Journal of Experimental Psychology:
Human Perception and Performance
2008, Vol. 34, No. 2, 460–474

0096-1523/08/$12.00 DOI: 10.1037/0096-1523.34.2.460

Perceptual Learning of Noise Vocoded Words: Effects of Feedback and Lexicality

Alexis Hervais-Adelman and Matthew H. Davis
Medical Research Council Cognition and Brain Sciences Unit

Ingrid S. Johnsrude
Queen's University

Robert P. Carlyon
Medical Research Council Cognition and Brain Sciences Unit

Speech comprehension is resistant to acoustic distortion in the input, reflecting listeners' ability to adjust perceptual processes to match the speech input. This adjustment is reflected in improved comprehension of distorted speech with experience. For noise vocoding, a manipulation that removes spectral detail from speech, listeners' word report showed a significantly greater improvement over trials for listeners that heard clear speech presentations before rather than after hearing distorted speech (clear-then-distorted compared with distorted-then-clear feedback, in Experiment 1). This perceptual learning generalized to untrained words suggesting a sublexical locus for learning and was equivalent for word and nonword training stimuli (Experiment 2). These findings point to the crucial involvement of phonological short-term memory and top-down processes in the perceptual learning of noise-vocoded speech. Similar processes may facilitate comprehension of speech in an unfamiliar accent or following cochlear implantation.

Keywords: speech, perception, adaptation, vocoding

Speech comprehension is robust under a great variety of different degraded listening conditions—for example, everyday speech is almost invariably subject to the effects of room acoustics or background noise, and speech transmitted over a telephone, radio, or public address system is degraded but, for the most part, readily comprehensible. The human speech-processing system is also able to take account of wide variations in speech rate (e.g., J. L. Miller & Liberman, 1979), speaker size (Ives, Smith, & Patterson, 2005; Smith, Patterson, Turner, Kawahara, & Irino, 2005), and accent (Clarke & Garrett, 2004; Clopper & Pisoni, 2004; Evans & Iverson, 2004; Weill, 2001), such that acoustically very different forms of the same utterances are perceived as tokens of the same speech sound. Speech perception remains robust even when challenged with extreme and artificial forms of distortion. For example, speech remains comprehensible when formants are resynthesized as sinusoids (Remez, Rubin, Berns, Pardo, & Lang, 1994; Remez, Rubin, Pisoni, & Carrell, 1981); this manipulation removes most of the natural qualities of human voices from the speech signal. Other researchers have shown that dramatic alterations to both the temporal (Mehler et al., 1993; Saberi & Perrott, 1999) and the spectral (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995) properties of speech do not substantially decrease intelligibility, at least in the absence of background noise.

In many of the situations described above, comprehension is poor on initial presentation but improves with continued exposure to distorted speech. This is a form of perceptual learning—"relatively long-lasting changes to an organism's perceptual system that improve its ability to respond to its environment and are caused by this environment" (Goldstone, 1998, p. 586). For instance, comprehension of heavily accented speech improves rapidly with exposure (Clarke & Garrett, 2004; Weill, 2001), and listeners readily learn to compensate for novel shifts in phoneme category boundaries (Kraljic & Samuel, 2006; Norris, McQueen, & Cutler, 2003). A similar perceptual learning process has also been observed with more artificial speech manipulations. For instance, comprehension of poor-quality synthetic speech improves with exposure (Fenn, Nusbaum, & Margoliash, 2003), as does comprehension of artificially distorted recordings of natural speech (such as speech that has been time compressed (Mehler et

Alexis Hervais-Adelman, Matthew H. Davis, and Robert P. Carlyon, Medical Research Council Cognition and Brain Sciences Unit, Cambridge, United Kingdom; Ingrid S. Johnsrude, Department of Psychology, Queen's University, Kingston, Ontario, Canada.

Preliminary reports of these experiments were presented at the British Society of Audiology, Short Papers Meeting, University College London, September 2004; at the Plasticity in Speech Perception Workshop, University of London, June 2005; and at the joint meeting of the Experimental Psychology Society and the Canadian Society for Brain, Behaviour and Cognitive Science in Montreal, July 2005. We acknowledge the financial support of the U.K. Medical Research Council (U.1055.04.013.00001.01), the Canadian Institute for Health Research, and the Canada Research Chairs Programme (to Ingrid S. Johnsrude). We would like to thank our volunteers for their participation, and Eleni Orfanidou and Kadia Acres for providing recordings of the word and nonword stimuli. We thank Arty Samuel and Howard Nusbaum for their comments and suggestions on a previous version of the article.

Correspondence concerning this article should be addressed to Alexis Hervais-Adelman or to Matthew H. Davis, Medical Research Council Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom. E-mail: alexis.hervais-adelman@mrc-cbu.cam.ac.uk or matt.davis@mrc-cbu.cam.ac.uk

al., 1993; Pallier, Sebastian-Galles, Dupoux, Christophe, & Mehler, 1998; Peelle & Wingfield, 2005).

Perceptual Learning of Noise-Vocoded Speech

In the present work, we explore the perceptual learning of noise-vocoded (NV) speech. Noise vocoding is an artificial distortion that removes much of the spectral information from speech while preserving much of the slowly varying temporal cues (Shannon et al., 1995). Noise vocoding is considered to be a simulation of sound transduced by a cochlear implant (CI), and investigators have explored how different parameters of this distortion affect speech intelligibility (Faulkner, Rosen, & Smith, 2000; Loizou, Dorman, & Tu, 1999; Shannon et al., 1995). For instance, variations in the number and spacing of the frequency bands used in creating NV speech affect intelligibility in normally hearing listeners in a manner that resembles the effect of changing the number and placement of electrodes in CI users (Fu & Galvin, 2003; Rosen, Faulkner, & Wilkinson, 1999; Shannon et al., 1995).

Here we build on a recent study that demonstrated robust and rapid perceptual learning for NV speech in normally hearing listeners (Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005). In an initial experiment, Davis and colleagues showed that over the course of a 25-min experiment, word report scores increased from near zero to around 70% correct for NV sentences. This finding suggests a striking form of perceptual learning as report scores were enhanced for sentences containing words that participants had never previously heard in vocoded form. The observation that perceptual learning of NV speech generalized to novel words seems to imply a sublexical locus of learning through the retuning of perceptual representations of phonetic units that are shared over many words.

However, in order to endorse this proposal for a sublexical locus of learning, further evidence that perceptual learning generalizes to novel vocoded words would be valuable. In all the studies reported by Davis et al., perceptual learning was assessed via report scores for NV sentences. As is well known, report scores for sentences can be enhanced by knowledge of semantic and syntactic content (cf. G. A. Miller, Heise, & Lichten, 1951). It is therefore possible that listeners were using some generic knowledge of the typical syntactic or semantic content of earlier sentences to guide their interpretation of later sentences. Subsequent experiments showed that some learning was possible with NV "jabberwocky" sentences (i.e., nonwords combined with English function words) and that training with NV syntactic prose sentences (i.e., syntactically correct sentences that lack sentence-level meaning; e.g., "The effect supposed to the consumer") produced perceptual learning that was equivalent to training with NV normal English. Although these results might argue against learning that is solely due to knowledge of typical sentence content, it remains possible that familiarity with vocoded function words, or syntactic prediction (cf. Grosjean, 1980; Salasoo & Pisoni, 1985), could produce improved report scores for novel words in later sentences without requiring a sublexical locus for learning. In the present study, we therefore assessed changes in the comprehension of NV speech by using isolated words. Should we again observe that perceptual learning generalizes to novel vocoded words, we can be more confident that learning involves changes to sublexical representations.

Davis and colleagues (2005) also conducted a series of studies that explored the role of feedback about sentence content on perceptual learning of NV speech. They showed that improvements in comprehension occur more rapidly if listeners hear distorted sentences when they already know the identity of the original sentence (as a consequence of receiving either clearly spoken or written presentation of each sentence prior to hearing the distorted speech). Perceptual learning was more rapid for listeners who received this "clear-then-distorted" feedback than for listeners who received equivalent exposure to distorted speech but with feedback presented only after distorted speech presentation ("distorted-then-clear" feedback). Similarly, Schwab, Nusbaum, and Pisoni (1985) and Greenspan, Nusbaum, and Pisoni (1988) have shown significant and robust learning of synthetic speech with training that consisted of orthographic feedback on the identity of the synthetic speech, presented concurrently with the repetition of a test item. Davis et al. (2005) also observed effective learning with written feedback presented alongside distorted sentences.

