Sensorimotor Adaptation of Speech I: Compensation and Adaptation

John F. Houde
Department of Otolaryngology
University of California
San Francisco, CA 94143-0732

Michael I. Jordan
Division of Computer Science and Department of Statistics
University of California
Berkeley, CA 94720-1776

Running Title: Speech Adaptation

Abstract

When motor actions (e.g., reaching with your hand) adapt to altered sensory feedback (e.g., viewing a shifted image of your hand through a prism), the phenomenon is called sensorimotor adaptation (SA). In the study reported here, SA was observed in speech. In two two-hour experiments (adaptation and control), subjects whispered a variety of CVC words. For those words containing the vowel [ɛ], subjects heard auditory feedback of their whispering. A DSP-based vocoder processed the subjects' auditory feedback in real-time, allowing the formants of subjects' auditory speech feedback to be shifted. In the adaptation experiment, formants were shifted along one edge of the vowel triangle. For half the subjects, formants were shifted so subjects heard [ɑ] when they produced [ɛ]; for the other half, the shift made subjects hear [i] when they produced [ɛ]. During the adaptation experiment, subjects altered their production of [ɛ] to compensate for the altered feedback, and these production changes were retained when subjects whispered with auditory feedback blocked by masking noise. In a control experiment, in which the formants were not shifted, subjects' production changes were small and inconsistent. Subjects exhibited a range of adaptations in response to the altered feedback, with some subjects adapting almost completely, and other subjects showing very little or no adaptation.

Introduction

The role of auditory feedback in speech production has been a topic of research since the early studies of Lombard, who showed that speech was possible with masking noise that blocked a speaker's ability to hear (Lombard, 1911; Lane & Tranel, 1971). This result is consistent with later investigations of the speech of adults deafened after learning to speak, which showed that even after decades of being deaf, these speakers were still able to produce intelligible speech (Cowie & Douglas-Cowie, 1983; Lane & Webster, 1991). Significantly, however, children that are prelingually deaf exhibit only a brief period of babbling and do not naturally learn to speak (Ross & Giolas, 1978; Oller & Eilers, 1988; Borden, Harris, & Raphael, 1994). Even with intensive speech therapy, it is difficult for such children to develop intelligible speech (Smith, 1975; Levitt, Stromberg, Smith, & Gold, 1980; Osberger & McGarr, 1982). Such evidence would suggest that auditory feedback is critical for learning to speak but is thereafter not needed. But does this mean that auditory feedback is ignored after speaking is learned? Several studies suggest otherwise. In the early 1950's, Lee showed that listening to a delayed version of their auditory feedback severely disrupted most speakers' speech (Lee, 1950). The effect of delayed auditory feedback (DAF) on speaking became the subject of intensive research (Yates, 1963), and inspired Fairbanks to propose a model of speech motor control based on auditory feedback (Fairbanks, 1954).

Speech can be produced without immediate auditory feedback, but there is evidence that at least some aspects of speech may require auditory feedback to be correctly produced. Although deafened adults do continue to produce intelligible speech, it is also known that certain aspects of their speech begin to deteriorate soon after deafness. The production of sibilant fricatives (such as /s/) rapidly deteriorates, along with the control of pitch (Cowie & Douglas-Cowie, 1983; Matthies, Svirsky, Perkell, & Lane, 1996; Lane et al., 1997). These results are consistent with more recent experiments of Kawahara, who applied perturbations to the pitch of speakers' auditory feedback. Kawahara found that speakers will make compensatory adjustments to their pitch within about 100-150ms of a pitch feedback perturbation (Kawahara, 1993). Other manipulations of auditory feedback have also been shown to alter speech. For example, a spectral shift of a speaker's auditory feedback will cause the speaker to shift the spectrum of his produced speech (Gracco, Ross, Kalinowski, & Stuart, 1994).

