Journal of Experimental Psychology: Human Perception and Performance

Adaptive Plasticity in Speech Perception: Effects of External Information and Internal Predictions

Sara Guediche, Julie A. Fiez, and Lori L. Holt

Online First Publication, February 8, 2016. http://dx.doi.org/10.1037/xhp0000196

CITATION

Guediche, S., Fiez, J. A., & Holt, L. L. (2016, February 8). Adaptive Plasticity in Speech Perception: Effects of External Information and Internal Predictions. *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication. http://dx.doi.org/10.1037/xhp0000196

Journal of Experimental Psychology:
Human Perception and Performance

© 2016 American Psychological Association
0096-1523/16/$12.00 http://dx.doi.org/10.1037/xhp0000196

Adaptive Plasticity in Speech Perception: Effects of External Information and Internal Predictions

Sara Guediche and Julie A. Fiez
University of Pittsburgh and Center for the Neural Basis of Cognition, Pittsburgh, Pennsylvania

Lori L. Holt
Carnegie Mellon University and Center for the Neural Basis of Cognition, Pittsburgh, Pennsylvania

When listeners encounter speech under adverse listening conditions, adaptive adjustments in perception can improve comprehension over time. In some cases, these adaptive changes require the presence of external information that disambiguates the distorted speech signals, whereas in other cases mere exposure is sufficient. Both external (e.g., written feedback) and internal (e.g., prior word knowledge) sources of information can be used to generate predictions about the correct mapping of a distorted speech signal. We hypothesize that these predictions provide a basis for determining the discrepancy between the expected and actual speech signal that can be used to guide adaptive changes in perception. This study provides the first empirical investigation that manipulates external and internal factors through (a) the availability of explicit external disambiguating information via the presence or absence of postresponse orthographic information paired with a repetition of the degraded stimulus, and (b) the accuracy of internally generated predictions; an acoustic distortion is introduced either abruptly or incrementally. The results demonstrate that the impact of external information on adaptive plasticity is contingent upon whether the intelligibility of the stimuli permits accurate internally generated predictions during exposure. External information sources enhance adaptive plasticity only when input signals are severely degraded and cannot reliably access internal predictions. This is consistent with a computational framework for adaptive plasticity in which error-driven supervised learning relies on the ability to compute sensory prediction error signals from both internal and external sources of information.

Keywords: perceptual learning, lexical adaptation, perceptual recalibration, sensorimotor adaptation, speech adaptation

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Native listeners are remarkably attuned to subtle distributional regularities of the language community. Yet, it is common to encounter speech that deviates from these expected patterns. Accent, dialect, background noise, and other sources can shift acoustic speech signals relative to speech community norms. This can compromise the mapping of the acoustic signal onto meaningful sounds and words, leading to poor comprehension (Brouwer & Bradlow, 2014; Cooper, Brouwer, & Bradlow, 2015; Kalikow, Stevens, & Elliott, 1977).

Nonetheless, in many cases, only brief experience with such "distorted" speech input is needed to boost comprehension (Altmann & Young, 1993; Francis, Baldwin, & Nusbaum, 2000; Greenspan, Nusbaum, & Pisoni, 1988; Liss, Spitzer, Caviness, & Adler, 2002; Norris, McQueen, & Cutler, 2003; Pallier, Sebastian-Galles, Dupoux, Christophe, & Mehler, 1998; Schwab, Nusbaum, Pisoni, 1985). The improvements in speech perception generalize to similarly distorted utterances not heard previously; this adaptive plasticity[1] is observed across many different types of signal distortion, including time-compressed, noise-vocoded, foreign accented, and synthetic speech

Sara Guediche, Center for Neuroscience at the University of Pittsburgh and Center for the Neural Basis of Cognition, Pittsburgh, Pennsylvania. Julie A. Fiez, Department of Psychology and Center for Neuroscience at the University of Pittsburgh and Center for the Neural Basis of Cognition, Pittsburgh, Pennsylvania. Lori L. Holt, Department of Psychology, Carnegie Mellon University and Center for the Neural Basis of Cognition, Pittsburgh, Pennsylvania.

This work was funded by the University of Pittsburgh, Kenneth P. Dietrich School of Arts and Sciences, Andrew Mellon Predoctoral Fellowship; RO1 DC 00467 to Lori L. Holt; RO1 MH 59256 to Julie A. Fiez; and National Science Foundation Grant 1125719. We thank Fu and colleagues for providing Tigris, a program used to apply different levels of distortion to speech recordings. We also acknowledge the contributions made by Christi Gomez in synthesizing the stimuli and testing participants in Experiment 1. We are also thankful for the efforts of Rob Schwartz, Anthony Kelly, and Christi Gomez in data collection.

Correspondence concerning this article should be addressed to Sara Guediche, who is now at the Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI 02912. E-mail: Sara_Guediche@brown.edu

[1] We use the term *adaptive plasticity* to describe flexibility in responding and adapting to distortions in the acoustic realization of speech, after a brief period of exposure. Norris and colleagues use the term perceptual learning (Norris et al., 2003) to describe the phenomenon, and others have referred to the phenomenon as lexical adaptation (Maye, Aslin, & Tanenhaus, 2008), perceptual adaptation (Li & Fu, 2007), and perceptual recalibration (Vroomen, van Linden, de Gelder, & Bertelson, 2007). Different terminologies have often been used in association with specific theoretical or computational accounts, or simply with different tasks. Regardless of the label, it is important to note that the learning mechanisms are believed to be distinct from those associated with the phenomenon of selective adaptation (Norris et al., 2003; Vroomen et al., 2007): selective adaptation is a change in the perception of a phonetic boundary that is thought to reflect the fatigue of feature detectors or auditory contrast mechanisms (Diehl, Elman, & McCusker, 1978; Eimas & Corbit, 1973).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

(e.g., Altmann & Young, 1993; Bradlow & Bent, 2008; Clarke & Garrett, 2004; Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Hervais-Adelman, Davis, Johnsrude, & Carlyon, 2008; Schwab et al., 1985).

The mechanisms that support adaptive plasticity in speech perception are not yet understood in detail. However, several factors appear to play important roles. Chief among them is the presence of information that disambiguates the distorted speech. For example, a period of experience with distorted spoken sentences accompanied by external information consisting of a written version of the sentence, or a clear acoustic presentation of the sentence, results in greater comprehension improvements than exposure without such disambiguating information (Davis et al., 2005). Supportive information of this sort can be provided through visual presentation of the orthographic form (e.g., Francis, Nusbaum, & Fenn, 2007; Loebach, Pisoni, & Svirsky, 2010; Schwab et al., 1985), or presentation of an undistorted acoustic instance of the speech, either subsequent or prior to presentation of the distorted speech (Hervais-Adelman et al., 2008).

Nonetheless, adaptive plasticity is sometimes apparent when there are no external disambiguating information sources. Mere exposure to systematic distortions to natural speech acoustics—such as those arising from a nonnative accent, dysarthria, or time-compressed speech—may also lead to improvements in listeners' processing efficiency and comprehension (Altmann & Young, 1993; Clarke & Garrett, 2004; Liss et al., 2002; Pallier et al., 1998). What guides adaptive changes in perception in such cases?

Vroomen and colleagues (2007) note that that the disambiguating auditory information from a spoken lexical context and the visual information from an articulating face both lead to adaptive changes in speech perception. They suggest that a common mechanism may account for each case. Specifically, they propose that when disambiguating information is available with distorted speech input, a mismatch or "conflict" may be detected. The proposal is that an internal error signal is generated when there is a discrepancy between the input predicted by the disambiguating information and the actual (distorted) speech that is experienced (see Guediche, Blumstein, Fiez, & Holt, 2014). The internally generated error signal may in turn guide adaptive plasticity.

If distorted speech input is sufficient to at least partially access prior knowledge (e.g., possible words, sound categories, etc.), it may generate internal predictions about the typically expected acoustic input associated with those categories or words, serving as a basis for calculating error signals. *Internally generated* predictions may then provide a basis for adaptive plasticity when there is no external disambiguating information.

