, leaving 148 items per talker. The values under Present in this table have been recalculated accordingly.

2.1.4. Procedure

Each subject heard a total of 296 words, presented over headphones at a comfortable listening level of 75 dB. Half of the participants heard 296 words spoken by two female talkers, and half heard those 296 words spoken by two male talkers. The participants heard each word-type twice. Three factors were examined: Difficulty (Easy vs. Hard) × Talker identity (2 Talkers) × Gender (Female vs. Male) in which Difficulty and Talker are within-subjects variables and Gender is a between-subjects variable. Subjects were instructed to type the word they thought they heard. Each subject received six practice trials with feedback, using words that did not occur in the test set.

2.2. Results and discussion

A mixed-design analysis of variance (ANOVA) was conducted on the data, with lexical difficulty and talker identity as within-subjects factors and gender as a between-subjects factor. The Easy–Hard effect was the only significant main effect, [$F(1,18) = 281.30$; $p < .0001$]. There was a significant interaction for Talker × Gender [$F(1,18) = 104.56$; $p < .0001$]. A post-hoc means test indicated that the performance of the two male talkers was elevated compared to that of Female Talker 2. The recognition accuracy means and standard errors for Difficulty, Talker and Gender in Experiment 1 are presented in Table 2.

The overall recognition accuracy in this open-set experiment (9.9%) is lower than that for other noise-band speech studies, explained by the challenging nature of monosyllables in an open-set test with only six practice trials (along with a relatively large, and so variable, word set). However, the theoretical question addressed by the present experiments—whether recognition performance is influenced by Easy–Hard status for noise-band monosyllables under different test formats—does not require replication of performance levels of prior, noncomparable studies, but rather performance above the floor and below a ceiling. The

Table 2
The means and standard errors for accuracy data in percent correct for subjects in Experiment 1

	Recognition accuracy	
	PC	SE
Difficulty		
Easy	14.8	.8
Hard	4.9	.4
Total	9.9	
Talker		
M1	8.0	.8
M2	13.3	1.0
F1	12.0	.8
F2	6.0	1.0
Talker (M1, M2, F1, F2) × Difficulty (E,H)		
M1E	12.8	1.4
M2E	17.3	1.4
F1E	19.1	1.2
F2E	9.9	1.2
M1H	3.2	.7
M2H	6.6	.7
F1H	7.6	.9
F2H	2.2	.9
Gender		
Male	10.0	.6
Female	9.7	.7
Total	9.9	

Note: PC = percent correct; SE = standard error.

recognition performance in the present experiment is well above chance performance.[2] Piloting also attempted to ensure that, using the same test words, performance would not hit a ceiling. Recognition performance in the closed-set task was over 50% on a 10-item list.

In light of this open-set recognition rate, it appears that whatever advantage is conferred by low talker-variability is swamped by substantial word variability. Even at these low performance levels, though, Easy words conferred a significant advantage on spoken word recognition processes. The difficulty of the open-set task does not seem to derive from the processing load added by sensitivity to talker differences; there was no main effect of talker. Moreover, participants only heard talkers of the same gender, and so their results would not reflect sensitivity to gender-based differences in Talker, such as differences in fundamental frequency or in frequency range of formants partially represented in the 5-channel noise-band speech.

An account of word recognition that characterizes effect size as depending on completion of input analysis predicts that the Easy–Hard effect would subside or vanish in closed-set tasks. Because the Neighborhood Activation Model (or NAM; Luce, 1986), for example, assumes that the frequency information that biases decision units is a flexible rather than intrinsic feature of the activation levels of the acoustic–phonetic patterns, the size of the Easy–Hard effect depends upon contextual factors, such as the frequencies of other words. This means that in order to exploit the relevant aspects of neighborhood structure to produce the Easy–Hard effect, a sufficiently complete analysis of the input must be carried out. A closed-set task truncates this analysis and should result in a diminished Easy–Hard effect.

3. Experiment 2: A closed-set recognition experiment

The Easy–Hard effect investigated in the open-set Experiment 1 is here examined in a closed-set format. The extent of the effect of lexical difficulty, talker, and gender, was evaluated.

3.1. Method

3.1.1. Subjects

Twenty introductory psychology students from Loyola University Chicago (different from those in Experiment 1) were given partial course credit to serve as participants in the experiment. All participants were native speakers of American English, and none reported any history of speech or hearing problems.