These results suggest that the availability of higher level knowledge of the content of speech at the time that distorted speech is presented facilitates perceptual learning. This is consistent with a top-down component to the perceptual learning process (cf. Norris et al., 2003) in which learning is driven by a comparison between a target representation of a distorted speech token and the lexical or sublexical representations generated by hearing those same items. Clear presentation prior to, or written presentation concurrent with, distorted speech presentation can assist learning by providing the auditory system with the correct target representation for distorted speech. This information allows intermediate representations to be adjusted in order to map distorted speech input onto the correct higher level representation more accurately (for a computational implementation of this form of learning in the context of the TRACE model, see Mirman, McClelland, & Holt, 2006).

However, on the basis of existing results using distorted sentences, it is difficult to distinguish a top-down account from an account in which the order of clear and distorted presentation affects learning because of differences in the memorability of clear compared with distorted speech. Whereas comprehensible, clear speech has a rich linguistic structure and can be easily retained in auditory-verbal short-term memory (STM), distorted speech is less comprehensible and thus cannot engage verbal STM processes as effectively. Listeners are therefore likely to rely more on an auditory echoic store (Crowder & Morton, 1969) to retain poorly comprehended, distorted sentences without additional support from phonological or lexical representations. Reliance on an auditory echoic store might prevent effective comparisons between distorted sentences and subsequently presented clear feedback presentations as the intervening several seconds of speech would overwrite auditory echoic representations. Such a difference in memorability might therefore result in listeners learning less effectively when they had to retain a distorted sentence (i.e., distorted-then-clear feedback) than when they had to retain a clear sentence to compare with a subsequently presented distorted one (clear-then-distorted feedback). In the present study, we therefore seek to replicate the enhanced learning that Davis and colleagues (2005) observed for clear-then-distorted feedback, by using single words. If storage of representations of distorted speech in memory

was the limiting factor in permitting perceptual learning with distorted-then-clear feedback, we might expect that the asymmetry of feedback order would be substantially reduced by using shorter spoken stimuli and a shorter stimulus onset asynchrony. If the advantage for clear-then-distorted feedback is replicated with single-word presentations, this would provide further support for the proposition that top-down feedback is critical for learning.

A further striking observation of Davis and colleagues (2005) was that perceptual learning depended on the lexical content of the training sentences. Listeners exposed to NV sentences composed entirely of nonwords showed dramatically reduced perceptual learning compared with listeners trained with normal NV English sentences. Indeed, in both experiments that included this comparison, listeners trained with nonword sentences were no better than naive listeners at the task of reporting English NV sentences, although as described earlier, some learning was possible from jabberwocky and syntactic prose sentences. These results suggest that lexical information is critical for efficient learning, providing further support for a top-down mechanism being involved in learning. This finding also contrasts with the results of previous studies exploring perceptual learning of time-compressed speech (Altmann & Young, 1993; Pallier et al., 1998; Sebastian-Galles, Dupoux, Costa, & Mehler, 2000) in which jabberwocky and noncomprehended foreign-language stimuli could provide for effective learning.

However, the apparent importance of lexical information in Davis et al.'s (2005) investigation may again have been due to the demands of retaining the nonword stimuli in STM, rather than to any dependence of perceptual learning on lexical information per se. If perceptual learning is enhanced by comparisons between clear and distorted presentations (as suggested earlier), then learning from clear-then-distorted presentation of nonword sentences would require listeners to maintain an accurate representation of a clearly presented nonword sentence (typically 6 to 17 syllables long) in phonological STM. This task of remembering long nonword sequences would be well beyond the capabilities of most normal listeners (cf. Gathercole, Willis, Baddeley, & Emslie, 1994). In an attempt to rule out the influence of phonological STM on learning of nonword sentences, one of the studies reported in Davis et al. (2005) combined auditory presentation of vocoded nonword sentences with written presentation of an orthographic transcription of the same sentence. This manipulation failed to demonstrate perceptual learning, suggesting that phonological STM capacity was not the critical limiting factor in learning from nonword sentences. However, it can be argued that reading a nonword sentence is rather challenging, and so further converging evidence on this issue would be valuable. With this goal in mind, we sought to replicate the effect of lexical information on perceptual learning of NV speech in a study comparing training using isolated words and nonwords presented in clear-then-distorted form. As neither isolated words nor nonwords present more than a minimal load on phonological STM, this experiment therefore provides a test of whether perceptual learning of NV speech is indeed dependent on lexical information (as suggested for other forms of distortion; cf. Norris et al., 2003).

Experiment 1: Effects of Feedback Order

In this experiment, we sought to replicate and extend the findings of Davis et al. (2005) in two ways. First, by training and testing with single NV words, we could more accurately assess generalization to novel lexical items and rule out the possibility that general information about the structure and content of the training sentences could lead to improved word report. A second motivation for this study was to test the effect of feedback order (clear-then-distorted [henceforth referred to as CD] feedback versus distorted-then-clear [henceforth referred to simply as DC] feedback) via shorter single-word stimuli and a shorter interstimulus interval (ISI) between clear and distorted presentations. Compared with previous studies using sentences, the DC feedback condition places a reduced demand on echoic memory for distorted speech, allowing for a stronger test of the role for top-down feedback in perceptual learning.

Method

Participants. Twenty volunteers from the volunteer panel of the Medical Research Council Cognition and Brain Sciences Unit took part in the experiment (4 of these participants were men; all were right-handed; and the average age was 20 years and 8 months, with a range from 18 years and 2 months to 22 years). Participants had no history of hearing impairment or dyslexia.

Materials. We selected 60 monosyllabic and 60 bisyllabic English nouns and divided them into two groups that were matched on number of syllables ($M = 1.5$ for each group), number of phonemes ($M = 4.1$ for each group), uniqueness point (from the CELEX database [Baayen, Piepenbrook, & Gulikers, 1995]: $M = 2.4$ phonemes for each group), word-form frequency (from CELEX: Group A, $M = 17.11$ per million; Group B, $M = 16.7$ per million), and imageability (obtained from the Medical Research Council psycholinguistic database [Coltheart, 1981]: $M = 548$ for each group). There were no significant differences in any of these properties across the two groups. The words were originally recorded by a 20-year-old female speaker of southern British English for a previous study (Orfanidou, Marslen-Wilson, & Davis, 2006). For the full list of words, see Appendix A. Words were recorded in mono onto digital audiotape at a sampling rate of 44.1 kHz and were then digitally transferred to a computer hard disk as 16-bit wave audio files. Stimuli were then noise vocoded following the procedure described by Shannon et al. (1995), with Praat software (Boersma & Weenink, 2003) and a modified version of a script (first written by Chris Darwin) that implements the processing steps described below. The words were first filtered into six logarithmically spaced frequency bands between 50 and 8000 Hz. Contiguous band-pass filters were constructed in the frequency domain: Passbands were 3 dB down at 50, 229, 558, 1161, 2265, 4290, and 8000 Hz with a roll-off of 22 dB per octave (cutoff frequencies chosen to simulate equal distances along the basilar membrane; Greenwood, 1990). For each spoken word, the amplitude envelopes of the energy contained within each frequency band were extracted via the standard Praat algorithm (squaring intensity values and convolving with a 64-ms Kaiser-20 window, removing pitch-synchronous oscillations above 50 Hz). The resulting envelopes were then applied to band-pass filtered noise in the same frequency ranges. Finally, the resulting bands of modulated noise were recombined to produce the distorted word.

Procedure. Stimuli were presented over Sennheiser (Wedemark, Germany) HD250SP headphones through a QED (Veda Products, Bishop's Stortford, Hertfordshire, United Kingdom)

headphone amplifier, from a desktop PC fitted with a Soundblaster (Creative Labs, Singapore) Live sound card, using DMDX software (Forster & Forster, 2003). Listeners were asked to listen carefully to each of the 120 NV words and to repeat each word as quickly and accurately as possible. Responses were recorded to a computer hard disk using an AKG (AKG Acoustics, Vienna, Austria) table-mounted microphone. Listeners received feedback after each test item. Ten listeners received feedback in which the clear word (C) preceded a single repetition of the distorted word (D) (the DCD group), and 10 received feedback in which the repetition of the distorted word (D) preceded the clear presentation of the word (C) (the DDC group). Listeners were provided with a 10-s silent period in which to respond after presentation of the test word, and the feedback presentations were separated by a 200-ms gap. Every trial was preceded by a brief warning tone. The feedback presentations were separated by a 200-ms gap. The structure of the trials is illustrated in Figure 1.