Consistent with this hypothesis, more intelligible speech generally leads to greater adaptive plasticity (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Liss et al., 2002; Pallier et al., 1998; Peelle & Wingfield, 2005). Moreover, this hypothesis has grounding in other domains. For example, otherwise unrecognizable low spatial frequency images can be recognized as an exemplar of an object category (Bar, 2003) or a general scene "gist" (Oliva & Torralba, 2007) when other aspects of the sensory input are sufficient to guide predictions about the identity of the image from context (Panichello, Cheung, & Bar, 2013).

The possibility that predictions from external disambiguating information and internally generated predictions from prior knowledge will each evoke error signals suggests that each will contribute to the degree of adaptive plasticity. Guediche et al. (2014) note that many studies that report adaptive plasticity in the absence of external information or feedback during training have tended to use less severe distortion manipulations that have a less detrimental effect on stimulus intelligibility (e.g., Altmann & Young, 1993), as compared with studies that provide external information sources during training (e.g., Schwab et al., 1985). However, the relative contributions of internally generated predictions and external disambiguating information in evoking adaptive plasticity has yet to be examined explicitly.

In the present study, we manipulate speech intelligibility as a means of influencing the accuracy of internally generated predictions based on prior linguistic knowledge (e.g., lexical information). The ability to map speech input onto linguistic knowledge involves a complex alignment of many acoustic properties and the relationships among them. The greater the alignment between the incoming speech signal and these mappings, the more intelligible the signal (see Gentner, 1983; Gentner & Markman, 1997, for similar examples from another domain). Highly distorted speech violates the alignment of incoming acoustic information with linguistic knowledge, reducing intelligibility. We suggest that adaptive plasticity, which can be thought of as a realignment of the mapping from input to linguistic knowledge, is supported by the incremental introduction of the distortion.

The logic is that mild distortions better align with established mappings than severe distortions. This results in greater success in accessing linguistic knowledge and deriving an internally derived "prediction" of the typical mapping. By introducing the distortion incrementally, the probability of successful predictions is increased and progressive alignment (Church, Mercado, Wisniewski, & Liu, 2013; Sumner, 2011) allows transfer as the realignment of the mapping is extended to successively greater levels of distortion. With this, there is the opportunity to derive an error signal from the discrepancy of the current and predicted mappings guiding adaptive plasticity at higher levels of distortion.

The alignment of highly distorted speech to linguistic knowledge is poor. As a result, highly distorted signals violate the mappings established by long-term experience, and evoke less accurate internal predictions with which to drive adaptive plasticity. In this case, external information that supports the mapping to linguistic knowledge provides an alternate means of determining discrepancies and should exert an especially strong effect on adaptive plasticity.

In this way, we predict an interaction between stimulus intelligibility and the effectiveness of external disambiguating information in adaptive plasticity of speech perception. Incremental introduction of the distortion allows for progressive alignment of the mappings and more accurate internal predictions. For this reason, we expect the influence of external disambiguating information (e.g., a visual orthographic presentation of the word) to be less influential for incremental introduction of distortions compared to the abrupt introduction of the severe distortion.

To examine this prediction, we exploit a signal processing technique from the cochlear implant literature that systematically affects speech intelligibility and allows for incremental introduction of increasingly severe signal distortions (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Zeng, 2004). In Experiment 1, we characterize the relationship between the severity of a noise-vocoded, spectrally shifted speech distortion[2] and intelligibility, as measured by word recognition accuracy. Using these stimuli in Experiments 2 and 3, we assess adaptive plasticity via pre- to posttraining improvements in word recognition accuracy for the most severe acoustic signal distortion. Experiment 2 employs a standard paradigm to measure adaptive plasticity in which participants experience the same severe distortion (severe condition) throughout training that is tested at the pretest and posttest, with a unique word presented on every trial. In Experiment 3, we employ a paradigm in which the severity of the distortion is gradually incremented (incremental condition) every 10 trials. Thus, intelligibility systematically decreases, and consequently the accuracy of internally generated predictions systematically decreases, until it eventually reaches the severe distortion level assessed at pre- and posttests. Here, too, all words within the experiment are unique. In both Experiments 2 and 3, we assess the influence of external disambiguating information by providing concurrent postresponse presentation of the written form of the stimulus paired with a repeated auditory presentation of the distorted stimulus (e.g., Fenn, Nusbaum, & Margoliash, 2003; Greenspan et al., 1988; Schwab et al., 1985). This form of feedback allows the discrepancy to be determined without resorting to the preresponse memory of the acoustic signal. We predict that the availability of external sources of disambiguating lexical information during exposure will evoke the greatest adaptive plasticity when internally generated predictions are unlikely to be accurate (i.e., when the signal distortion is severe) and will be less powerful when internally generated predictions are more accurate (i.e., when the signal distortion is incremental). We expect both the presence of external disambiguating information and the signal distortion type to contribute to adaptive plasticity in a potentially redundant fashion revealing an interaction on adaptive plasticity between the two factors.

Experiment 1

The goal of Experiment 1 is to establish the nature of the relationship between distortion severity and speech intelligibility for a noise-vocoded, spectrally shifted distortion manipulation of single-syllable English words, as measured through open-set word recognition across 25 levels of distortion severity. This establishes the relationship of the distortion severity manipulation to be used in Experiments 2 and 3 with speech intelligibility, as a proxy for the accuracy of internally generated predictions.

Method

Participants. Twenty undergraduate-aged native-English participants with normal hearing from the Pittsburgh area participated in a brief listening test. Participants were paid $7/hr and the protocol followed Carnegie Mellon University Institutional Review Board (IRB) regulations.

Stimuli. Stimuli were based on the phonetically balanced word lists described by Egan (1948; Lists 1–7). Each list was composed of 50 phonetically balanced English monosyllabic words. A female monolingual English talker (Lori L. Holt) uttered each word into an Electrovoice RE 20 microphone connected to a digital Marantz PMC670 recorder with 16-bit resolution at a sampling rate of 22,050 Hz (Fairport, NY). Stimuli were saved as individual *.wav files, and equated across all lists for root mean square amplitude.

Distortions to these natural speech signals were introduced using Tiger Speech (http://www.tigerspeech.com/tst_tigercis.html; see Fu & Galvin, 2003, and Li and Fu, 2007, for detailed methods). Each spoken word was band-pass filtered into 20 frequency bands using eighth-order Butterworth filters with 24 dB/octave falloff. The frequency characteristics of the bands were based on the Nucleus-24M cochlear implant (tiling frequencies from 116–7871 Hz). Each band was half-wave rectified to extract the temporal envelope and low-pass filtered at 160 Hz. This envelope served to modulate a carrier band. The following equation was used to calculate the frequency range of the carrier band where P0 is the insertion depth of the theoretical implant equivalent, and p(i) is the corresponding frequency shift introduced to the 20 frequency bands:

$$p(i) = P0 + 0.75 * i, \ i = 0, 1, 2, 3 \ldots 20$$

The carrier band frequencies assumed a 35-mm cochlea (Greenwood, 1990) and were calculated using the following equation:

$$f(i) = 165.4 * (10p(i) * 0.06 - 0.88)$$

The two equations were combined to determine the corner frequencies of carrier bands for a given insertion depth. Cross-over attenuation between adjacent bands was −3 dB. The carrier bands were summed to create the spectrally shifted speech tokens. For the purposes of the present experiment, P0 varied between 9.25 and 15.25 mm from the apex of the cochlea, simulating a range of 9.25-mm to 15.25-mm insertion depths and shifting speech spectra incrementally upward across the frequency dimension. Carrier bands incrementing in steps of 0.25-mm insertion depth were calculated between these endpoint values. For example, at an insertion depth of 9.25 mm, speech was shifted upward in frequency such that there was no spectral energy below 448 Hz (see Figure 1). At the most severe distortion, 15.25 mm, there was no spectral energy below 1214 Hz (see Figure 1). These signal distortions create a complex mapping challenge for word recognition because a great deal of information in the speech signal is carried below 2000 Hz (Fant, 1949).