[2] The lowest recognition rate is for Hard words in the open-set, at 4.9%. In the 125,000 word CMU dictionary, 13.5% of the words are monosyllables, for a total of 16,875 monosyllables. If we restrict the reference set to singular, distinct monosyllables, there are no less than 5000 root monosyllables. Chance performance should yield a 1/5000 identification rate, but the participants in Experiment 1, in the most challenging condition, performed well above that level, at 245/5000. Therefore, a 4.9% recognition rate (for Hard words) on the open-set task is well above a performance floor.

3.1.2. Speech material

The same two sets of 296 auditory stimuli were used in the present experiment as were used in Experiment 1.

3.1.3. Signal processing

The same set of audio stimuli were used in Experiment 2 as was used in Experiment 1.

3.1.4. Procedure

In the closed-set, each trial presented participants with an audio noise-band token over headphones at a comfortable listening level of 75 dB, in the same manner as in Experiment 1. In Experiment 2, however, participants were presented with a video display of 10 different words on each trial simultaneously with the audio stimulus onset. The display on each trial included the correct answer, and subjects were instructed to click the mouse on the menu display word that they thought they heard. The 10 response alternatives were quasi-randomly selected from the Easy–Hard database, with the constraint that no two response alternatives shared an initial phoneme. Each subject received six practice trials with feedback, using words that did not occur in the test set.

3.2. Results and discussion

A mixed-design ANOVA was conducted on the data, with lexical difficulty and talker identity as within-subjects factors and gender as a between-subjects factor. There was a significant main effect of Difficulty [$F(1,18) = 8.98$; $p < .008$]. There were two significant interactions: Talker × Gender [$F(1,18) = 14.97$; $p < .001$], and Difficulty × Talker × Gender [$F(1,18) = 4.87$; $p < .04$]. No other main or interaction effects were significant. Post-hoc tests indicated that the Talker × Gender interaction is largely explained by the poor performance the second female talker when compared to that of both male talkers. The recognition accuracy means and standard deviations for Lexical Difficulty and Gender in Experiment 2 are presented in Table 3.

The moderate overall recognition accuracy in this experiment (51.7%) shows that participants found the closed-set task on this relatively large word set challenging, but not unduly so. With the correct item among the 10 response alternatives, chance performance is 10%.[3] Even at these fair performance levels, Easy words still conferred a significant advantage on spoken word recognition processes. Whether due to the challenging character of 5-channel noise-band speech, the substantial word variability across the test set, or the increased confusability of response alternatives that were all drawn from the test set, the task is sufficiently difficult to permit the benefit of lexical ease to emerge. Table 3 represents the small mean difference in correct performance between Easy and Hard words on the closed-set task (53.2% vs.

Table 3
The means and standard errors for accuracy data in percent correct for subjects in Experiment 2

	Recognition accuracy	
	PC	SE
Difficulty		
Easy	53.2	1.5
Hard	50.2	1.9
Total	51.7	
Talker		
M1	50.1	2.6
M2	55.8	2.7
F1	54.5	2.6
F2	46.5	2.7
Talker (M1, M2, F1, F2) × Difficulty (E,H)		
M1E	52.0	2.3
M2E	56.8	2.8
F1E	55.3	2.3
F2E	48.9	2.8
M1H	48.2	3.0
M2H	54.9	2.9
F1H	53.8	3.0
F2H	44.1	2.9
Gender		
Male	52.3	1.8
Female	51.1	1.9
Total	51.7	

Note: PC = percent correct; SE = standard error.

[3] This is an idealization, because the response alternatives on each trial in the closed-set task were drawn from the 148 word-types in the test set, rather than from the entire lexicon. Although participants were not informed of this fact, implicit learning might raise the outcome of a "guessing" strategy slightly above the chance performance of 10%.

50.2%), compared to the much larger mean difference on the open-set task (14.8% vs. 4.9%). However, as in the open-set Experiment 1, the difficulty of the closed-set task does not seem to derive from the processing load added by implicit accommodation of talker-differences; there was no main effect of talker. Moreover (again as in Experiment 1), participants only heard talkers of the same gender, and so their results would not reflect sensitivity to gender-based differences in Talker, such as differences in fundamental frequency or in frequency range of formants partially represented in the 5-channel noise-band speech.