The presentation order of Word Sets A and B was counterbalanced across subjects within groups (i.e., 5 listeners in each group heard all the items of Set A followed by all the items of Set B, and the other 5 heard B followed by A). The order of items within word sets was randomized across subjects. The recorded spoken responses were scored for accuracy both in terms of the percentage of phonemes correct and word recognition. Reaction times were also collected for all responses on the basis of an off-line analysis of the recorded wave form of the subjects' vocal responses. However, because no reliable differences in response times were observed, we will not report or discuss these data.

Results

Figure 2 shows performance scores for both phonemes correct and word recognition averaged over two 60-word blocks. Because two groups of participants were tested on the same word blocks in different orders, we included in our statistical analysis an additional dummy variable code for the order in which the two word groups were presented to account for differences in overall report score between these two item groups; however, effects of this variable will not be reported (Pollatsek & Well, 1995). Although Pollatsek and Well (1995) suggest that one should conduct analyses of results averaged by items and by participants in order to fully account for effects of interitem variability, Raaijmakers and colleagues have argued persuasively that conducting both these analyses is not necessary in a fully counterbalanced design such as that used here (Raaijmakers, 2003; Raaijmakers, Schrijnemakers, & Gremmen, 1999). We will therefore report only the outcome of analyses by participants.

Evidence for learning. To assess the extent of any learning, we carried out mixed analyses of variance (ANOVAs) on the accuracy data. Word-identification and phoneme-identification scores were analyzed separately. Accuracy scores were averaged by participants, with test block (early or late) as a within-participants factor and condition (DCD or DDC) as a between-participants factor.

The results show that learning took place over the course of the experiment: Participants performed significantly better on the second block of 60 words than on the first—phonemes correct, $F(1, 16) = 19.096$, $p < .001$, $\eta^2 = 0.544$; words correct, $F(1, 16) = 10.039$, $p < .006$, $\eta^2 = 0.386$—indicating that participants' understanding of NV speech improved over the course of the experiment. Listeners' word recognition performance was worse on isolated words than on complete sentences. After training with 15 sentences with CD feedback, containing approximately 120 words, Davis et al. (2005) showed that participants' word identification scores were around 60%, whereas our DCD participants were identifying an average of only 39% of the isolated words. This is consistent with the fact that the rich contextual information in normal sentences (not present in isolated words) can be used by listeners to improve their word report scores for NV speech (cf. Grosjean, 1980; Salasoo & Pisoni, 1985).

The DCD group performed significantly better than the DDC group—phonemes correct, $F(1, 16) = 3.421$, $p < .083$, $\eta^2 = 0.176$; words correct, $F(1, 16) = 6.956$, $p < .018$, $\eta^2 = 0.303$—replicating the results of Davis et al. (2005). This suggests that knowledge of the identity of a distorted item prior to its second presentation (as happens in DCD but not DDC) increases the efficiency of learning—that is, this kind of feedback is more effective. Although the Block × Condition interaction was not significant—phonemes correct, $F(1, 16) = 1.403$, $p < .253$, $\eta^2 = 0.081$; words correct, $F(1,16) = 2.874$, $p < .109$, $\eta^2 = 0.152$—the significant main effect of training condition must reflect a difference in the rate of learning, as all participants, regardless of training condition, were naive to the distortion at the beginning of the experiment and must therefore have had equal levels of initial performance. The lack of a significant interaction between block and condition may be due to substantial learning in the DCD condition occurring fairly early within the first block,

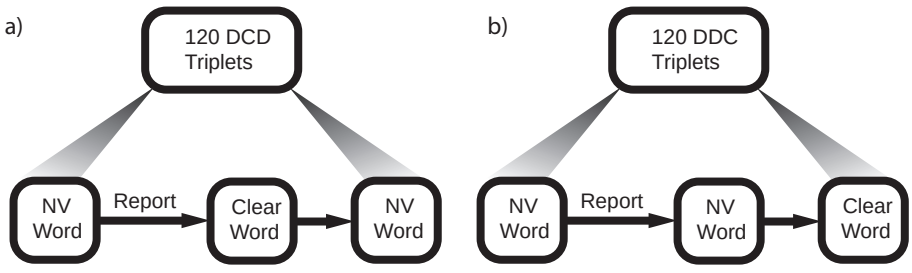

Figure 1. Structure of trials in Experiment 1. Each trial comprised a test noise-vocoded (NV) word after which listeners were provided with a 10-s silent period in which to respond. They then received feedback: either the clear word followed by the identical distorted word (distorted-clear-distorted [DCD] condition; Panel a), or the identical NV word followed by its clear counterpart (distorted-distorted-clear [DDC] condition; Panel b).

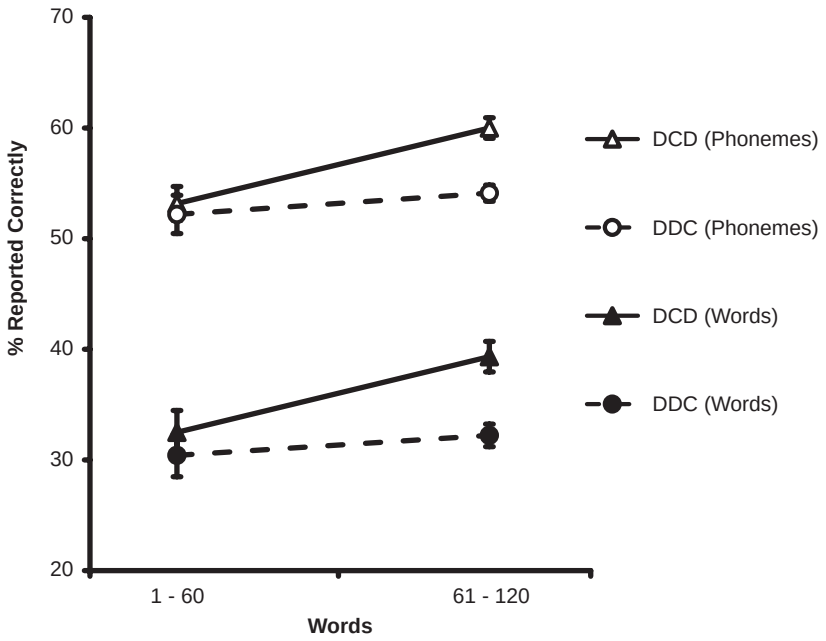

Figure 2. Mean scores in Blocks 1 and 2 of Experiment 1 for both distorted-distorted-clear (DDC) and distorted-clear-distorted (DCD) conditions, scored by the percentage of words correctly reported and the percentage of phonemes correctly reported. Each block comprised 60 test items. The error bars represent the standard error of the mean.

thereby increasing performance in both blocks relative to the DDC condition.

In combination with the main effect of condition described before, these results demonstrate a significant difference in the effectiveness of the two training conditions used. Because both groups of participants reported vocoded words after only a single presentation and heard the same number of repetitions of vocoded and clear words subsequently, the difference in report score can be attributed to the order in which word repetitions were presented. Report scores for the two groups of participants did not differ for the very first test word, before listeners heard any repetitions, $t(22) = 1.1086$, $p < .292$, but did differ between the two conditions for the second block of words. This therefore suggests a difference in the rate of learning between the DCD and DDC conditions.

Effects of interitem variability. Despite the randomization of presentation order and the careful matching of words on the features described above, there was still considerable variability in the apparent difficulty of the NV words. Many words were not reported correctly by any participants in the study, and several were reported correctly by at least 19 of the 20 participants.