Procedure. Listeners were seated in individual sound-attenuated booths in front of a computer monitor. On each trial, they heard a single acoustic presentation of a distorted word over headphones (Beyer DT-150; approximately 70 dB) and typed what they heard on a computer keyboard. Participants were informed that a real, monosyllabic English word would be presented on each trial and they were instructed to guess if they were unsure of the identity of the word; their responses were otherwise completely

[2] The noise-vocoded speech distortion used in the current study differs from noise-vocoding approaches that have been used in prior studies of adaptive plasticity in speech perception (e.g., Dahan & Mead, 2010; Hervais-Adelman et al., 2008; Shannon et al., 1995). In addition to separating the speech into band-delimited channels and using these channels to modulate noise bands, as has been the case in prior studies, a second step in signal processing is introduced. In this step, we frequency-shift the band-limited channels (Shannon et al., 1995; Zeng, 2004). This manipulation allows us to systematically manipulate the degree of stimulus distortion; the greater the difference between the original band-delimited channels and the frequency-shifted channels, the greater the signal distortion and the poorer the intelligibility (see Experiment 1).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

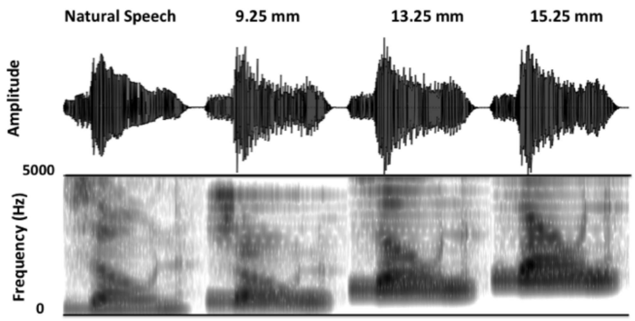

Figure 1. Example of waveforms (amplitude, top) and spectrograms (frequency, with amplitude in gray scale, bottom) at different levels of distortion for one of the words in the set, "zone." From left to right: undistorted stimulus is on the left, stimulus at a mild distortion level of (9.25 mm), stimulus at the moderate distortion (13.25 mm), and stimulus at the severe distortion (15.25 mm).

unconstrained. Participants were allowed up to 4 s to type their responses after each item was presented. Responses were echoed on the computer monitor and participants were allowed to backspace in order to correct responses in the event of a typing error. Once participants were satisfied with each response, they pressed the "Enter" key to register the response and the next stimulus was presented over the headphones.

Participants heard words at distortion levels from 9.25-mm to 15.25-mm insertion depths (sampled in 0.25-mm steps), randomly ordered. Words were randomly selected from the total set of 350 words, presented once, and never repeated across trials. At each of the 25 distortion levels, listeners heard 14 words for a total of 350 randomly ordered trials. Following the scoring approach of previous work (Greenspan et al., 1988), participants' responses were scored as correct only if the response exactly matched the spoken word. Homonyms were accepted, but responses were scored as incorrect if listeners heard any missing or additional phonemes. For example, when *road* was presented a response of *road* or *rode* or *rowed* would be scored as correct, but *row*, *crowed*, *load*, or *roam* would be scored as incorrect.

Results and Discussion

Listeners' word recognition accuracy was superior for less severe, compared with more severe, distortions. Distortion severity, quantified as millimeters of simulated insertion depth as a proxy for the degree of mismatch between input and output filters in the signal processing algorithm, was significantly negatively correlated with word recognition accuracy, $R = -.78$, $p < .001$. This relationship between distortion severity and word recognition is depicted as an intelligibility curve in Figure 2. Curve estimation shows that the relationship between the insertion depth and intelligibility is significant for several equation models including quadratic ($R = -.96$) and linear functions ($R = -.78$), $p < .001$, thus demonstrating a systematic relationship between intelligibility and distortion severity. At the 9.25-mm insertion depth, open set mean word recognition accuracy was 56% ($SEM = 4.38$), whereas at the most severe 15.25-mm insertion depth, word recognition accuracy was only ($M = 11\%$, $SEM = 1.41$).

These findings provide quantitative support that it is possible to systematically influence speech intelligibility through a noise-vocoded, spectrally shifted distortion, and assure that greater shifts in input-output channel frequencies produce less intelligible speech. This manipulation of speech input thus serves as a testing ground for examining the effect of intelligibility on adaptive plasticity. In particular, it allows for the severity of the distortion to be systematically incremented to affect intelligibility and, correspondingly, listeners' ability to make accurate linguistic predictions; at the 15.25-mm insertion depth, internal predictions (e.g., lexical predictions) will be least accurate, whereas at the 9.25-mm depth internal predictions are more likely to be accurate.

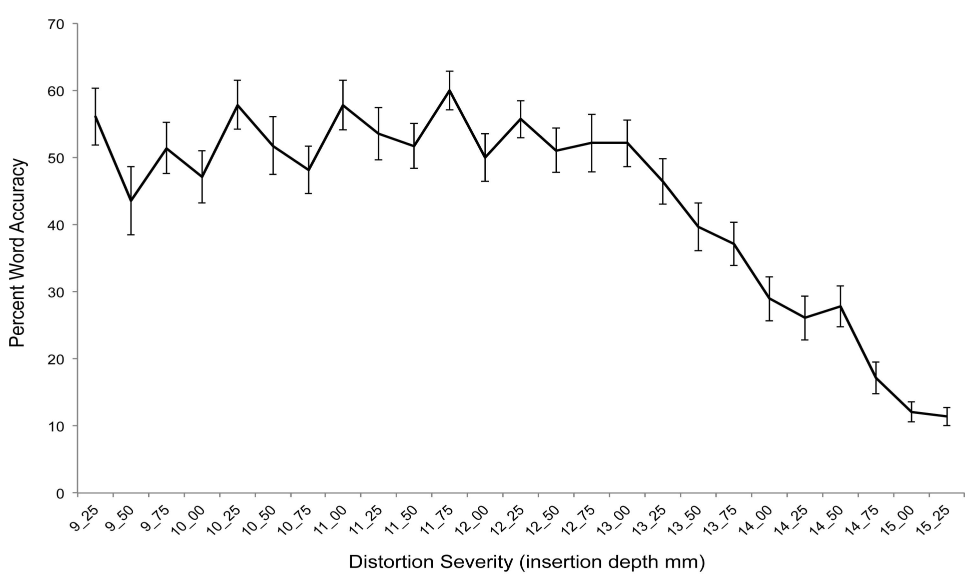

Figure 2. Experiment 1 results showing percent word recognition at each distortion level. Error bars represent standard error of the mean over subjects.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

In the following experiments, we use these stimuli to test the hypothesis that the accuracy of linguistic predictions during exposure, as indexed by the intelligibility of the speech distortion, impacts adaptive plasticity.

Experiment 2

Experiment 2 examines adaptation to a severe distortion using a traditional exposure paradigm whereby distortion severity is held constant during the entire period. Specifically, we used the most severe distortion (15.25 mm) from Experiment 1, for which accurate predictions were least likely. Thus, mere exposure to this severe distortion should yield poor internal predictions and limited adaptive plasticity. However, we predict that the presence of external information during exposure will boost adaptive plasticity (as measured by improvements in word recognition performance from pre- to posttest) compared with exposure without such information. Inspired by previous studies of adaptive plasticity (Fenn et al., 2003; Greenspan et al., 1988; Schwab et al., 1985), exposure to the written form of the acoustically presented word paired with a repetition of the distorted stimulus following each response serves as the external disambiguating information source.

Method

Participants. Seventy-eight native English participants with normal hearing were recruited and paid for their participation. All participants gave informed consent prior to participation, in accordance with the IRB regulations of Carnegie Mellon University. Four participants were excluded due to technical errors, three participants were excluded for failure to properly follow the instructions, and three participants were excluded for not meeting the initial exclusion criteria of native English speaker. Two additional participants were excluded because they were bilinguals. Of the remaining 66 participants, 33 participants were in the *external-present* lexical information training group and 33 participants were in an *external-absent* generated lexical information training group.

Stimuli. Stimuli were based on the phonetically balanced word lists described by Egan (1948; Lists 1–9). Each list was composed of 50 phonetically balanced English monosyllabic words. Stimuli that were pronouns or plurals were removed from the lists across the different distortion levels, leaving 380 words in the selection pool. From this pool, 350 words were selected and presented in a random order, with each item presented only once for each participant. Only the most severe distortions synthesized at the 15.25-mm insertion depth were used in this experiment.