To explain many of these results, Perkell and Lane proposed that the production of phonemes is modified by "postural parameters," such as pitch, which determine the manner in which the phonemes are spoken (Lane et al., 1997). Part of their theory postulates that the neural delays in processing auditory feedback probably make it unusable for the control of fast speech movements (Perkell, 1996). Postural parameters, which are slowly changing speech features, could be controlled directly by auditory feedback, and is why their control deteriorates rapidly in deafened speakers. On the other hand, phonetic parameters, which determine what phoneme is being spoken, are rapidly changing speech features that must be relatively insensitive to immediate auditory feedback.

If the Perkell/Lane theory is correct, how are phonetic parameters learned and maintained? Similar questions are asked about the control of reaching movements, because a reach to a target point, like the production of a phoneme, can be made correctly without immediate feedback (in this case, visual) to guide the reach (Polit & Bizzi, 1979; Sainburg, Ghilardi, Poizner, & Ghez, 1995). The role of visual feedback in reaching has, for more than a century, been studied using altered visual feedback (Welch, 1978), as exemplified by the prism adaptation experiments of Helmholtz (Helmholtz, 1962). In these experiments, a subject reaches to targets while wearing image-shifting prism glasses. Initially, the subject misses the targets, but soon learns to compensate and reach accurately. This compensation is retained beyond the time that the glasses are worn: when the glasses are removed, the subject's reaches now overshoot targets in the direction that was compensated. This retained compensation is called adaptation, and its generation from exposure to altered sensory feedback is called *sensorimotor adaptation* (SA)

To explain SA in reaching, Held postulated that visual feedback, when present, was compared with predicted visual feedback, and that any consistent differences caused gradual changes in the parameters controlling reaches (Held & Hein, 1958; Hein & Held, 1962; Welch, 1986). In Held's account, predicted visual feedback is generated internally from "reafference," or "efference copy" of the motor efferent commands controlling the reach. Held's account has been elaborated by Jordan, who postulated that predicted sensory consequences of actions were provided by an internal model of motor-sensory relations (Jordan & Rumelhart, 1992). This "forward model" was first learned by simply observing the sensory consequences of random reaches, and later used to generate sensory predictions of motor acts.

These theories of SA suggest that visual feedback, while not needed for controlling any particular reach, is nonetheless critical for the maintenance and learning of reaches. Could the same be said about the role of auditory feedback in the production of phonemes? In our study, we addressed this question by investigating whether phoneme production could also be made to exhibit SA. We altered subjects' auditory feedback in a phonetically relevant fashion by shifting formant frequencies. We then examined whether extended exposure to this altered feedback caused subjects to compensate their vowel productions, and, most importantly, whether these compensations were retained after exposure. Some of the results that we report here previously appeared in (Houde & Jordan, 1998). The current paper provides additional findings as well as a more detailed exposition of the methodology and results presented there.

Methods

Subjects

The experiment was run on eight male native speakers of North American English who were either undergraduate or graduate students at MIT. All were naive to the purpose of the study. All subjects were first run in a pretest procedure in which their vowel formants were measured, and then, on a separate day, in the adaptation experiment in which the formants of their audio feedback were shifted as described below. About a month later, all subjects were run in a control experiment that was identical in every aspect to the adaptation experiment (e.g. same data acquisition and measurement procedures, same headphone configuration, acoustic environment) with the sole difference that formants were not shifted.

Apparatus

For all parts of the study – the pretest procedure, the adaptation experiment, and the control experiment – the same apparatus was used to prompt subjects to speak and to provide audio feedback of their speech. The key components of the apparatus are shown in Figure 1. The subject sits in front of a PC video monitor wearing a head-mounted microphone and earphones. Words are presented on the monitor screen for the subject to pronounce. To minimize bone conduction effects, the subject whispers rather than speaks his pronunciations of these words. The whispered speech is transduced by the microphone and fed as input to a digital signal processing board (called the DSP system) inside the PC. The DSP system implements a formant shifting acoustic transformation (described below) and returns the altered feedback to the subject via the insert earphones. It also records the formants of subject's utterances.