In order to investigate the source of the variability, we correlated the variables on which the stimuli were matched (number of syllables and phonemes, uniqueness point, word-form frequency and imageability), as well as neighborhood density and acoustic duration, with the proportion of participants (out of 20) who reported each word correctly. Significant bivariate correlations were found for acoustic duration (Pearson's $R = .387$, $p < .001$), number of phonemes ($R = .314$, $p < .001$), neighborhood density ($R = -.272$, $p = .003$), and uniqueness point ($R = .208$, $p < .023$). Significant correlations were not observed with any of the other parameters on which the stimuli had been matched (word-form frequency, $R = .037$, $p < .727$; imageability, $R = .075$, $p < .451$; number of syllables, $R = .147$, $p < .109$). Acoustic duration, number of phonemes, uniqueness point, and neighborhood density were, however, all significantly intercorrelated (see Table 1 for correlation coefficients). Hence, we calculated partial correlation coefficients that controlled for the effects of each of these significantly intercorrelated variables in turn.

When controlling for the effects of the intercorrelated variables, we found that only the effect of acoustic duration remained sig-

Table 1
Pearson Correlation Coefficients of Variables That Are Significantly Correlated With Word Difficulty in Experiment 1

Variable	1	2	3	4	5
1. Mean score					
R	—	.387	.314	−.272	.208
p		.000	.000	.003	.023
2. Acoustic duration					
R	.387	—	.433	−.252	.213
p	.000		.000	.006	.020
3. Number of phonemes					
R	.314	.433	—	−.496	.457
p	.000	.000		.000	.000
4. Neighborhood density					
R	−.272	−.252	−.496	—	−.338
p	.003	.006	.000		.000
5. Uniqueness point					
R	.208	.213	.457	−.338	—
p	.023	.020	.000	.000	

Note. $N = 120$.

nificant (partial correlation coefficient = .290, $p = .002$); uniqueness point (partial correlation coefficient = .056, $p = .546$), number of phonemes (partial correlation coefficient = .077, $p = .407$), and neighborhood density (partial correlation coefficient = −.125, $p = .181$) were not significantly correlated with performance. This suggests that of the variables examined, only acoustic duration can significantly predict the difficulty of NV words, and even this accounts for approximately only 8.4% of the observed variance. We speculate that this weak effect could result if there is some additional compensation mechanism that requires a period of input to become fully operational after the onset of the word. A similar mechanism has been proposed in the comprehension of accented speech (Floccia, Goslin, Girard, & Konopczynski, 2006) and would potentially explain more accurate identification of distorted target words with longer acoustic durations.

Discussion

The results of Experiment 1 showed a clear and significant improvement in listeners' word report performance over the course of the experiment even in the absence of sentence content and the repetition of function words such as "and" or "the" as occurred in previous experiments using sentence materials. Because improved performance was observed from one block to the next without any repetition of stimuli, we conclude that exposure to NV speech results in changes to the sublexical processing of distorted input, which permits a greater proportion of words to be reported correctly. We will return to this point in the General Discussion.

The results of this experiment also replicate the earlier finding that clear-then-distorted presentation (DCD condition) provides for more effective perceptual learning than distorted-then-clear feedback (DDC condition; cf. Davis et al., 2005). In line with the proposal made previously by Davis and colleagues (2005), we suggest that knowledge of the identity of the target word when hearing the second presentation of the distorted word provides for a more accurate "teaching signal" (the correct interpretation of distorted speech input), which can drive learning more rapidly. It is important to note that this learning process cannot operate as effectively when presentation of distorted speech precedes presentation of clear speech (and hence the availability of the target signal). Whereas previous work used sentence stimuli that, when distorted, easily exceeded listeners' capacity to maintain auditory representations of them, it appears likely that participants in the DDC condition could retain some auditory representation of the NV word until subsequent clear presentation occurred. However, this echoic representation was not sufficient to support perceptual learning that was as efficient as in the DCD condition, where clear presentation preceded distorted presentation.

It remains possible that an echoic representation of the distorted word in the DC feedback condition was lost due to backward masking by the subsequent clear stimulus, or even that this auditory representation simply decayed over the ISI. However, backward masking is poorly understood, and the amount of backward masking can vary, depending on task-specific factors such as the amount of practice listeners have had (for further discussion, see Moore, 1997, p. 129). It is therefore difficult to determine whether backward masking of auditory stimuli is a critical factor in explaining the present results. However, given the asymmetry in the effect of DCD and DDC training, we must assume either that there is an asymmetry in backward masking (with clear stimuli masking distorted stimuli, but not vice versa) or that learning is supported by a higher level representation that can derive only from clear speech. In the absence of relevant evidence concerning asymmetries in backward masking from clear and distorted speech, at present we would favor our original hypothesis that perceptual learning depends on top-down feedback from higher level representations that can be effectively accessed only from clear speech. We will return to the possible role of backward masking in the General Discussion.

Experiment 2: Training With Words and Nonwords

The results of Experiment 1 suggest that a process by which distorted speech is compared with prior knowledge of speech content is critical for learning. In Experiment 2, we sought to establish whether the presence of lexical information is also critical for effective perceptual learning of NV speech. Davis et al. (2005) showed that sentences made up of real words are significantly more effective at producing learning than are sentence-length strings of nonwords, a finding that they interpreted as evidence for the involvement of top-down lexical feedback in the perceptual learning process. However, an alternative explanation is that the long nonword strings used by Davis et al. could not be retained in phonological STM and therefore prevented the comparison of the clear and distorted sentences in the nonword training condition. In Experiment 2, we therefore compared the perceptual learning produced by DCD training with single NV words and nonwords by using a crossover design. For these single-word stimuli, the load on phonological STM for both words and nonwords is minimal, and both types of material should be equally well retained over the short interval between presentation of the clear and distorted stimulus.

Method

Participants. Thirty-six participants, naive to NV speech (18 of these participants were men; 31 were right-handed; and the mean age was 20 years and 3 months, with a range from 18 years to 26 years and 1 month), all with normal hearing and no history of language impairment, took part in the study. All volunteers were paid for their participation.

Materials. We generated three groups (A, B, and C) of 40 words (20 monosyllables and 20 bisyllables) by regrouping the stimulus set from Experiment 1. These three sets of items were matched for neighborhood density (data from the CELEX database; means: Group A = 9.28, Group B = 9.35, Group C = 9.20), spoken word-form frequency per million (means: Group A = 5.53, Group B = 5.48, Group C = 5.43), acoustic duration (means: Group A = 0.62 s, Group B = 0.63 s, Group C = 0.62 s), and numbers of phonemes (means: Group A = 4.08, Group B = 4.05, Group C = 4.08). Using the word report data from Experiment 1, we also equated mean performance over the three groups (means: Group A = 35.9%, Group B = 32.7%, Group C = 31.1%). There were no significant differences between the groups on any of these properties. These three groups of words constituted the test blocks in the experiment. The stimuli were noise vocoded via Praat (Boersma & Weenink, 2003), as in Experiment 1.

Another group of 120 words (60 monosyllables and 60 bisyllables) and a matched group of 120 nonwords were selected from

materials used by Orfanidou et al. (2006; see Appendix B) to serve as training stimuli. The nonwords were matched to the 120 word stimuli on the basis of overall phonemic composition of the two groups, such that each of the 120-item word and nonword groups contained the same overall distribution of phonemes. These words and nonwords were recorded by the same female speaker of southern British English in the same session as the stimuli used for Experiment 1 (see Orfanidou et al., 2006) and were noise vocoded in the same way as Groups A, B, and C. The stimuli in these two training groups were then assembled to make triplets providing feedback for the listeners, as in the DCD condition of Experiment 1. Each NV stimulus (D) was followed by a clear version (C) followed by the distorted version again (D), with a 1-s gap between the repetitions of the stimuli, to make sets of 120 word (W) and nonword (N) DCD triplets.

Procedure. We conducted our assessment in a sound-treated booth using a custom computer program written in Visual Basic, run on the same PCs used previously. We were concerned that listeners would have difficulty reporting nonword stimuli, and so we resorted to a train-then-test procedure used previously by Davis et al. (2005) in order to compare the efficacy of NV nonword and word stimuli at producing learning. The experiment consisted of the following five phases: Phase 1 was an initial test session, assessing comprehension of NV words before any training; Phase 2 comprised a DCD training session (with either NV words or NV nonwords), in which subjects simply listened to stimuli and were not expected to respond; in Phase 3, we conducted another test session to assess improvements in NV word comprehension after the initial training session; Phase 4 comprised DCD training with the stimulus type (words or nonwords) that was not presented in Phase 2; and in Phase 5, there was a final assessment of comprehension of NV words. As reaction time data were not a useful measure of learning in Experiment 1, we did not collect verbal responses. Participants were instead asked to type their responses on a standard QWERTY computer keyboard. Prior to testing, all participants heard five nonwords and were asked to transcribe each one as they heard it. This screening ensured that listeners were able to retain isolated nonwords in STM.