4. Summary and conclusions

This study was designed to address two issues: the magnitude of an Easy–Hard effect in the face of the severe obstacles to signal access posed by "noise-band speech," and the magnitude of the Easy–Hard effect across the different processing conditions of open- and closed-set test presentation. The findings of this study demonstrate that

(1) There was an Easy–Hard effect for normal listeners of noise-band speech.
(2) The Easy–Hard effect emerged in both open and closed-set test formats, but there was a substantially smaller Easy–Hard effect in the closed-set.

The results of these two experiments identify the influence of lexical difficulty, as well as a number of properties of spoken language settings that attenuate the benefits of Easy words. We will discuss these findings in the order presented above.

These experiments further confirm and extend the findings concerning the benefits of Lexical Ease. The Easy–Hard effect can survive the shift from a difficult open-set task to the easier closed-set task. In the present experiments, while talker-specific information may ease the burden of word recognition, word variability is a powerful barrier to word recognition. Bradlow and Pisoni (1999) uncovered a related, compensatory connection between various sources of processing difficulties and outcomes. There, the processing burden imposed by lexically Hard monosyllables was eased by the listener's experience with the speech of a specific talker. Finally, the property of lexical ease can be used to compensate for substantial perceptual degradation. On the basis of 5-channel noise-band speech—in which participants are only receiving information about the frequency placement of five noisy, amplitude envelope-modulated bands—participants still perform well above chance levels, and still benefit from lexical ease.

The second chief finding of the present set of experiments was that the influence of Difficulty in the closed-set experiment is fragile. The common forces of lexical processing might explain why, although our closed-set experiment departed slightly from the findings of (Sommers et al., 1997) by reaching significance in the subjects-analyses, and noise-masked and noise-band speech have interestingly different characteristics,[4] we did nevertheless find a similar pattern: The effect size was smaller in the closed-set experiments than in their open-set counterparts, indicating that reliance on lexical ease decreases in the presence of other useful cues, such as a visual menu of 10 options in the closed-set test. Generally, this "limited

[4] The lexical frequency figures for their database were not included in (Sommers et al., 1997), so we are unable to isolate this difference as a possible explanation for our divergent results on the respective closed-set experiments. Type of difficulty may also be a factor in explaining this effect-size pattern across open- and closed-set tasks. Test set size, confusability of target and response alternatives, difference between the means of the relevant characteristics of particular classes of Easy and Hard words: All of these factors may (differentially) influence the difficulty of the task. Sommers et al. (1997) reports only the neighborhood frequency and neighborhood density of the class of Modified Rhyme Test words selected, and the present experiments are comparable along these dimensions. However, neighborhood frequency and neighborhood density are not the only characteristics that contribute to lexical ease; lexical (as opposed to neighborhood) frequency has been shown to be a potent contributor to accessibility (Luce and Pisoni, 1998). There is a large mean difference between the lexical frequency of the Easy and Hard words used in the present experiments (Easy: 312.1 vs. Hard: 11.1). This difference alone could explain why the Easy–Hard effect emerges in the closed-set test format of Experiment 2, but whether or not it does explain the effect is unclear. Because Easy–Hard effects depend upon the neighborhood characteristics of the designated samples of words, further experiments are needed to identify the specific contributions of neighborhood characteristics to the effects of lexical difficulty.

processing demand'' account predicts that the Easy–Hard effect will decrease as processing demands decrease. So whether or not the effect disappears altogether will be determined by the perceptual availability of identifying information in each stimulus item. This feature of perceptual availability varies across task (open-set, closed-set, reaction time, error rate), item (phonological characteristics, neighborhood characteristics, etc.), and masker (white, Gaussian noise at +5 dB, speech shaped noise, amplitude-envelope modulated and bandpass-filtered noise, etc.). Our results on the closed-set task differed only slightly from those of Sommers et al. Sommers et al. found that the processing advantage for lexically Easy words vanished on the closed-set task. If this ''limited processing demand'' explanation for the vanishing closed-set effect is correct, then closed-set recognition effects on such tasks should be more fragile than those gleaned from open-set formats. So it is not the character of the closed-set task per se, but the full range of specific processing demands, which determines performance on the closed-set format. The Difficulty effect size (partial Eta squared, η_p^2) in the subjects-analysis for the identification test in the open-set (Experiment 1) is .940, and in the closed-set (Experiment 2) is 0.333. Using the scale defined by Cohen (1988), the open-set effect size for Difficulty is large, while the closed-set d is a medium-to-small effect size.[3]