Figure 1 also shows an overview of the key signal processing steps running on the DSP which implements the acoustic transformation. It is an analysis-synthesis process which repeatedly: (a) Captures from the microphone an 8ms frame of the subject's whispered speech (64 time samples at an 8KHz sampling rate); (b) Performs a 64-channel spectral analysis of this frame, retaining only a smoothed magnitude spectrum of it; (c) Estimates the first four formants from the magnitude spectrum; (d) Alters the frequencies of the three lowest formants; (e) Resynthesizes a new 8ms frame of whispered speech from the altered formants. This process incurred a feedback delay of only 16ms.

The formant estimation method implemented in the DSP was determined by the special characteristics of whispered speech. In whispered speech, like breathy speech, the open glottis introduces pole-zero pairs in the spectrum that result from coupling to the subglottal cavities (Stevens, 1999). This results in a pole-zero pair in the neighborhood of F1 that distorts this formant into two spectral peaks. It can also result in the apparent shift of F1 frequency (Kallail & Emanuel, 1984). These distorting effects of subglottal coupling are seen to a greater or lesser degree for any given subject, which appears to relate to how open the subject keeps his glottis while whispering (Houde, 1997b). As a result, there was great variability accross subjects in the spectral distortion around F1, and so a simple peak finding algorithm for F1 was quite unreliable. Instead, we used a method robust to distortion effects in which F1 frequency was estimated as the center of spectral mass of a frequency interval containing F1. This frequency interval was determined in the subject pretest session. Higher formants were estimated as peaks in the spectrum of the whispered speech.

To check the consistency of our formant frequency estimates, formant frequencies, as estimated by the DSP were compared with peaks in the LPC spectrum computed by the ESPS/Waves+ speech analysis software on the same speech signal. By this method, we were able to verify that our formant estimates matched those of the ESPS/Waves+ software.

In the resynthesized speech output by the DSP, formant amplitudes were approximately maintained by scaling the synthesizer output to have the same RMS amplitude as the input speech. Formant bandwidths were also approximately preserved. Since only four formant peaks were used to resynthesize the output speech, there was a loss of detail in the valleys between formant peaks in the spectrum of the output speech. This was especially true in the region of F1, since we did not resynthesize the pole-zero pair found in this region in the input whispered speech (see below). However, all subjects, as well as the experimenters, rated the resynthesized output as perceptibly quite similar to the original speech.

Pretest Procedure

For a number of reasons, all subjects were run in a pretest session several days before being run in the experiments. First, a region over which F1 varied was needed for F1 estimation based on center-of-spectral-mass. Second, the formant-shifting transformations, described in more detail below, were subject-specific and based on the formant frequencies of a subject's normal whispered production of the vowels [i], [ɩ], [ɛ], [æ], and [ɑ]. Third, the pretest session insured that all subjects run later in the experiments had strong formants for most vowels, and that the transformationed versions of their vowels that would be used in the experiments sounded correct to the experimenter.

In the pretest session, subjects were prompted to whisper a series of CVC words containing the vowels mentioned above. Their audio feedback was provided by the apparatus described above, but their formants were not shifted. In the first part of the session, all formants were estimated by a simple peak finding algorithm, and the resulting data was used to estimate the range over which F1 varied accross all the vowel productions. Having determined this range, in the second part of the session, F1 was estimated using the center-of-spectral mass method, and the data collected was used to estimate the formants of the subject's vowel productions. The formants of each vowel were estimated from 60 productions of that vowel, which resulted in a mean standard error of about 7Hz for both the F1 and F2 formant frequency estimates.

Having measured the subject's vowel formants, the feedback transformations that would shift the subject's formants in the adaptation experiment could then be created. In the final part of the pretest session, the correctness of these transformations (i.e., their ability to make one vowel sound like another) was tested by the experimenter listening by himself to the output of the feedback shifting apparatus while the subject spoke into a microphone.