In Phases 1, 3, and 5, all listeners were tested on their ability to report 40 NV words, each preceded by a warning tone. A period of 10 s was allowed for a typed response to be made after each word. Each distorted item was presented only once and was not accompanied by feedback; we wished to minimize any learning during this phase. Phases 2 and 4 were training phases, during which subjects were told simply to listen to the stimuli and to make no response. Participants were initially presented with either word (18 listeners) or nonword (18 listeners) stimuli, as DCD triplets. The subsequent training session consisted of the other group of training items that had not previously been presented. The two conditions are illustrated in Figure 3. The use of a crossover design was intended to allow measurement of the effectiveness of perceptual learning from words and nonwords either by using between-subjects comparisons of data from before and after the first block of training (similar to the design used by Davis et al., 2005) or by comparing performance following the first and second block of training for groups trained with words following nonwords and vice versa. Although this latter comparison may potentially be affected by carry-over effects (as the group that received word training first may not subsequently show any additional benefit as a result of further training), the results of the study in fact suggest that carry-over effects did not significantly alter our interpretation of the results.

The order of presentation of the words within both testing and training blocks was randomized for each subject. The order of presentation of the three test stimulus sets was counterbalanced across subjects. Responses were scored in terms of words transcribed correctly (homophones of the target were scored as correct) and percentage of phonemes transcribed correctly for each word.

Results

Figure 4 shows the results averaged across participants and over all the test words in a block for the three test blocks of words, for both conditions. Analyses of the data averaged across participants were carried out separately for the results scored by percentage of words and phonemes reported correctly. As in Experiment 1, we included an additional dummy variable in all analyses to code for the order of presentation of the test blocks, although main effects and interactions involving this dummy variable will not be reported (cf. Pollatsek & Well, 1995).

To establish whether there was learning over the course of the experiment, we entered scores averaged by participants into mixed ANOVAs, with test block as a three-level within-participants factor and training order (words then nonwords or nonwords then words) and stimulus order (ABC, ACB, BAC, BCA, CAB, CBA) as between-participants factors, although effects of this stimulus-order factor will not be reported (cf. Pollatsek & Well, 1995).

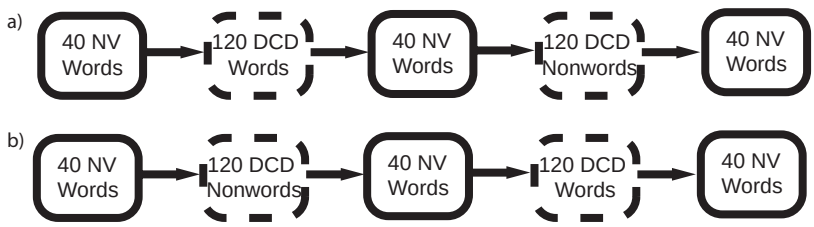

Figure 3. Structure of Experiment 2. Solid boxes denote test sessions during which listeners heard and reported noise-vocoded (NV) words without feedback; dashed boxes show training sessions during which listeners heard distorted words and nonwords with clear-then-distorted feedback (as in the distorted-clear-distorted [DCD] condition of Experiment 1). Panel a shows the word–nonword training condition; Panel b shows the nonword–word training condition.

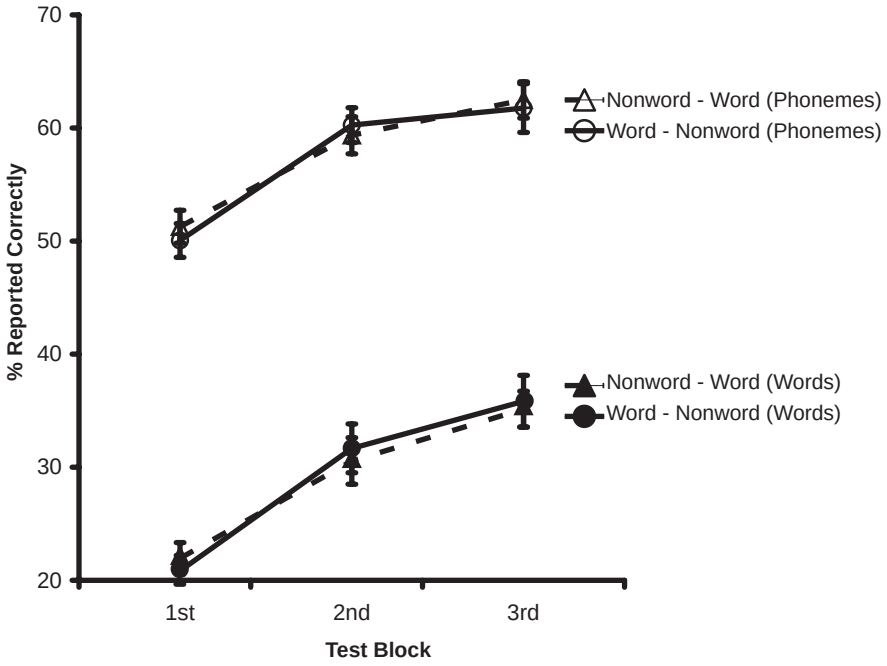

Figure 4. Mean scores in Test Blocks 1, 2, and 3 of Experiment 2 for both nonword–word and word–nonword conditions, scored by both the percentage of words correctly reported and the percentage of phonemes correctly reported. Each block comprised 40 test items. The error bars represent the standard error of the mean.

There was a significant effect of block, indicating that participants' performance improved over the course of the experiment; words correct: Block 1 = 21.4%, Block 2 = 31.1%, Block 3 = 35.5%, $F(2, 48) = 52.28$ $p < .001$, $\eta^2 = 0.685$; phonemes correct: Block 1 = 52.4%, Block 2 = 57.9%, Block 3 = 58.5%, $F(2, 48) = 6.58$, $p < .003$, $\eta^2 = 0.215$. There was no significant effect of training order (words followed by nonwords vs. nonwords followed by words), suggesting that there was no difference in effectiveness of the two types of training stimuli; words correct: $F(1, 24) = 0.012$, $p < .913$, $\eta^2 = 0.0005$; phonemes correct: $F(1, 24) = 0.035$, $p < .852$, $\eta^2 = 0.01$.

To assess the effects of training with words or nonwords, we entered the differences in report performance from Test Block 1 to Test Block 2 and from Test Block 2 to Test Block 3 into mixed ANOVAs with the intervening training type (word or nonword) entered as a within-subject factor and training session and stimulus order entered as between-subjects factors. There was no significant main effect of training type; scored by percentage of words correct: mean improvement due to training with words = 7.7%, due to training with nonwords = 6.4%, $F(1, 24) = 0.231$, $p < .63$, $\eta^2 = 0.010$; scored by percentage of phonemes correct: mean improvement due to training with words = 6.7%, mean improvement with nonwords = 4.8%, $F(1, 24) = 0.053$, $p < .820$, $\eta^2 = 0.002$. There was no reliable effect of training position (first or second) on the improvement in performance due to training; scored by percentage of words correct: $F(1, 24) = 0.291$, $p < .594$, $\eta^2 = 0.012$; scored by percentage of phonemes correct: $F(1, 24) = 1.241$, $p < .276$, $\eta^2 = 0.049$. The analysis does not indicate any significant difference between the effects of the first or second training block, suggesting that carry-over effects did not impact on the present results.

Discussion

In this experiment, no significant difference was observed in the efficacy of word and nonword training for participants learning to understand single NV words. This finding of effective perceptual learning from nonword DCD presentations appears at odds with the results of Davis et al. (2005), who showed that listeners exposed to 20 NV nonword sentences with DCD presentation were no better at reporting English sentences than were naive listeners. The present results demonstrate equivalent perceptual learning from word and nonword sequences and therefore suggest that the supposed lexicality effect observed by Davis et al. (2005) was not directly due to the presence or absence of lexical information in the training stimuli, but was rather due to the difficulty of maintaining a sentence-length nonword string in phonological STM in order to compare clear presentations with subsequent presentation of distorted speech. The isolated nonwords used here can be easily and accurately retained in phonological STM following clear presentation and can therefore provide a suitable teaching signal during presentation of distorted speech. These results therefore appear to be consistent with the findings of Experiment 1 in pointing to the critical involvement of a comparison process between clear and distorted stimuli for perceptual learning. We will expand on this account in the General Discussion.