Procedure. Each listener participated in pretest, training and posttest segments of the experiment. The duration varied according to participant response times, lasting approximately 40 min.

The experimental session began with a 50-trial pretest. Stimuli were drawn randomly from the total set of 380 words, with each selected item removed from the stimulus pool to preclude its selection on another trial. Following the pretest participants progressed on to the exposure phase of the experiment. During this segment each participant heard 250 words acoustically distorted at the 15.25-mm distortion level. The items were randomly drawn from the total set of items that remained in the stimulus pool, with each item removed from the pool after its selection. Participants were randomly assigned to one of two training conditions that were defined by the presence or absence of external disambiguating lexical information. On each trial, participants from each condition heard a single word and typed their responses on a standard keyboard. Participants were informed prior to the experiment that all trials would be real, monosyllabic English words, but responses were otherwise open set. After indicating satisfaction with their word recognition response by pressing "Enter," participants in the *external-present* condition saw the correct printed form of the word and heard a repeated presentation of the same distorted auditory word. Following a brief delay, the next stimulus item was presented. This mode of external information has been used in many previous studies of the adaptive plasticity of speech perception (Fenn et al., 2003; Francis et al., 2000; Francis et al., 2007; Greenspan et al., 1988; Schwab et al., 1985). Participants in the *external-absent* condition also heard 250 randomly selected words processed at the 15.25-mm distortion depth, but received no external information about the accuracy of their responses (no printed form of the word nor a repeated auditory presentation). The experiment simply progressed to the next trial after participants pressed the "Enter" key on each trial.

The experimental session concluded with a 50-trial posttest of words processed at the 15.25-mm depth without the provision of external information in the form of the written word. Stimuli were drawn randomly from the remaining pool of stimulus items, with each selected item removed from the stimulus pool to preclude its selection on another trial. Learning resulting from the intermediate training segment was assessed as a change in word recognition accuracy from the pretest to the posttest. Each word encountered during pretest, training, and posttest was unique for a given participant, therefore, changes in word recognition accuracy cannot arise due to word repetition, memorization of prior responses or feedback, or learning an explicit mapping from the distorted input to a particular lexical item. Instead, changes in word recognition from pre- to posttest must reflect adaptation at the level of prelexical representations that generalizes across different words (see Schwab et al., 1985; Greenspan et al., 1988).

Results

The results of Experiment 2 are summarized in Table 1 and Figure 3. Mean performance on the pretest in this experiment was lower than observed for the same distortion level in Experiment 1 (9% vs. 11%, respectively). However, this difference did not reach significance. Because the purpose of Experiment 1 was to establish the general relationship between performance and distortion level, any minor baseline difference in individual performance between Experiment 1 and Experiment 2 should not have an adverse impact on the ability to draw conclusions from Experiment 2.

To probe for learning effects, we conducted a 2×2 analysis of variance (ANOVA) with test (pretest, posttest) as a within-subject factor, external information (present, absent) as a between-subjects factor, and proportion correct word recognition as the dependent measure. We found a main effect of test, $F(1, 64) = 134.41$, $p < .001$, $\eta_p^2 = .68$, a main effect of external information, $F(1, 64) = 5.38$, $p = .024$, $\eta_p^2 = .078$, and a significant interaction between test and external information, $F(1, 64) = 14.79$, $p < .001$, $\eta_p^2 = .19$. The degree of adaptive plasticity, measured as the difference between word recognition accuracy on posttest versus pretest, when listeners were provided with external information during

Table 1
Percent Word Recognition Accuracy in Experiments 2 and 3

Experiment	Condition	Pretest	Posttest
Experiment 2 (severe)	External-absent	8.0% (1.16)	16.30% (1.81)
	External-present	9.27% (1.53)	25.82% (2.48)
Experiment 3 (incremental)	External-absent	8.81% (1.19)	22.88% (2.48)
	External-present	10.41% (1.99)	25.82% (1.73)

Note. The table shows the mean percentage of correctly recognized words for the pretest and posttest in each of the conditions in Experiments 2 and 3. Standard errors of the means over participants are reported in parentheses.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

exposure (M = 16.55%, SEM = 1.85) was significantly more than when external information was not provided (M = 8.30%, SEM = 1.08), $t(64) = 3.85$, $p < .001$, 95% confidence interval (CI) [.04, .13]. This suggests that external information during training can boost adaptive plasticity when intelligibility of the speech distortion is poor.

In order to examine the relationship between the accuracy of internally generated lexical predictions and the degree of adaptive plasticity, we performed a correlation analysis using individual measures of improvements in word recognition between pretest and posttest and word recognition accuracy during the training phase. We found a significant correlation ($R = 0.56$, $p = .001$) in the external-absent and ($R = 0.60$, $p < .001$) in the external-present condition (see Figure 4).

Discussion

In this experiment, we examined whether the presence of external disambiguating information during training (following each response) affected the degree of adaptive plasticity, as measured by word recognition accuracy to severely distorted speech before versus after training. The same severe distortion level (15.25 mm) was used in the pretest, training, and posttest. The results show that adaptive plasticity (as measured by improvements in word recognition) was enhanced by the presence of external disambiguating information during exposure to severely distorted stimuli. These findings indicate that externally provided sources of information contribute to adaptive plasticity, at least when speech distortions are severe, intelligibility is low, and the distortion is presented abruptly.

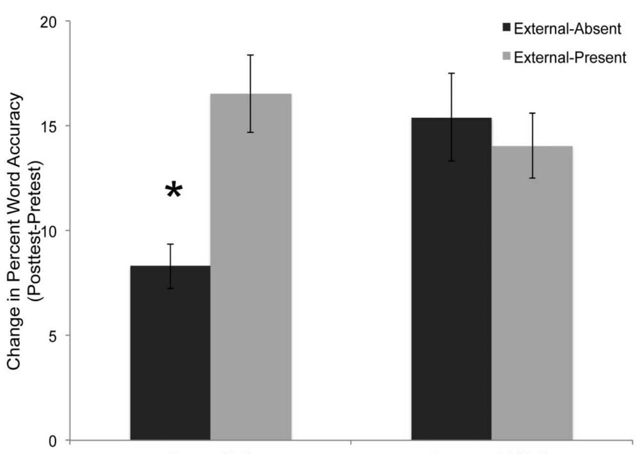

Figure 3. The degree of adaptive plasticity across conditions, measured as the difference in the percent of correctly recognized severely distorted words on the posttest compared with the pretest. Error bars represent standard error of the mean over subjects. E2 = Experiment 2; E3 = Experiment 3.

The form of external feedback provided in the external-present condition has been commonly used in studies of adaptive plasticity in speech perception (Schwab et al., 1985; Greenspan et al., 1988). It allows the degraded acoustic stimulus to be paired concurrent with the orthography, enabling the opportunity to detect the discrepancy between the intended and actual speech input (the error signal that is hypothesized to drive adaptive plasticity). A potential concern is that external information includes both the written form of the word and a repeated presentation of the item. The repeated presentation of the item may create more opportunity for internal sources of information to drive adaptive plasticity in external-present, as opposed to external-absent, conditions and make the contribution between internal and external factors difficult to disentangle. However, if repeated exposure within the external-present condition contributes substantively to adaptive plasticity then the effect of external information should not differ as a function of the signal distortion severity. We will test this directly with the results of Experiment 3.

In Experiment 3, we examine the extent of adaptive plasticity when the distortion severity is gradually introduced during training (incremental) as a function of whether external disambiguating information is present or absent. Experiment 1 showed a relationship between the distortion level and intelligibility: Lower distortion levels were more intelligible. Said another way, participants were better able to map mild distortions to linguistic knowledge than severe distortions. This is an indication that internally generated predictions would be more accurate for mild, as opposed to severe, distortions. An incremental presentation of the distortion (mild to severe) therefore would allow more accurate internal predictions during exposure (see Experiment 1). As the system realigns the mapping via supervised learning driven by accurate predictions, the likelihood of accurate predictions to severely distorted speech presented at posttest will increase. We expect that internally mediated predictions may be sufficient to drive adaptive plasticity and, further, that as a result the influence of external information that supports accurate predictions will be less influential.