Insofar as the organization of the emerging lexicon powerfully influences the acquisition of language by children with cochlear implants (Pisoni et al., 2000; Pisoni, 2005), knowledge about the lexical storage of fine, talker-specific details should advance research on the encoded components of speech that could be beneficially exploited in speech and hearing technology. In light of the results concerning cochlear implantation, the performance boosting provided by lexical ease suggests a practical strategy for the construction of training materials for cochlear implant patients, particularly pediatric patients. Shannon et al. (1995) points out that amplitude-modulated and bandpass-filtered speech simulates the stimulation pattern of a cochlear implant. Word-related features may indicate what speech sounds are potentially effective in training cochlear implant users. In particular, the original pediatric cochlear implant protocol was designed for children with acquired deafness, children who already had some exposure to language. Pisoni et al. found that the outstanding pediatric cochlear implant performers, or ''stars'', had more developed mental lexicons to begin with than their cohorts (Pisoni et al., 2000; Pisoni, 2005). It appears that their relatively developed lexicon conferred the expected processing advantages. Accordingly, training materials that exploit lexical advantages may ease the perceptual challenge of early implant use. Easy words afford just such lexical advantages, and it is an open empirical question whether the effects reported here are large enough that the processes underlying them could be recruited in those training efforts.

It still remains to be determined how the manipulation of the number of channels interacts with the influence of lexical ease. As well, we need to better understand the precise durability of such features as Difficulty. The difficulty of a word is not a single, intrinsic part of the activation levels of acoustic–phonetic patterns, and so does not simply act as a standing prime for intelligibility and recognition. But neither is its biasing influence on decision units indefinitely flexible. Rather than a single prior occurrence lowering the threshold of the token, the threshold of that token's node in the lexicon is chronically depressed by its lexical standing. The use of both open- and closed-set test formats on noise-band monosyllables allows us to examine this critical issue for spoken language processing.

The research reported here contributes to the continuing project of understanding the organization and activity of the lexicon. It remains an interesting theoretical question just how flexible are the properties of word units. Can their thresholds change and, if so, at what rate and for what reasons? Can the manipulation of neighborhood parameters—frequency, density, etc.—contribute to this process? These questions remain unanswered. One thing is clear, however. Top-down, lexical information offers potent compensation for degraded bottom-up information. In the fast and noisy commerce of spoken language processing, the words that stand out from the crowd are the ones most easily recognized.

Acknowledgments

The author would like to thank Alex de Miranda and Ryan Haas for their help running the experiments, Megan Deiger and Lauree Garvin for technical assistance, and Ann Bradlow, Richard Fay, Philip Loizou, David Pisoni, Mitchell Sommers, and William Yost, for specific discussion throughout. Special thanks are due to Stan Sheft for his assistance in generating stimuli. This research was supported by Grant PO1-DC00293-15 awarded to the Parmly Hearing Institute by the National Institute on Deafness and Other Communication Disorders.

References

Bashford, J.A., Warren, R.M., Brown, C.A., 1996. Use of speech-modulated noise adds strong "bottom-up" cues for phonemic restoration. Percept. Psychophys. 58, 342–350.

Bradlow, A., Pisoni, D., 1999. Recognition of spoken words by native and non-native listeners: talk-, listener-, and item-related factors. J. Acoust. Soc. Amer. 106, 2074–2085.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences, second ed. Hillsdale, NJ, Lawrence Erlbaum Associates.

Dirks, D.D., Takayanagi, S., Moshfegh, A., Noffsinger, P.D., Fausti, S.A., 2001. Examination of the neighborhood activation theory in normal and hearing-impaired listeners. Ear Hearing 22, 1–13.

Dorman, M., Loizou, P., 1998. Identification of consonants and vowels by cochlear-implant patients using a 6-channel continuous interleaved sample processor and by normal-hearing subjects using simulations processors with two to nine channels. Ear Hearing 19, 162–166.

Dorman, M., Loizou, P., Rainey, D., 1997. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. J. Acoust. Soc. Amer. 102, 2403–2411.

Dorman, M., Loizou, P., Fitzke, J., Tu, Z., 2000a. The recognition of NU-6 words by cochlear implant patients and by normal-hearing subjects listening to NU-6 words processed in the manner of CIS and SPEAK strategies. Ann. Otol. Rhinol. Laryngol. 109 (Suppl. 185), 64–66.