Some of the nonwords used in training are similar to real words. These may have been confused with real words, and it has been suggested that this could provide listeners with a potential source of feedback about the phonological content of the nonwords, which might have increased the efficacy of nonword training. However, if nonwords were frequently confused with similar-sounding words, then at least one phoneme in the distorted non-

word would be miscategorized. If this happened repeatedly for different nonwords, it is likely that erroneous mappings of NV phonemes onto internal phonological representations would be learned. This would not serve to enhance learning but would in fact be to the detriment of performance after training with nonwords.

As in Experiment 1, we again observed significantly improved performance in later test blocks despite different word stimuli being used in each test block. This finding further confirms that learning generalized to untrained words and supports the proposal made by Davis et al. (2005) that perceptual learning of NV speech must be occurring at a level of representation that is prelexical and is therefore applicable to all words heard in distorted form.

General Discussion

The results of the experiments reported here demonstrate a robust and rapid perceptual learning process at work for single NV words. For instance, in Experiment 1, participants in the DCD group reported 32.5% of the first 60 NV words correctly, whereas the second group of 60 words was reported with 39.3% accuracy. In Experiment 2, in which test performance was assessed without feedback presentations and hence without concurrent learning, report scores showed an even clearer improvement (from 21.4% initially to 35.5% for the final NV set of 40 test words). Although the numerical magnitude of these performance improvements is less dramatic than those previously reported for NV sentences by Davis et al. (2005), learning effects with words were statistically significant; therefore, the current results again reflect the operation of powerful mechanisms that substantially alter the perception of this form of distorted speech. In this discussion, we will consider what the current results add to the existing understanding of perceptual learning of speech, for both NV words and sentences, and draw comparisons, where relevant, between perceptual learning of NV and other forms of distorted speech.

The Locus of Perceptual Learning of NV Speech

The results of the two experiments show that learning to understand NV words is a process that is not specific to the trained items but that readily generalizes to novel words that are subject to the same distortion. In previous studies using sentences (Davis et al., 2005), generalization to novel words could conceivably have arisen from supralexical influences on sentence report. However, for the single NV words used here, we can be confident that the perceptual learning process has altered the speech processing system in a manner that is applicable to all NV words. Our observation that perceptual learning generalizes to novel words is difficult to explain in terms of changes to lexical or supralexical processes (which would be specific to trained words) but rather suggests that a sublexical level of processing is altered by perceptual learning.

We make no claims here about the nature of this sublexical representation, but given that learning of NV speech generalizes to novel monosyllables, we can be confident that it involves subsyllabic units. For instance, changes to the representations that encode phonemes or phonetic features would be applicable to all NV words and could therefore produce learning that generalizes to novel words. Future experiments testing for generalization between trained and untrained phonemes will be valuable in order to distinguish between perceptual learning at a phonemic or more peripheral level. Kraljic and Samuel (2006) demonstrated generalization of learning from one continuum of ambiguous phonemes to another continuum that varied on the same phonetic feature (/d/ vs. /t/ and /b/ vs. /p/, both of which critically vary in terms of voice-onset time). This suggests that perceptual learning of ambiguous phonemes occurs at a subphonemic featural level, as learning at a phonemic level would not be expected to generalize between continua. It would be of considerable interest to apply similar methods to the investigation of perceptual learning of NV words.

One piece of evidence suggests that a nonperipheral locus of learning comes from applying a "psychoanatomical" method popular in visual perception. This method uses the degree of generalization of perceptual learning (between different eyes, retinal positions, etc.) to assess the neural and functional locus of perceptual learning (Ahissar & Hochstein, 2004; Hochstein & Ahissar, 2002). In recent work, Hervais-Adelman, Davis, Johnsrude, Taylor, and Carlyon (2007) have demonstrated that perceptual learning of NV sentences generalizes over frequency regions. Because early stages of auditory processing are highly frequency selective, this generalization indicates that the learning process probably does not result from changes to these earliest (peripheral) stages of sound processing. Instead, this finding suggests that perceptual learning occurs at a non-frequency-selective level of processing, abstracted from simple acoustic analysis and probably involving cortical regions beyond core auditory cortex.

The Role of Feedback Order in Perceptual Learning

Another finding from Experiment 1 that replicates earlier results using sentences (Davis et al., 2005) is that perceptual learning of NV words proceeds more rapidly when feedback is provided via CD (clear-then-distorted) presentation than with DC (distorted-then-clear) presentation. Indeed, in Experiment 1, there was surprisingly little evidence of reliable perceptual learning for the DDC condition—in contrast to previous results with sentences. As discussed previously, the effect of feedback order points to a top-down component of the learning process and a temporal asymmetry such that clear feedback before but not after distorted speech presentations enhances learning (previous results with sentences similarly showed that learning with DC feedback was no more effective than D feedback alone; Davis et al., 2005). We suggest that in the case of previous studies using sentences, these results may have occurred because distorted sentences cannot be retained in auditory memory or STM long enough for a comparison to be carried out between the representation of the distorted sentence and the representation of the clear target (as furnished by subsequent presentation). The present replication of this effect of feedback order with shorter stimuli and hence a reduced requirement for long-term storage of unanalyzed auditory representations might therefore be taken to suggest that the difference between the conditions is not merely due to limitations in listeners' ability to store representations of distorted targets.

We interpret the superiority of learning in the DCD condition as evidence that performance can be improved when listeners have a clear phonological representation of the target word to compare with the vocoded speech. As discussed earlier, one way in which this could happen would be via the phonological representation providing a clear "teaching signal" that helps the auditory system

to map future distorted input onto the correct internal representation. Another factor, which also relies on the effectiveness of comparisons between vocoded speech and a clear phonological representation, is backward masking. Backward masking decays very rapidly in detection tasks, and even for tasks requiring listeners to compare the memory traces of two sounds, such as in "pitch recognition masking," it decays substantially over the 200-ms ISI used here (Massaro, 1975). However, it is possible that backward masking persists over longer time courses for speech stimuli than for the pure tones used to study it previously. If we further assume that listeners compare clear and distorted speech after the final word has been presented, then it is possible that in the DDC condition, the final (clear) word masked the memory trace of the previous distorted word, thereby impairing performance. In contrast, the final (distorted) word in the DCD condition may not have masked the previous (clear) word, presumably because the latter had been recoded into a phonological form. Note that this explanation, like ours, rests on the idea that efficient learning depends on the distorted signal being effectively compared with a clear phonological representation of the same word. The explanations differ only in whether DCD performance is good because the clear speech is available in memory before the final distorted word is present or whether DDC performance is bad because the representation of the penultimate (distorted) word is overwritten by the subsequent clear speech.

At present, then, existing data would strongly suggest that higher level information must be present in auditory memory concurrently with a representation of the distorted target items for effective perceptual learning to occur. We therefore propose that the difference in effectiveness of CD and DC feedback suggests an online learning process by which responses to distorted speech are tuned on the basis of concurrent feedback from a target representation derived from clear speech presentations. As discussed previously by Davis et al. (2005), one might conceive of this learning process as being akin to the supervised learning incorporated into back-propagation network models of spoken word recognition (e.g., Gaskell & Marslen-Wilson, 1997; Norris, 1993), although other supervised learning accounts are also plausible (see, for instance, Mirman et al., 2006). Whichever computational account one might favor, our results suggest that online comparisons between heard speech and a phonological target representation play an important role in supporting perceptual learning of NV speech.

The Role of Lexical Knowledge in Perceptual Learning

One aspect of the current results that is harder to reconcile with previous findings concerns the effect of lexical content on perceptual learning. Previous studies using sentences (Davis et al., 2005) produced no evidence of perceptual learning from NV sentences composed of nonwords. However, Experiment 2 demonstrates that words and nonwords are equally effective as training stimuli for NV speech: This finding appears at odds with the results of Davis et al. (2005). One possible explanation of these discrepant results concerns the impact of STM capacity for nonword sequences. It is very difficult for participants to maintain a representation of a sentence-length string of nonwords in STM because the lack of familiar phonological units prevents effective maintenance and rehearsal. Thus, comparisons between incoming distorted speech and previous clear presentations of nonword sentences are not possible—thwarting an important component of the perceptual learning process. In contrast, isolated, clearly spoken monosyllabic and bisyllabic nonwords place a smaller load on phonological STM (Gathercole et al., 1994) and can be easily retained for the purposes of comparison with distorted speech during DCD presentation.