Feedback Transformations

To maximize the chances of observing compensation and adaptation in the experiments, the design of the audio feedback transformations was guided by two constraints. First, the transformation should introduce a change to the subject's feedback that is potentially perceivable by the subject. For example, changes to the speech signal above a subject's hearing range would presumably not be perceived by a subject, and no compensatory response would be expected. Second, the subject, upon detecting a feedback change, should be able to make some production change that compensates for the effect of the feedback change. Delayed auditory feedback is one example of a feedback change that is easily perceived by a subject, but for which there is no possible compensatory production adjustment.

In order to satisfy these two constraints, the audio feedback transformtions used were developed as an audio analog of the image-shifting action of prisms, as shown in Figure 2. The situation for reaches made while viewing the hand through prisms is shown in Figure 2(a). The image shift of the prisms is shown by the gray arrows. When a subject makes a quick reach to an imaginged target location, e.g., point P1 or P2 in the figure, he instead sees his hand arrive at a shifted location – P1' or P2'. Over continued exposure to this image shift, a subject responds by shifting his reaches to compensate. The black arrow in the figure shows a hypothetical example of this compensation. Intially, when the subject reaches for point P2, his hand actually arrives at point P2, but he sees his hand arrive at point P2'. However, after several reaches, when the subject reaches for point P2, his hand actually arrives at point P2c. As a result, the subject now sees his hand arriving at point P2c' – i.e., closer to his original intended target point P2.

To aid in constructing an audio analog of this situation, the image shift of the prisms can be described with respect to an numbered reference path oriented along the shift direction (dashed line in figure). Points can be described in terms of their deviation from this path and their distance along it. These quantities are shown for point P1 in the figure. The image shift changes the hand's apparent distance along the path but preserves the how much the subject's hand deviates from the path. In the example shown here, the image shift subtracts 2.0 units from the hand's apparent distance along the path: actual hand positions P1 and P2 are centered around position 3; their images (P1' and P2') are centered around position 1.

An audio analog of this situation is shown in abstract form in Figure 2(b). The gray arrows show the action of an audio feedback transformation that shifts vowel sounds along a perceptually salient direction: the edge of the vowel triange containing a subject's production of the vowels [i], [ɩ], [ɛ], [æ], and [ɑ]. These vowels are numbered to correspond to the numbered path described above for the hand image shift of prisms, and the resulting path is refered to as the subject's [i]–[ɑ] path. Vowel sounds can be described in terms of their deviation from this path and their distance along it. These quanatities are indicated for vowel sound V1. When a subject produces a vowel sound, e.g., vowel sound V1 or V2 near [ɛ] in the figure, the audio transformation subtracts 2.0 units from the vowel sound's distance along the [i]–[ɑ] path, but preserves its deviation from the path. As a result, the subject's audio feedback of his actual vowel production is instead a different vowel sound – vowel sounds V1' or V2' near [i] as shown in the figure. Such a change in vowel sound (i.e., from sounding similar to [ε] to instead sounding similar to [i]) is likely to be perceived by the subject. Furthermore, it is possible for the subject to compensate by shifting his vowel productions in a direction opposite the audio feedback shift. An example is shown by the black arrow. Intially, when the subject produces vowel sound V2, his audio feedback is of vowel sound V2'. However, by shifting his production to V2c, his audio feedback will instead be vowel sound V2c' – i.e., closer to his original, unaltered audio feedback of vowel sound V2.