Experiment 3

Adaptive plasticity in speech perception is sometimes apparent even when no external sources of information are available (e.g., Adank & Janse, 2009; Altmann & Young, 1993; Clarke & Garrett, 2004; Liss et al., 2002). In such cases, accurate internally gener-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

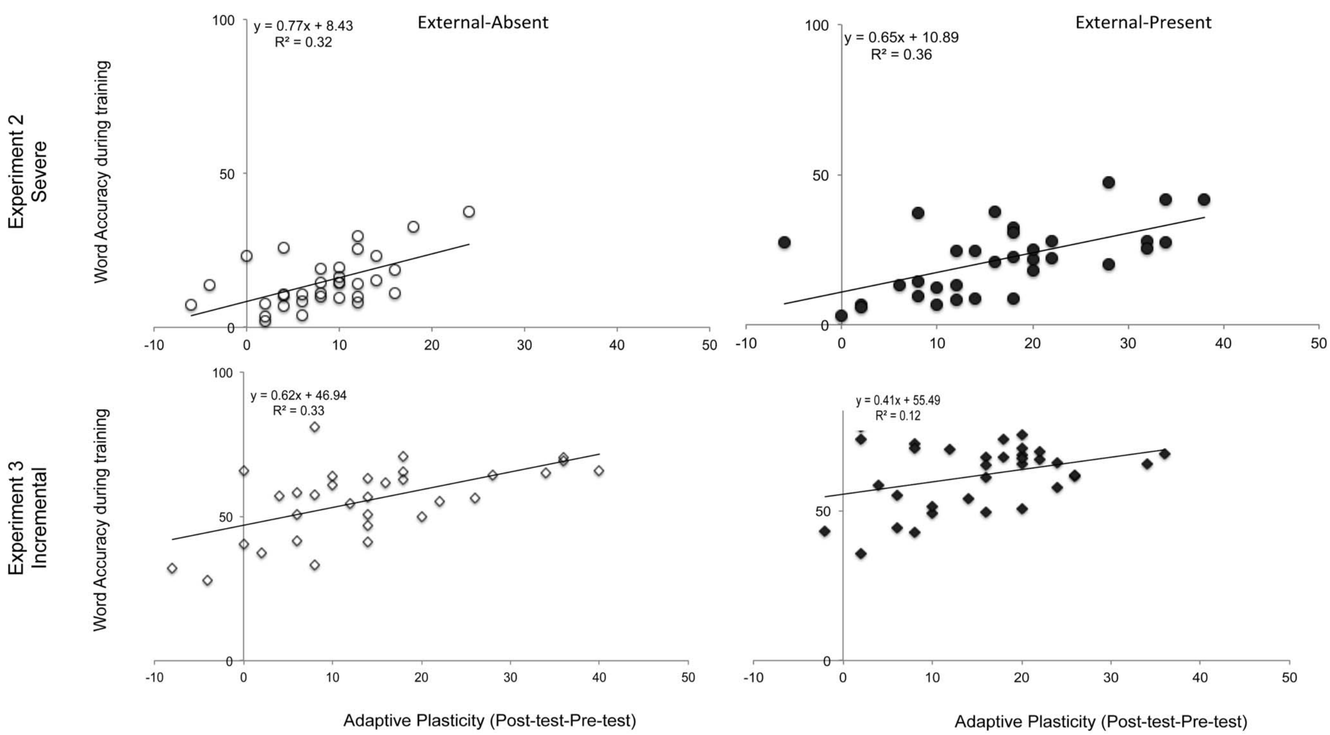

Figure 4. Plots show correlations between adaptive plasticity (measured as the improvement in word recognition accuracy from pretest to posttest, *x*-axis) and word recognition accuracy during training (as a measure of accurate lexical predictions, *y*-axis) for individual participants, for each condition. Top panel shows the results of Experiment 2. Bottom panel shows the results of Experiment 3.

ated predictions may mediate adaptive plasticity. In Experiment 3, we directly examine the role of internally generated linguistic predictions by manipulating the intelligibility of the stimuli experienced in the training segment of the experiment. We incrementally introduce the speech distortion such that intelligibility decreases gradually across training. Based on the systematic relationship between distortion severity and intelligibility observed in Experiment 1, we expect this to result in more accurate predictions than was possible for the severely distorted speech experienced in training in Experiment 2. This approach permits accurate predictions to be generated at the outset of training, which we hypothesize provide internal information that can support adaptive plasticity. Consequently, continuous adaptation allows for continued intelligibility despite successively greater distortion levels. The goal of this experiment is to examine a speech distortion that promotes accurate predictions while also manipulating the presence of external (present, absent) sources of information during training. If accurate internally generated predictions contribute to adaptive plasticity via the same kinds of prediction-driven learning mechanisms as those supported by external disambiguating information, then the presence of external information may provide little additional adaptation benefit when accurate internal information is already available.

Method

Participants. Seventy-two native-English participants with normal hearing were recruited to participate in this study. All participants gave informed consent prior to participation, in accordance with the IRB regulations of Carnegie Mellon University. Two participants were excluded due to technical errors and three were excluded for not meeting the initial exclusion criteria of nonnative English speaker or for not properly following instructions. One additional participant was excluded because he or she was a proficient mandarin bilingual (prekindergarten). Of the remaining 66 participants, 34 were in the *external-present* provided condition and 32 participants were in the *external-absent* condition.

Stimuli. The same base words that were used in Experiments 1 and 2 were also used in this experiment. However, each word was distorted at each of the 25 different insertion depths. A total of 350 base word stimuli were randomly selected from the set of 380 words. The distortion level for each trial chosen is described below.

Procedure. Each listener participated in pretest, training and posttest segments of the experiment using the same basic procedures described in Experiment 2. The only difference was the procedure used for stimulus selection during training, as described below. Pre- and posttests were identical to those of Experiment 2.

Following the pretest segment, in which each participant heard 50 words processed with a 15.25-mm insertion depth, each participant was assigned to either the *external-present* or *external-absent* condition. The two training conditions were the same as those described in Experiment 2, with the exception that the stimuli were presented at incrementally increasing levels of distortion across the

250 training trials. The first 10 words were processed at the 9.25-mm insertion depth (for which intelligibility is relatively high; mean 56% in Experiment 1). After each 10 trials, signal distortion was increased by an insertion depth of 0.25 mm, ending with 10 trials at the 15.25-mm depth. As in Experiment 1, no words were repeated across pre- and posttests or training. The words presented at each insertion depth were randomly selected (without replacement) from the larger pool.

Results and Discussion

The results of Experiment 3 are summarized in Table 1 and Figure 3. Mean performance on the pretest in this experiment was once again slightly lower than observed for the same distortion level in Experiment 1 (10% vs. 11%, respectively), however this difference failed to reach significance. To examine learning effects, we conducted a 2 × 2 ANOVA using test (pretest, posttest) as a within-subject factor, external information (present, absent) as a between-subjects factor, and word recognition accuracy as the dependent measure. We observed a main effect of test (pretest, posttest), $F(1, 64) = 129.60$, $p < .001$, $\eta_p^2 = .67$. Neither the effect of external information, $F(1, 64) = 0.95$, $p = .33$, $\eta_p^2 = .015$, nor the interaction between test and external information, $F(1, 64) = 0.27$, $p = .60$, $\eta_p^2 = .004$ were significant. This indicates that adaptive exposure to the distortion without disambiguating external information during training indeed can drive adaptive plasticity, ($M_{Post\text{-}Pre} = 14.06\%$, $SEM = 2.01$). Moreover, the presence of external disambiguating information during training did not confer additional benefits beyond exposure alone, ($M_{Post\text{-}Pre} = 15.41\%$, $SEM = 1.57$), $t(64) = 0.52$, $p = .6$, 95% CI [−.04, .07].

In order to directly test the relationship between the accuracy of internally generated lexical predictions and the degree of adaptive plasticity, we performed a correlation analysis using individual measures of improvements in word recognition between pretest and posttest and word recognition accuracy during the exposure phase. We found a significant correlation ($R = 0.57$ 0.3, $p = .001$) in the external-absent condition and ($R = 0.34$, $p < .05$) in the external-present condition. This suggests that the accuracy of lexical predictions during exposure is related to the degree of adaptive plasticity, as evidenced by improvements in speech intelligibility to severe acoustic distortions.