This observation of effective perceptual learning for isolated nonwords but not for nonword sentences would again be consistent with the theory that perceptual learning is driven by an error signal representing the discrepancy between the auditory input and a phonological representation of the true form of a distorted speech stimulus. Thus, perceptual learning of NV speech may depend on lexical information only to the extent that lexical information is necessary for the maintenance in phonological STM of a target representation prior to presentations of distorted speech.

However, although an STM explanation of these results is appealing and suggests some ways in which seemingly discrepant results with words and sentences can be reconciled, this account is not without its faults. Indeed, there are two existing findings that challenge an account of perceptual learning based on a top-down, phonological (but not lexical) comparison between clear and distorted speech. First and foremost, in the case of sentences, perceptual learning is possible even in the absence of this comparison process. Robust perceptual learning was observed without any feedback presentations for NV sentences in Experiment 1 of Davis et al. (2005). This finding suggests that comparison between clear and distorted speech presentations are not the only mechanism that can support perceptual learning of NV sentences. Instead, we propose that normal sentence stimuli permit a form of top-down perceptual learning to operate without external feedback through ongoing prediction of sentential elements. Clearly this would depend on sentences being of a form that permits ongoing prediction processes to operate, and so we might predict that in the absence of external feedback, learning would be challenged by sentences that included unfamiliar words or lacked higher level meaning.

A further complicating result is that in Experiment 4 of Davis et al. (2005), it was shown that concurrent presentation of written feedback that accompanied presentation of equivalent NV nonword sentences did not permit any perceptual learning. As these written presentations were specifically intended to provide a phonological target representation without placing an excessive load on STM, an account in which STM restrictions were solely responsible for the failure of learning with nonword sentences is not feasible. It might be that some additional factor—such as difficulties in segmenting nonword sentences into isolated phonological units for comparison with written text—might be responsible for the failure of perceptual learning in this case. Consistent with an explanation in terms of segmentation, Davis et al. observed some limited perceptual learning for jabberwocky sentences (sentences with English function words, but with nonwords replacing content words)—a stimulus that provides very little additional lexical support but that does permit some degree of lexical segmentation.

Further experiments therefore seem to be required to explore the role of lexical information in perceptual learning both of NV speech and of other forms of distortion that might rely on similar learning mechanisms. For instance, Altmann and Young (1993) observed that effective training of listeners to understand time-compressed speech was phonologically, rather than lexically, determined. They found that listeners learned to understand time-

compressed speech when trained with jabberwocky sentences or sentences of a nonnative (and therefore entirely incomprehensible) language if the phonology of the two languages was sufficiently similar, a finding supported by later studies of cross-linguistic learning of time-compressed speech (Pallier et al., 1998; Sebastian-Galles et al., 2000). These results appear difficult to reconcile with Davis et al.'s (2005) results on NV sentences; however, in light of the present results with isolated words, we would perhaps suggest that both findings point to a role for phonological information in supporting perceptual learning.

Another form of perceptual learning that has been shown to depend on lexical knowledge occurs for artificially modified speech that contains an ambiguous fricative (midway between /s/ and /f/). Perceptual learning has recently been shown to depend on the presence of lexical information that supports one or other interpretation of the ambiguous fricative (Norris et al., 2003). Thus, this result suggests that lexical knowledge plays a critical role in individuals learning the correct interpretation of ambiguous speech sounds (see Eisner & McQueen, 2005; Kraljic & Samuel, 2005). However, the current results with isolated nonwords suggest that it is the phonological information that automatically accompanies lexical identification, rather than lexical identification per se, that might be critical for perceptual learning. It would be of interest to explore whether visual feedback that disambiguates an ambiguous speech sound would be sufficient to support perceptual learning in the absence of lexical information.

Conclusion

The experiments reported here further extend our understanding of the nature and locus of perceptual learning of NV words. The generalizability of the learning to untrained words strongly suggests a sublexical locus for perceptual learning. Significant effects of feedback order on perceptual learning suggest that mechanisms involved in comparing distorted speech to a phonological representation of the intended form of speech are critical for learning. However, in contrast to previous results with sentences, this study indicated that lexical information does not seem to be involved in the learning of NV words. Overall, the findings point to an important role for top-down processes in guiding the learning process, but by mechanisms that can operate via different sources of information depending on the nature and content of the training materials. In combination, then, we hypothesize that the presence or absence of external feedback may not be so crucial as the presence of some constraint on the interpretation of distorted speech that permits listeners to reinforce accurate perceptual hypotheses and make alterations that can correct inaccurate hypotheses.

References

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Science, 8,* 457–464.

Altmann, G., & Young, D. (1993, September). *Factors affecting adaptation to time-compressed speech.* Paper presented at the meeting of the Eurospeech 9, Berlin, Germany.

Baayen, R. H., Piepenbrook, R., & Gulikers, L. (1995). *The CELEX Lexical Database*. Retrieved from http://www.ru.nl/celex/subsecs/section_psy.html

Boersma, P., & Weenink, D. (2003). Praat: Doing phonetics by computer (Version 4.1) [Computer program]. Retrieved October 2003, from http://www.praat.org

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America, 116,* 3647–3658.

Clopper, C. G., & Pisoni, D. B. (2004). Effects of talker variability on perceptual learning of dialects. *Language and Speech, 47,* 207–239.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 33A,* 497–505.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception & Psychophysics, 5,* 365–373.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A. G., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General, 134,* 222–241.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics, 67,* 224–238.

Evans, B. G., & Iverson, P. (2004). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America, 115,* 352–361.

Faulkner, A., Rosen, S., & Smith, C. (2000). Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: Implications for cochlear implants. *Journal of the Acoustical Society of America, 108,* 1877–1887.

Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003, October 3). Consolidation during sleep of perceptual learning of spoken language. *Nature, 425,* 614–616.

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1276–1293.

Forster, K. L., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35,* 116–124.

Fu, Q. J., & Galvin, J. J., III. (2003). The effects of short-term training for spectrally mismatched noise-band speech. *Journal of the Acoustical Society of America, 113,* 1065–1072.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12,* 613–656.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory, 2,* 103–127.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49,* 585–612.

Greenspan, S. L., Nusbaum, H. C., & Pisoni, D. B. (1988). Perceptual learning of synthetic speech produced by rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 421–433.

Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years later. *Journal of the Acoustical Society of America, 87,* 2592–2605.

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

Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., Taylor, K., & Carlyon, R. P. (2007). *Generalisation of perceptual learning of vocoded speech.* Manuscript in preparation.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36,* 791–804.

Ives, D.T., Smith, D. R., & Patterson, R. D. (2005). Discrimination of speaker size from syllable phrases. *Journal of the Acoustical Society of America, 118,* 3816–3822.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology, 51,* 141–178.

Kraljic, T., & Samuel, A. G. (2006). How general is perceptual learning for speech? *Psychonomic Bulletin & Review, 13,* 262–268.

Loizou, P. C., Dorman, M., & Tu, Z. (1999). On the number of channels needed to understand speech. *Journal of the Acoustical Society of America, 106*(4, Pt. 1), 2097–2103.

Massaro, D. W. (1975). Backward recognition masking. *Journal of the Acoustical Society of America, 58,* 1059–1065.

Mehler, J., Sebastian, N., Altmann, G., Dupoux, E., Christophe, A., & Pallier, C. (1993). Understanding compressed sentences: The role of rhythm and meaning. In C. von Euler, R. R. Llins, A. M. Galaburda, & P. Tallal (Eds.), *Annals of the New York Academy of Sciences: Vol. 682. Temporal information pressing in the central nervous system: Special reference to dyslexia and dysphasia* (pp. 272–282). New York: New York Academy of Sciences.

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology, 41,* 329–335.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later occurring information on the perception of stop consonants and semi-vowels. *Perception & Psychophysics, 25,* 457–465.

Mirman, D., McClelland, J. L., & Holt, L. L. (2006). An interactive Hebbian account of lexically guided tuning of speech perception. *Psychonomic Bulletin & Review, 13,* 958–965.