Figure 2(c) shows how the real situation differs from the abstract audio analog of prism shift described above. The figure shows the positions, in terms of formant 1 and 2 frequencies, of subject OB's productions of the vowels [i], [ɩ], [ε], [æ], and [ɑ]. As in Figure 2(b), these vowels are numbered, and the dashed line joining the vowels is OB's [i]–[ɑ] path. This path is distorted from the idealized [i]–[ɑ] path of Figure 2(b): the path is not straight, and the distance between vowels on the path is variable. Nevertheless, audio transformations can still be defined as shifts with respect to this path. Vowel sounds can still be described in terms of their deviation from this path and their distance along it, as is shown for vowel sound V1. A vowel sound's distance along the path – which will be called *path projection* – is defined as the nearest point on the [i]–[ɑ] path to the vowel and this distance is normalized to be the same between all adjacent vowels on the path. In this figure, the gray arrows show the action of the -2.0 transformation – one of the two formant-shifting audio transformations used in the experiments. When a subject produces a vowel sound, e.g., vowel sound V1 or V2 near [ε] in the figure, the -2.0 transformation subtracts 2.0 units from the vowel sound's path projection, but preserves its path deviation. As a result, the subject's audio feedback of his actual vowel production is instead a different vowel sound – vowel sounds V1' or V2' near [i] as shown in the figure. By the same argument used in the previous figure, such a vowel change is likely to be perceived by the subject, and the subject can still compensate by shifting his vowel productions along the [i]–[ɑ] path in a direction opposite the audio feedback shift. A hypothetical example of compensation for the -2.0 transformation is shown by the black arrow. Intially, when the subject produces vowel sound V2, his audio feedback is of vowel sound V2'. However, by shifting the formants of his production to V2c, his audio feedback will instead be vowel sound V2c' – i.e., closer to the formants of his original, unaltered audio feedback of vowel sound V2.

The other feedback transformation used in the experiments, but not shown in Figure 2(c), was the +2.0 transformation, which added 2.0 units to a vowel sound's path projection. In this case, for example, when a subject produces a vowel sound near [ɛ] the audio feedback will be a vowel sound near [ɑ]. As with the -2.0 transformation, a subject could compensate for the +2.0 transformation by shifting production along the [i]–[ɑ] path in the opposite direction, e.g., by shifting the formants of his production of [ɛ] towards [i].

Experimental Design

Two experiments were performed with each subject: an adaptation experiment with either the -2.0 or +2.0 transformation, and, about a month later, in a control experiment. This control experiment was identical to the adaptation experiment except that a strength 0.0 feedback transformation (no feedback alteration) was used throughout the experiment. In the adaptation experiment, half the subjects were run with the +2.0 transformation (subjects RS, CW, TY, and VS), and half were run with the -2.0 transformation (subjects AH, OB, RO, and SR).

Figure 3 shows the design used in all of the experiments. The PC video screen prompted the subject to whisper the printed word with a 300ms target duration (as indicated by the length of the bracket beneath the word). Dots appearing as the subject whispered indicated when he had reached this target duration. The word promptings came in groups of 10 called *epochs*. The first 6 prompted words came from a training word set ($\mathbf{W_{train}}$) and the last 4 came from a test word set ($\mathbf{W_{test}}$). Via the earphones, the subject heard feedback of his first 5 utterances but heard only masking noise after that. Thus the subject never heard his production of the test words, which were used to test how training word adaptation generalized. Data from $\mathbf{W_{train}}$ were used to assess compensation and adaptation, and it is the analysis of these data that is the focus of the current paper. $\mathbf{W_{train}}$ was a set of four CVC [ε] words: "pep," "peb," "bep," and "beb." These words were chosen for several reasons. First, both the beginning and ending consonants are bilabials (produced with the lips), which implied less interfering coarticulation with the vowel (produced with the tongue body). This tended to result in longer, cleaner steady-state vowel portions of the utterances. Data from $\mathbf{W_{test}}$ were used to assess generalization of any adaptation seen in the training words to other vowels and other consonant contexts. $\mathbf{W_{test}}$ consisted of "pep," "peg," "gep," "teg," "peep," "pip," "pap,", and "pop." (Analysis of generalization will be covered in a followup paper.)