Cross-Experiment Analyses

A primary aim of the study was to investigate whether internally generated predictions and externally provided disambiguating information contribute to adaptive plasticity. To address this question, we conducted analyses across Experiments 2 and 3. We first conducted a one-way ANOVA on the pretest accuracy scores with the four possible experimental conditions spanning Experiments 2 and 3 as between-subjects factors severe/external-present, severe/external-absent, incremental/external-present, incremental/external-absent. There was no significant difference in pretest accuracy across conditions in Experiment 2 and 3 ($p = .68$), supporting posttest versus pretest difference scores across conditions as a meaningful index of adaptive plasticity.

To investigate the statistical interaction between the presence of external disambiguating lexical information and the degree of signal distortion, we conducted a 2 × 2 × 2 ANOVA with test (pretest, posttest) as a within-subjects variable, external information (present, absent) as a between-subjects variable and distortion condition (severe, Experiment 2; incremental, Experiment 3) as a between-subjects variable. The analysis revealed a main effect of test, $F(1, 128) = 261.24$, $p < .001$, $\eta_p^2 = .67$, indicative of comprehension gains as a function of training. There was a main effect of external information, $F(1, 128) = 5.42$ $p = .02$, $\eta_p^2 = .041$, consistent with the supportive influence of externally provided disambiguating lexical information. In addition, there was a significant interaction between test and external information, $F(1, 128) = 8.14$ $p = .005$, $\eta_p^2 = .06$, indicating that the degree of adaptive plasticity observed from pre- to posttest was modulated by the presence of external lexical information. The main effect of distortion condition, $F(1, 128) = 1.68$, $p = .20$, $\eta_p^2 = .013$, the interaction between external information and distortion condition, $F(1, 128) = 0.90$, $p = .35$, $\eta_p^2 = .007$, and the interaction between test and distortion condition, $F(1, 128) = 1.89$, $p = .17$, $\eta_p^2 = .015$, were not significant.

We hypothesized that if internally generated lexical predictions support adaptive plasticity in speech perception, then the accuracy of lexical predictions (manipulated as a function of distortion severity) should interact with the presence of external disambiguating information. The significant three-way interaction supports this hypothesis, $F(1, 128) = 4.21$, $p = .04$, $\eta_p^2 = .032$. This indicates that the degree of adaptive plasticity is affected by the accuracy of internal lexical predictions (manipulated via distortion severity, Experiments 2 vs. 3) and the presence of external disambiguating lexical information. Indeed the accuracy during training in Experiment 2 ($M = 18.27\%$, $SEM = 1.32$) was significantly lower than in Experiment 3 ($M = 58.81\%$, $SEM = 1.48$), $t(130) = -20.44$, $p < .001$. 95% CI [−.44, −.36]. The two appear to contribute in a redundant fashion such that either source of information is sufficient for learning, yet not mutually additive in their effect on the degree of adaptive plasticity. Specifically, when the input is severely distorted such that internal lexical predictions are less accurate, adaptive plasticity in speech perception is supported by the presence of external disambiguating information. However, external information does not benefit adaptive plasticity when internal lexical predictions are fairly accurate. The adaptive plasticity observed for the incremental introduction of the speech distortion with no external disambiguating information to support adaptive plasticity (Experiment 3, external-absent) was indistinguishable from that observed when a severe speech distortion was accompanied by external disambiguating information (Experiment 2, external-present), $t(63) = -.89$, $p = .38$, 95% CI [−.08, .03]. Taken together, the results across Experiment 2 and 3 indicate that accurate lexical predictions can provide an internally generated source of information induced by the stimulus that contributes to adaptive plasticity and diminishes the additional benefit of externally provided disambiguating information.

General Discussion

Adult listeners are experts at mapping complex and highly variable acoustic speech signals to meaning, but signal distortions from accent, regional dialect, speech impairment and other sources can negatively impact speech intelligibility. However, under some conditions listeners rapidly adapt to systematic distortions in speech input; intelligibility improves for subsequent exposure to

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

the distortion. Although a great deal of research has been directed toward understanding the nature of this adaptive plasticity, there remain important open questions.

In the present study we examined the possibility that *internally generated* sources of disambiguating information derived from prior knowledge such as lexical knowledge or other linguistic sources of information, may serve a role similar to external disambiguating lexical information, such as the presence of a printed form of the word. We reasoned that if distorted speech input is sufficient to at least partially access prior knowledge about sound categories and words, it will be used to derive internal predictions about the typically expected input, which, in turn, may generate an error signal when the actual (distorted) input mismatches the prediction. Because lower signal distortions should lead to greater accuracy of internally generated predictions (observed as greater intelligibility), external information sources should play a less prominent role in driving adaptive plasticity for less severe, compared with more severe, signal distortions. This would provide a basis for adaptive plasticity when no external disambiguating information is available, as has been observed in studies involving mere presentation of certain types of signal distortions, such as time-compressed speech (Altmann & Young, 1993; Peelle & Wingfield, 2005). However, when internally generated predictions are less accurate as is the case for the severe condition, external information can be used to determine discrepancies and generate error signals that drive learning.

To address this hypothesis, we examined the interaction between signal distortion severity and the presence of external disambiguating lexical information. If accurate lexical predictions (whether internally derived, or externally provided) contribute to adaptive plasticity, we expected that the impact of external lexical information should be diminished when signal distortions were incrementally introduced, preserving more accurate internal lexical predictions across training. We observed this predicted interaction. To the extent that internal lexical predictions were more accurate, external disambiguating lexical information had less of a role in driving adaptive plasticity.

In Experiment 2, the presence of external information greatly boosted the degree of adaptive plasticity relative to that observed without external information. However, in Experiment 3 the presence of external information had no significant impact on the degree of adaptive plasticity. This difference was driven by the intelligibility of the stimuli experienced during the training period in Experiments 2 versus 3. Because all other aspects of the study were equivalent across experiments, other factors that could potentially contribute, including any potential adaptive plasticity that occurs over the course of the pretest or posttest and the specific type of external information provided, can be ruled out. Indeed, gradually introducing the signal distortion such that more accurate lexical predictions were preserved during training had an effect on adaptive plasticity comparable, and statistically indistinguishable from, the availability of an external source of lexical information. Moreover, significant positive correlations between word recognition performance during training and the improvements in word recognition between pretest to posttest show a consistent positive relationship between the accuracy of lexical predictions and adaptive plasticity, in each of the conditions, across *both* Experiments 2 and 3 (see Figure 4). Taken together, these results confirm the hypothesis that internal and external information sources contribute to adaptive changes in speech perception.

It is important to note that adaptive plasticity in this study, like that observed in some previous studies (Schwab et al., 1985), can be localized to a prelexical level. Each and every word encountered in the experiment was unique for any given participant. Therefore, improvements in word recognition accuracy of the severely distorted signals from pretest to posttest requires generalization to novel lexical items and, therefore, indicates adaptive plasticity in the mapping of the distorted acoustic input at a prelexical level.

Beyond the prelexical locus of remapping, the mechanisms responsible for adaptive plasticity are not well understood. However, Guediche et al. (2014; see also Vroomen and Baart, 2012) make the case that the factors significant in driving adaptive plasticity in speech perception bear some similarity to those involved in driving adaptation in other domains. Thus, the mechanisms involved in other forms of adaptation (e.g., sensorimotor adaptation) may be informative in guiding the development of a deeper mechanistic understanding of adaptive plasticity in speech perception. In sensorimotor adaptation tasks like prism goggle adaptation, for example, changes in motor output result after exposure to distorted sensory input (Redding & Wallace, 2006). In sensorimotor adaptation, internally generated sensory outcome predictions are generated through efference copies of the motor plan. These predictions, in turn, are used to derive sensory prediction error signals that denote the difference between the expected and actual sensory consequences of a planned movement mismatch, and can be used to guide adaptive changes in motor output. A great deal of research links this supervised, error-driven learning to a specific neural system involving the cerebellum (Kawato & Wolpert, 1998; Wolpert, Miall, & Kawato, 1998; Wolpert, Diedrichsen, & Flanagan, 2011).