Moore, B. C. J. (1997). *An introduction to the psychology of hearing* (4th ed.). London: Academic Press.

Norris, D. (1993). Bottom up connectionist models of "interaction." In G. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: Second Sperlonga meeting* (pp. 211–233). Hove, United Kingdom: Erlbaum.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47,* 204–238.

Orfanidou, E., Marslen-Wilson, W. M., & Davis, M. H. (2006). Neural response suppression predicts repetition priming of spoken words and pseudowords. *Journal of Cognitive Neuroscience, 18,* 1237–1252.

Pallier, C., Sebastian-Galles, N., Dupoux, E., Christophe, A., & Mehler, J. (1998). Perceptual adjustment to time-compressed speech: A cross-linguistic study. *Memory and Cognition, 26,* 844–851.

Peelle, J. E., & Wingfield, A. (2005). Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1315–1330.

Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 785–794.

Raaijmakers, J. G. (2003). A further look at the "language-as-fixed-effect fallacy." *Canadian Journal of Experimental Psychology, 57,* 141–151.

Raaijmakers, J. G., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41,* 416–426.

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review, 101,* 129–156.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981, May 22). Speech perception without traditional speech cues. *Science, 212,* 947–950.

Rosen, S., Faulkner, A., & Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *Journal of the Acoustical Society of America, 106,* 3629–3636.

Saberi, K., & Perrott, D. R. (1999, April 29). Cognitive restoration of reversed speech. *Nature, 398,* 760.

Salasoo, A., & Pisoni, D. B. (1985). Interaction of knowledge sources in spoken word identification. *Journal of Memory and Language, 24,* 210–231.

Schwab, E. C., Nusbaum, H. C., & Pisoni, D. B. (1985). Some effects of training on the perception of synthetic speech. *Human Factors, 27,* 395–408.

Sebastian-Galles, N., Dupoux, E., Costa, A., & Mehler, J. (2000). Adaptation to time-compressed speech: Phonological determinants. *Perception & Psychophysics, 62,* 834–842.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995, October 13). Speech recognition with primarily temporal cues. *Science, 270,* 303–304.

Smith, D. R., Patterson, R. D., Turner, R., Kawahara, H., & Irino, T. (2005). The processing and perception of size information in speech sounds. *Journal of the Acoustical Society of America, 117,* 305–318.

Weill, S. A. (2001). *Foreign accented speech: Adaptation and generalization.* Unpublished master's thesis, Ohio State University, Columbus.

(*Appendixes follow*)

Appendix A

Stimuli in Experiment 1

Group	Word	Group	Word	Group	Word
A	bristle	A	movie	B	blossom
A	moth	A	hoof	B	harp
A	panic	A	vessel	B	maple
A	towel	A	cuisine	B	title
A	glove	A	vest	B	sultan
A	shovel	A	chrome	B	spool
A	rat	A	hero	B	wig
A	sardine	A	rung	B	saloon
A	beard	A	gem	B	shark
A	kitten	A	tomb	B	sleeve
A	wolf	A	nephew	B	beef
A	elm	A	toy	B	bone
A	rogue	A	corpse	B	keg
A	herb	A	beak	B	moss
A	spade	B	human	B	mule
A	heaven	B	petal	B	wool
A	meal	B	brass	B	thorn
A	siren	B	lung	B	prop
A	spirit	B	calf	B	garlic
A	knee	B	mast	B	stub
A	ribbon	B	turtle	B	polo
A	wallet	B	gallon	B	ash
A	tunic	B	snail	B	palace
A	chapel	B	spark	B	willow
A	nozzle	B	pest	B	diet
A	speck	B	honey	B	tennis
A	haze	B	cedar	B	hermit
A	latch	B	verb	B	tunnel
A	cult	B	monkey		
A	basket	B	guy		
A	walrus	B	basin		
A	fable	B	barn		
A	muzzle	B	crystal		
A	rabbit	B	noun		
A	wing	B	fellow		
A	column	B	web		
A	helmet	B	harvest		
A	bourbon	B	tulip		
A	verse	B	balloon		
A	weed	B	gang		
A	frog	B	bison		
A	insect	B	card		
A	gravel	B	flora		
A	porch	B	pouch		
A	hurdle	B	volume		
A	span	B	biscuit		

Appendix B

Stimuli in Experiment 2

Group	Test words	Training words	Training nonwords
A	honey	oak	treak
A	fellow	zoo	breck
A	towel	earl	kerrow
A	nozzle	fee	sperrea
A	vessel	elf	prief
A	diet	ant	berrye
A	polo	thief	crett
A	movie	booth	rost
A	gravel	gown	spope
A	column	bush	purps
A	panic	badge	frup
A	monkey	noose	yease
A	hermit	wharf	freeve
A	gallon	vase	circue
A	wallet	fog	croot
A	basket	van	peash
A	biscuit	dirt	seash
A	crystal	king	broy
A	tulip	goat	provel
A	helmet	pup	poth
A	ash	shawl	koth
A	moss	vine	vorce
A	wool	rib	hybrack
A	beef	chin	sodge
A	porch	cave	provail
A	rogue	dame	stipe
A	gem	kite	nopple
A	keg	pope	cogue
A	wig	doll	kollow
A	bone	rug	ush
A	gang	hen	lomb
A	beard	whiff	feam
A	tomb	gin	cong
A	sleeve	wheat	correr
A	spade	goal	pottle
A	spark	lard	cupe
A	mast	lane	neckrel
A	stub	wick	kise
A	chrome	cone	gloth
A	brass	lad	pite
B	fable	ranch	cousket
B	turtle	shrub	oap
B	title	scab	pisk
B	willow	tune	sharf
B	hurdle	cliff	looth
B	heaven	filth	adome
B	bison	tweed	kime
B	shovel	truce	harse
B	flora	plum	tove
B	cedar	broom	sightle
B	nephew	bench	umple
B	bristle	silk	rotch
B	garlic	crumb	dobe
B	ribbon	blouse	tooge
B	balloon	pint	nifle
B	cuisine	frill	ludge
B	spirit	troop	sar
B	harvest	blade	torm
B	volume	wand	dack
B	mountain	rust	sarak
B	knee	devil	parair
B	shark	bubble	mooth
B	haze	buckle	jettuce
B	lung	camel	nortle

(*Appendix continues*)

Appendix (*continued*)

Group	Test words	Training words	Training nonwords
B	verb	bible	phato
B	wing	whistle	blackarp
B	hoof	organ	corch
B	barn	noodle	gow
B	calf	channel	ingle
B	herb	atom	zomith
B	rat	autumn	ribe
B	elm	hockey	fackle
B	pouch	poet	cumic
B	verse	nickel	scuzzm
B	snail	chisel	stabe
B	corpse	lesson	taid
B	pest	axle	passon
B	glove	daisy	garl
B	cult	mutton	pla
B	span	cotton	drib
C	petal	orchid	yetteaze
C	muzzle	poison	scaider
C	kitten	python	relmick
C	tunnel	cable	craless
C	basin	barrel	rith
C	chapel	lemon	gicial
C	hero	herring	margy
C	maple	linen	wiom
C	sardine	canal	stuneck
C	saloon	pimple	shattel
C	tennis	pudding	arrtit
C	rabbit	bishop	crishine
C	bourbon	cabin	matten
C	palace	apron	royer
C	insect	pencil	wobing
C	sultan	havoc	gattle
C	blossom	soccer	plish
C	walrus	satchel	ellish
C	tunic	margin	slub
C	siren	spire	wartine
C	toy	tortoise	relat
C	guy	harpoon	stashy
C	noun	carpet	cardom
C	moth	furnace	thermen
C	meal	piston	blarmer
C	harp	planet	radick
C	thorn	mackerel	tiddle
C	latch	dungeon	rarn
C	rung	human	silloom
C	card	stomach	insipped
C	web	lantern	sorn
C	beak	squirrel	kank
C	weed	magnet	fown
C	wolf	reptile	bund
C	frog	mattress	iln
C	mule	sapphire	twiant
C	speck	canteen	gillar
C	prop	fountain	spinza
C	spool	velvet	ballan
C	vest	victim	feinten

Received June 12, 2006
Revision received March 13, 2007
Accepted May 28, 2007 ■