As Figure 3 shows, the experiment consisted of 422 epochs, organized into a sequence of five *phases*:

The warmup phase consisted of 36 epochs, which, for the average subject, had a duration of 10 minutes. In this phase, when the subject heard feedback of his whispering, the feedback was unaltered (i.e., the 0.0 feedback transformation was used). The purpose of the warmup phase was to provide time for the subject to acclimate to the experimental conditions. It was expected that 10 minutes would be sufficient time for this acclimation to occur and the subject's whisperings to stabilize.

The baseline phase consisted of 60 epochs, which, for the average subject, had a duration of 17 minutes. In this phase, when the subject heard feedback of his whispering, the feedback was unaltered (i.e., the 0.0 feedback transformation was used). The purpose of the baseline phase was to collect baseline data of the subject's whisperings before his feedback was altered. Utterance data collected in this phase were compared with data collected in later phases to assess whether the subject changed his whispering in response to the altered feedback.

The ramp phase consisted of 66 epochs, which, for the average subject, had a duration of 20 minutes. This phase was further subdivided into 11 *stages*, each of which was 6 epochs long (approximately 2 minutes in duration). Within each stage, the amount of feedback alteration was held constant. In the first stage (Stage 0), when the subject heard feedback of his whispering, it was unaltered (i.e., the 0.0 feedback transformation was used). Over the next 10 stages, however, the amount of feedback alteration was incremented linearly to its maximum magnitude. The maximum feedback alteration was either -2.0 or +2.0 vowel units, depending on the subject (see discussion of subjects below). The purpose of the ramp phase was to gradually introduce the feedback alteration to which the subject would be exposed for the rest of the experiment. The main reason for gradual introduction was to minimize the subject's awareness of the altered feedback.

The train phase consisted of 200 epochs, which, for the average subject, had a duration of 1 hour. In this phase, when the subject heard feedback of his whispering, the feedback was altered by either the -2.0 or +2.0 feedback transformation (depending on the subject). The purpose of the train phase was to give the subject roughly one hour of exposure to the full-strength feedback transformation.

The test phase consisted of 60 epochs, which, for the average subject, had a duration of 17 minutes. In this phase, when the subject heard feedback of his whispering, the feedback was altered by either the -2.0 or +2.0 feedback transformation (depending on the subject). Except for the altered feedback, the test phase was essentially a repeat of the baseline phase. The purpose of the test phase was to collect utterance data that could be compared with utterance data from the baseline phase. Formant changes seen in this comparison were used to assess whether the subject changed his whispering in response to the altered feedback.

At the end of the experiment, the subject was interviewed briefly. In this interview, subjects were asked a variety of questions about their experience in the experiment. These questions were: (1) Did the feedback sound correct? (2) Were there any problems with how it sounded? (3) Was the feedback too loud? The questions were posed to determine if subjects noticed anything unusual about the feedback or any change in their vowel productions.

Results

The results for an individual subject (OB) are shown in Figure 4. The points labeled [i] through [ɑ] represent the mean (F1,F2) values of OB's productions of these vowels in the pretest session; the dashed line connecting these points is OB's [i]–[ɑ] path; transformations of OB's audio feedback were based on this path. The vowels along this path are also labeled with their path projection values – their normalized distance along the [i]–[ɑ] path.

The gray arrow running along OB's [i]–[ɑ] path shows the action of the feedback transformation used in the adaptation experiment. The base of the arrow labeled "Baseline" represents the mean (F1,F2) values of OB's production of the vowel /eh/ in the baseline phase of the experiment.[1] Relative to OB's [i]–[ɑ] path, his baseline production of /eh/ has a path deviation of about 70 Hz and a path projection of about 3.2. The -2 feedback transformation was designed to shift vowel formants along OB's [i]–[ɑ] path, and, as the curved gray arrow shows, it does this by preserving path deviation values and subtracting 2 from path projection values. The arrow tip represents the effect of this transformation on OB's audio feedback of his baseline production of [ɛ] – it preserves the path deviation of 70 Hz, but

[1]The ellipse showing the F1 and F2 standard errors around this mean is too small to be seen in the figure.

reduces the path deviation to 1.2, resulting in a vowel sound close to [i].