Dorman, M., Loizou, P., Kemp, L., Kirk, K., 2000b. Word recognition by children listening to speech processed into a small number of channels: Data from normal-hearing children and children with cochlear implants. Ear Hearing 21, 590–596.

Fishman, K., Shannon, R.V., Slattery, W.H., 1997. Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. J. Speech Hearing Res. 40, 1201–1215.

Forster, K.I., 1976. Accessing the mental lexicon. In: Wales, R.J., Walker, E. (Eds.), New Approaches to Language Mechanisms. North Holland, Amsterdam.

Forster, K.I., 1979. Levels of processing and the structure of the language processor. In: Cooper, W.E., Walker, E.C.T. (Eds.), Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett. Erlbaum, Hillsdale, NJ.

Friesen, L., Shannon, R., Baskent, D., Wang, X., 2001. Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. J. Acoust. Soc. Amer. 110 (2), 1150–1163.

Kucera, F., Francis, W., 1967. Computational Analysis of Present-day American English. Brown University Press, Providence, RI.

Loizou, P., Dorman, M., Tu, Z., 1999. On the number of channels needed to understand speech. J. Acoust. Soc. Amer. 106, 2097–2103.

Luce, P., 1986. A computational analysis of uniqueness points in auditory word recognition. Percept. Psychophys. 39, 155–158.

Luce, P., Pisoni, D., 1998. Recognizing spoken words: the neighborhood activation model. Ear Hearing 19, 1–36.

Marslen-Wilson, W.D., Welch, A., 1978. Processing interactions and lexical access during word recognition in continuous speech. Cognit. Psychol. 10, 29–63.

Nusbaum, H.C., Pisoni, D.B., Davis, C.K., 1984. Sizing up the Hoosier mental lexicon: measuring the familiarity of 20,000 words. In: Research on Speech Perception Progress Report No. 10. Speech Research Laboratory, Indiana University, Bloomington, IN, pp. 357–376.

Paap, K.R., Newsome, S.L., McDonald, J.E., Schvaneveldt, R.W., 1982. An activation-verification model for letter and word recognition: the word superiority effect. Psychol. Rev. 89, 573–594.

Pisoni, D.B., 2005. Speech perception in deaf children with cochlear implants. In: Pisoni, D.B., Remez, R.E. (Eds.), Handbook of Speech Perception. Blackwell Publishers, pp. 494–523.

Pisoni, D.B., Cleary, M., Geers, A.E., Tobey, E.A., 2000. Individual differences in effectiveness of cochlear implants in prelingually deaf children: Some new process measures of performance. Volta Review 101, 111–164.

Remez, R., Rubin, P., Pisoni, D., Carrell, T., 1981. Speech perception without traditional speech cues. Science 212, 947–950.

Samuel, A.G., 1987. Lexical uniqueness effects on phonemic restoration. J. Memory Language 26, 36–56.

Schroeder, M., 1968. Reference signal for signal quality studies. J. Acoust. Soc. Amer. 44, 1735–1736.

Shannon, R., Zeng, F., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270, 303–304.

Sommers, M., Kirk, K., Pisoni, D., 1997. Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners. I. The effects of response format. Ear Hearing 18, 89–99.

Torretta, G.M., 1995. The easy–hard word multi-talker speech database: an initial report. In: Research on Spoken Language Processing, Progress Report No. 20. Speech Research Laboratory, Indiana University, Bloomington, IN, pp. 321–334.

Trout, J.D., Poser, W.J., 1990. Auditory and visual influences on phonemic restoration. Language Speech 33, 121–135.

Van der Horst, R., Leeuw, A.R., Dreschler, W.A., 1999. Importance of temporal-envelope cues in consonant recognition. J. Acoust. Soc. Amer. 105, 1801–1809.

Van Tasell, D., Soli, S.D., Kirby, V.M., Widin, G.P., 1987. Speech waveform envelope cues for consonant recognition. J. Acoust. Soc. Amer. 82, 1152–1161.

van Tasell, D.J., Greenfield, D.G., Logemann, J.J., Nelson, D.A., 1992. Temporal cues for consonant recognition: training, talker generalization, and use in evaluation of cochlear implants. J. Acoust. Soc. Amer. 92, 1247–1257.