Although the motor plan is the source of prediction in sensorimotor adaptation, there are intriguing parallels with adaptive plasticity in speech perception if one posits that *lexical* knowledge can serve as a basis for generating predictions of expected sensory input (Guediche et al., 2014). Instead of sensorimotor adaptation, adaptive plasticity in speech perception may be a case of *sensory-cognitive* adaptation or, more specifically in the present case, *auditory prelexical-lexical* adaptation. In previous work, we explored this possibility by examining whether neural systems that have been implicated in supervised learning in sensorimotor adaptation are involved in adaptive plasticity in speech perception. Using functional MRI (fMRI) and a behavioral paradigm with intermediate distortion levels and no disambiguating external lexical information, we found evidence for cerebellar involvement in adaptive plasticity in speech perception (Guediche, Holt, Laurent, Lim, & Fiez, 2015) consistent with the possibility that error-driven learning may play a role.

The possibility of common neurobiological supervised learning mechanisms supporting sensorimotor adaptation and adaptive plasticity in speech perception is largely by analogy, with empirical support thus far limited to the findings reported by Guediche et al. (2015). Moreover, the perspective may be viewed as controversial because the supervised learning mechanism proposed for sensorimotor adaptation is dependent on generating predictions from motor planning and relies on the cerebellum, which has been historically regarded as a motor structure (Glickstein et al., 2006).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Accordingly, cerebellar-dependent computational principles of adaptation have been incorporated into models of speech production (e.g., (Golfinopoulos et al., 2011; Guenther & Ghosh, 2003) but not speech perception (but see Kleinschmidt & Jaeger, 2015).

At the same time, there are reasons to favor the idea of a common supervised learning mechanism that supports sensorimotor adaptation and adaptive plasticity in speech perception. For instance, even sensorimotor adaptation to altered sensory input is not limited to changes in motor output (Bedford, 1999). Recent research demonstrates that sensorimotor adaptation can produce changes in sensory representations across traditionally distinct perceptual systems (Volcic, Fantoni, Caudek, Assad, & Domini, 2013), including studies reporting changes in speech perception after sensorimotor adaptation in speech production (e.g., Lametti, Rochet-Capellan, Neufeld, Shiller, & Ostry, 2014; Mattar, Nasir, Darainy, & Ostry, 2011; Nasir & Ostry, 2009; Shiller, Sato, Gracco, & Baum, 2009). For example, in response to distorted auditory feedback of one's own voice during speech perception (Shiller et al., 2009), listeners' perception of a category boundary is shifted in a direction that reduces the impact of the sensorimotor speech production perturbation. Thus, even sensorimotor adaptation has perceptual consequences. Further, there is increasing evidence that supports cerebellar involvement in purely auditory tasks that do involve motor processing (e.g., Petacchi et al., 2005) and that it may encode prediction error signals in speech perception (Rothermich & Kotz, 2013). Together, these findings indicate that cerebellar-mediated supervised error-driven learning is a plausible mechanism for adaptive plasticity in speech perception.

The present research demonstrates that gradual increments in the distortion result in adaptive plasticity comparable to that driven by external information, in line with the idea that the probability of successful predictions can transfer to higher levels through progressive alignment. Intelligibility affects the degree to which the speech signal makes contact with internal linguistic knowledge, thereby serving as a proxy for the accuracy of the prediction and the effectiveness of the resulting error signal in driving learning. This is consistent with models of supervised learning, in which the magnitude of learning depends on both the quality of the input (reliability of the error) in addition to the size of the error (Kleinschmidt & Jaeger, 2015). Predictions about the speech signal are driven by the interaction between these two factors to generate the error signal that drives learning.

Although at least one form of adaptive plasticity in speech perception can be modeled with Hebbian learning (Mirman, McClelland, & Holt, 2006), the relatively slow time-course of Hebbian learning is not a perfect fit with the rapid tuning of speech perception observed in studies of adaptive plasticity (see Guediche et al., 2014 for discussion). Supervised learning driven by prediction errors may better align with the time course of rapid adaptive plasticity. Although this point has been made with regard to adaptive plasticity related to perceptual tuning of specific phonetic category boundaries by disambiguating lexical or visual information (e.g., Bertelson, Vroomen, & de Gelder, 2003; Norris et al., 2003), it has not been considered as a potential mechanism driving more global adaptive plasticity like that observed in the current study or in studies of fluent speech whereby distortion affects all speech sounds. Our finding that internal predictions can drive adaptive plasticity, even in the absence of feedback, are consistent with supervised, error-driven learning for global speech distortions. This presents the possibility that a common mechanism may contribute to a variety of adaptive plasticity phenomena in speech perception (see Guediche et al., 2014; Kleinschmidt & Jaeger, 2015; but see Baart & Samuel, 2015).

In summary, this study supports the possibility that externally- and internally mediated adaptive plasticity may arise via common processes. Until now, the relationship between the accuracy of internally generated predictions and the degree of adaptive plasticity in speech perception has not been directly investigated. This is in part due to the fact that speech perception research on predictions and prediction error signals has yet to be fully integrated with studies on adaptive plasticity (see Guediche et al., 2014, for a review). The present findings extend prior perspectives on supervised learning in speech perception. Specifically, they provide the first demonstration that the accuracy of internally generated linguistic predictions is correlated with the degree of adaptive plasticity and that information that affects predictions, whether externally presented or internally generated, appears to contribute in a redundant fashion, to adaptive plasticity in speech perception. In other words, either source of information is sufficient for learning but the benefit of their contribution is not additive. Furthermore, these results suggest that adaptive plasticity can be induced through different sources of disambiguating lexical information that is needed for lexically mediated supervised learning. This may have important implications for rehabilitation of listeners who experience difficulty adapting to acoustic distortions to speech.

References

Adank, P., & Janse, E. (2009). Perceptual learning of time-compressed and natural fast speech. *The Journal of the Acoustical Society of America, 126,* 2649–2659. http://dx.doi.org/10.1121/1.3216914

Altmann, G. T. M., & Young, D. (1993, September). *Factors affecting adaptation to time-compressed speech*. Paper presented at the EUROSPEECH conference, Berlin, Germany.

Baart, M., & Samuel, A. G. (2015). Turning a blind eye to the lexicon: ERPs show no cross-talk between lip-read and lexical context during speech sound processing. *Journal of Memory and Language, 85,* 42–59. http://dx.doi.org/10.1016/j.jml.2015.06.008

Bar, M. (2003). A cortical mechanism for triggering top–down facilitation in visual object recognition. *Journal of Cognitive Neuroscience, 15,* 600–609.

Bedford, F. L. (1999). Keeping perception accurate. *Trends in Cognitive Sciences, 3,* 4–11. http://dx.doi.org/10.1016/S1364-6613(98)01266-2

Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science, 14,* 592–597. http://dx.doi.org/10.1046/j.0956-7976.2003.psci_1470.x

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106,* 707–729. http://dx.doi.org/10.1016/j.cognition.2007.04.005

Brouwer, S., & Bradlow, A. R. (2014). Contextual variability during speech-in-speech recognition. *The Journal of the Acoustical Society of America, 136,* 26–32. http://dx.doi.org/10.1121/1.4881322

Church, B. A., Mercado, E., III, Wisniewski, M. G., & Liu, E. H. (2013). Temporal dynamics in auditory perceptual learning: Impact of sequencing and incidental learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 270–276. http://dx.doi.org/10.1037/a0028647

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America, 116,* 3647–3658. http://dx.doi.org/10.1121/1.1815131

Cooper, A., Brouwer, S., & Bradlow, A. R. (2015). Interdependent processing and encoding of speech and concurrent background noise. *Attention, Perception, & Psychophysics, 77,* 1342–1357. http://dx.doi.org/10.3758/s13414-015-0855-z

Dahan, D., & Mead, R. L. (2010). Context-conditioned generalization in adaptation to distorted speech. *Journal of experimental psychology: Human perception and performance, 36,* 704.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., 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. http://dx.doi.org/10.1037/0096-3445.134.2.222

Diehl, R. L., Elman, J. L., & McCusker, S. B. (1978). Contrast effects on stop consonant identification. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 599–609. http://dx.doi.org/10.1037/0096-1523.4.4.599

Egan, J. P. (1948). Articulation testing methods. *Laryngoscope, 58,* 955–991.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology, 4,* 99–109. http://dx.doi.org/10.1016/0010-0285(73)90006-6

Fant, G. (1949). *Analys av de Svenska Konsonantljuden: Talets Allmänna Svängningsstruktur* [Analysis of Swedish Consonant]. Stockhom, Sweden: LM Ericsson.

Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003, October 9). Consolidation during sleep of perceptual learning of spoken language. *Nature, 425,* 614–616. http://dx.doi.org/10.1038/nature01951

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics, 62,* 1668–1680. http://dx.doi.org/10.3758/BF03212164

Francis, A. L., Nusbaum, H. C., & Fenn, K. (2007). Effects of training on the acoustic phonetic representation of synthetic speech. *Journal of Speech, Language, and Hearing Research, 50,* 1445–1465. http://dx.doi.org/10.1044/1092-4388(2007/100)

Fu, Q. J., & Galvin, J. J., III. (2003). The effects of short-term training for spectrally mismatched noise-band speech. *The Journal of the Acoustical Society of America, 113,* 1065–1072. http://dx.doi.org/10.1121/1.1537708

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155–170. http://dx.doi.org/10.1207/s15516709cog0702_3

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52,* 45–56. http://dx.doi.org/10.1037/0003-066X.52.1.45

Glickstein, M. (2006). Thinking about the cerebellum. *Brain, 129,* 288–290.

Golfinopoulos, E., Tourville, J. A., Bohland, J. W., Ghosh, S. S., Nieto-Castanon, A., & Guenther, F. H. (2011). fMRI investigation of unexpected somatosensory feedback perturbation during speech. *NeuroImage, 55,* 1324–1338. http://dx.doi.org/10.1016/j.neuroimage.2010.12.065

Greenspan, S. L., Nusbaum, H. C., & Pisoni, D. B. (1988). Perceptual learning of synthetic speech produced by rule. *Journal of Experimental Psycholology: Learning, Memory, and Cognition, 14,* 421–433. http://dx.doi.org/10.1037/0278-7393.14.3.421

Greenwood, D. D. (1990). A cochlear frequency–position function for several species—29 years later. *The Journal of the Acoustical Society of America, 87,* 2592–2605. http://dx.doi.org/10.1121/1.399052

Guediche, S., Blumstein, S. E., Fiez, J. A., & Holt, L. L. (2014). Speech perception under adverse conditions: Insights from behavioral, computational, and neuroscience research. *Frontiers in Systems Neuroscience, 7,* 126. http://dx.doi.org/10.3389/fnsys.2013.00126

Guediche, S., Holt, L. L., Laurent, P., Lim, S. J., & Fiez, J. A. (2015). Evidence for cerebellar contributions to adaptive plasticity in speech perception. *Cerebral Cortex, 25,* 1867–1877. http://dx.doi.org/10.1093/cercor/bht428

Guenther, F. H., & Ghosh, S. S. (2003). A model of cortical and cerebellar function in speech. In *Proceedings of the XVth International Congress of Phonetic Sciences* (pp. 169–173). Barcelona, Spain: International Congress of Phonetic Sciences.

Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., & Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: Effects of feedback and lexicality. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 460–474. http://dx.doi.org/10.1037/0096-1523.34.2.460

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America, 61,* 1337–1351. http://dx.doi.org/10.1121/1.381436

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology, 9,* 718–727. http://dx.doi.org/10.1016/S0959-4388(99)00028-8

Kawato, M., & Wolpert, D. (1998). Internal models for motor control. *Novartis Foundation Symposium, 218,* 291–304.

Kleinschmidt, D., & Jaeger, T. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review, 122,* 148–203.

Lametti, D. R., Rochet-Capellan, A., Neufeld, E., Shiller, D. M., & Ostry, D. J. (2014). Plasticity in the human speech motor system drives changes in speech perception. *The Journal of Neuroscience, 34,* 10339–10346. http://dx.doi.org/10.1523/JNEUROSCI.0108-14.2014

Li, T., & Fu, Q. J. (2007). Perceptual adaptation to spectrally shifted vowels: Training with nonlexical labels. *Journal of the Association for Research in Otolaryngology, 8,* 32–41. http://dx.doi.org/10.1007/s10162-006-0059-2

Liss, J. M., Spitzer, S. M., Caviness, J. N., & Adler, C. (2002). The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *The Journal of the Acoustical Society of America, 112,* 3022–3030. http://dx.doi.org/10.1121/1.1515793

Loebach, J. L., Pisoni, D. B., & Svirsky, M. A. (2010). Effects of semantic context and feedback on perceptual learning of speech processed through an acoustic simulation of a cochlear implant. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 224–234. http://dx.doi.org/10.1037/a0017609

Mattar, A. A., Nasir, S. M., Darainy, M., & Ostry, D. J. (2011). Sensory change following motor learning. *Progress in Brain Research, 191,* 31–44. http://dx.doi.org/10.1016/B978-0-444-53752-2.00015-1

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science, 32,* 543–562. http://dx.doi.org/10.1080/03640210802035357

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. http://dx.doi.org/10.3758/BF03213909

Murray, M. M., Wallace, M. T., Vroomen, J., & Baart, M. (2012). Phonetic recalibration in audiovisual speech.

Nasir, S. M., & Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proceedings of the National Academy of Sciences of the United States of America, 106,* 20470–20475. http://dx.doi.org/10.1073/pnas.0907032106

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47,* 204–238. http://dx.doi.org/10.1016/S0010-0285(03)00006-9

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences, 11,* 520–527. http://dx.doi.org/10.1016/j.tics.2007.09.009

Pallier, C., Sebastian-Gallés, N., Dupoux, E., Christophe, A., & Mehler, J. (1998). Perceptual adjustment to time-compressed speech: A cross-linguistic study. *Memory & Cognition, 26,* 844–851. http://dx.doi.org/10.3758/BF03211403

Panichello, M. F., Cheung, O. S., & Bar, M. (2013). Predictive feedback and conscious visual experience. *Frontiers in Psychology, 3,* 620.

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. http://dx.doi.org/10.1037/0096-1523.31.6.1315

Petacchi, A., Laird, A., Fox, P., & Bower, J. (2005). Cerebellum and auditory function: An ALE meta-analysis of functional neuroimaging studies. *Human Brain Mapping, 25,* 118–128.

Redding, G. M., & Wallace, B. (2006). Generalization of prism adaptation. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1006–1022. http://dx.doi.org/10.1037/0096-1523.32.4.1006

Rothermich, K., & Kotz, S. A. (2013). Predictions in speech comprehension: FMRI evidence on the meter-semantic interface. *NeuroImage, 70,* 89–100. http://dx.doi.org/10.1016/j.neuroimage.2012.12.013

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.

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. http://dx.doi.org/10.1126/science.270.5234.303

Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America, 125,* 1103–1113. http://dx.doi.org/10.1121/1.3058638

Sumner, M. (2011). The role of variation in the perception of accented speech. *Cognition, 119,* 131–136. http://dx.doi.org/10.1016/j.cognition.2010.10.018

Volcic, R., Fantoni, C., Caudek, C., Assad, J. A., & Domini, F. (2013). Visuomotor adaptation changes stereoscopic depth perception and tactile discrimination. *The Journal of Neuroscience, 33,* 17081–17088. http://dx.doi.org/10.1523/JNEUROSCI.2936-13.2013

Vroomen, J., & Baart, M. (2012). Phonetic recalibration in audiovisual speech. In M. M. Murray & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes* (pp. 254–259). Boca Raton, FL: CRC Press.

Vroomen, J., van Linden, S., de Gelder, B., & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory-visual speech perception: Contrasting build-up courses. *Neuropsychologia, 45,* 572–577. http://dx.doi.org/10.1016/j.neuropsychologia.2006.01.031

Wolpert, D. M., Diedrichsen, J., & Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience, 12,* 739–751.

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences, 2,* 338–347. http://dx.doi.org/10.1016/S1364-6613(98)01221-2

Zeng, F. G. (2004). Trends in cochlear implants. *Trends in Amplification, 8,* 1–34. http://dx.doi.org/10.1177/108471380400800102

Received June 9, 2015
Revision received December 7, 2015
Accepted December 10, 2015 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.