The black arrow shows OB's compensation response to the feedback transformation: it shows OB altered his production of [ɛ] to oppose the action of the feedback transformation. The tip of the arrow labeled "Test" represents the mean (F1,F2) values of OB's production of [ɛ] by the end of the experiment, in the test phase. Relative to OB's [i]–[ɑ] path, his production of /eh/ in the test phase has a path deviation of about 100 Hz and a path projection of about 4.5. Thus, in response to the altered feedback, OB changed path deviation by about 30 Hz – which, as expected, is small since the feedback transformation did not shift path deviation. But OB increased path projection by 1.3 – i.e. directly opposing the distortion of the feedback transform that subtracted 2 from his path projections. Thus, if we measure mean compensation as:

$$\text{mean compensation} = \frac{(\text{mean path projection change})}{-\begin{pmatrix}\text{path projection shift of the}\\ \text{feedback transformation}\end{pmatrix}}$$

For subject OB, this value is: $(1.3)/(-(-2.0)) = 0.65$.

The white arrow shows OB's adaptation response to the feedback transformation: it shows how his production of [ɛ] without audio feedback (i.e., blocked by masking noise) changed from the baseline to the test phase of the experiment. The arrow base represents the mean (F1,F2) values of OB's production of [ɛ] in the baseline phase of the experiment. The base of the arrow shows that when OB was whispering without being able to hear himself, his vowel formants in the baseline phase were about halfway between [ɛ] and [æ] and about 40Hz distant from his [i]–[ɑ] path, which gives a path projection of about 3.5 and a path deviation of 40Hz. The arrow tip shows that OB's mean vowel formants under these same whispering conditions in the Test phase were a bit past [æ] and about 60Hz from his [i]–[ɑ] path, for a path projection of about 4.3 and a path deviation of 60Hz. Thus, under these whispering conditions, OB's vowel productions exhibited a mean path projection change of about 0.8 and a mean path deviation change of about 20Hz. We can again use the mean compensation fraction calculation described above to compare the mean path projection change observed under these whispering conditions with the magnitude of the feedback transformation. In this case, the value is: $(0.7)/(-(-2.0)) = 0.4$. We call this value OB's *mean adaptation* because it reflects how much of the subject's mean compensation is retained when he whispers in the absence of feedback.

On the other hand, OB showed virtually no production changes in the control experiment. The gray arrow near the vowel æ shows how the mean formants of [ɛ] produced by OB when hearing audio feedback changed from the baseline phase to the test phase. In fact, only the arrowhead of this arrow can be seen, as there was very little production change from baseline to test phase. OB's change in production of [ɛ] with feedback blocked by masking noise was similarly small, and the arrow representing this change is omitted for clarity.

Mean Compensation, Adaptation, and Path Deviation Data

Measures of path projection and path deviation were used to analyze the data for all subjects. Consider first the values derived from these measures – mean compensation, adaptation, and path deviation change – which most directly show how subjects responded to the altered feedback. Plots of these values are shown in Figure 5. In the figure, the left column shows subjects' compensation responses during the trials when they spoke while hearing feedback. Figure 5(a) shows calculated mean compensation for these responses, while Figure 5(b) shows the corresponding mean path deviations. The figures show that during the adaptation experiment (filled dots connected by the solid line) subjects consistently exhibited positive mean compensations, while in the control experiment (empty dots connected by dotted line) subjects' mean compensation was smaller and not consistently positive. Figure 5(b) shows that mean path deviation changes were inconsistent across subjects for both the adaptation and control experiments. The right column of Figure 5 shows that these same results are seen in adaptation response—the production changes that were retained when feedback was blocked by noise. Figure 5(c) shows that mean adaptation (mean compensation calculated from a subject's adaptation responses) is again larger and more consistently positive across subjects in the adaptation experiment than in the control experiment. Figure 5(d) shows that, again, the mean path deviations of these responses are not consistent across subjects in either the adaptation or the control experiments.

In all the plots of Figure 5, subjects are shown in order of decreasing mean adaptation with their data points joined by lines (the lines do not reflect any other correlations among the subjects). From this method of presentation, it can be seen that the subjects exhibit a great range of adaptations in response to the altered feedback, with some subjects adapting almost completely (CW, RS), and other subjects showing very little or no adaptation (VS,AH). It is also evident that each subject's mean adaptation roughly predicts his mean compensation, which is generally larger than his mean adaptation. On the other hand, a subject's mean adaptation is clearly not a good predictor of his responses in the control experiment or his mean path deviation in either the adaptation or the control experiments.

ANOVA tests of the effect of experiment phase (baseline vs. test) on path projection and path deviation data confirm the significance of the key trends seen in Figure 5.

Both compensation and adaptation responses were significant across subjects ($[F(1,7) = 22.325, p = 0.002]$ for compensation, $[F(1,7) = 11.590, p = 0.011]$ for adaptation) and significantly greater in the adaptation experiments than in the control experiments ($[F(1,7) = 15.362, p = 0.006]$ for compensation, $[F(1,7) = 8.369, p = 0.023]$ for adaptation). However, in no case was mean path deviation change significant across subjects (best case $[F(1,7) = 1.858, p = 0.215]$).

In sum, the plots of Figure 5 show that the only aspect of subjects' responses consistently affected by exposure to altered feedback is mean compensation and adaptation in the adaptation experiment. Because mean compensation is measured relative to the direction of the feedback transformation, the measure is positive for compensating production changes, independent of the transformation's shift direction. The plots show that this measure is generally positive across all subjects. This indicates subjects generally altered their production so as to oppose the shift of the feedback transformation, regardless of the direction of that shift. It is therefore unlikely that the observed compensatory behavior is the result of some drift in subjects' vowel productions not related to altered feedback exposure.

Path Projection Analysis

More aspects of subjects' responses can be seen by examining separately the baseline and testing phase path projection values. The results are shown in figures 6 and 7.

Figure 6 shows subjects' mean path projections for the baseline and testing phases of both the adaptation experiment (solid lines) and control experiment (dotted lines). Recall that path projection is measured in inter-vowel intervals along the subject's [i]–[ɑ] path. Normally, on this scale, 1.0 corresponds to [i], 2.0 to [ι], 3.0 to [ε], 4.0 to [æ], and 5.0 to [ɑ] (see figures 2(c) and 4). The left plots show mean path projection for subjects' compensation responses: the top plot shows testing phase values; the bottom plot shows baseline phase values. The right plots show the analogous values for subjects' adaptation responses.

In each plot, comparison of subjects' path projections is facilitated by a normalization that makes increases in path projection indicate compensation for all subjects. This normalization is needed because, depending on the feedback transform he is exposed to, a subject compensates by either increasing or decreasing vowel path projections:

- Subjects AH, OB, RO, and SR were exposed to the -2.0 feedback transformation in the adaptation experiment and are referred to as the *-2.0 subjects*. The -2.0 transformation decreases perceived path projection by 2 vowel units, and the -2.0 subjects compensated for it by increasing their vowel path projections.
- Subjects RS, CW, TY, and VS were exposed to the +2.0 feedback transformation in the adaptation experiment and are referred to as the *+2.0 subjects*. The +2.0 transformation increases perceived path projection by 2 vowel units, and the +2.0 subjects compensated for it by decreasing their vowel path projections.

To make the -2.0 and +2.0 subjects' responses comparable, the +2.0 subjects' path projections were calculated using an [i]–[ɑ] path with reversed numbering (i.e., with 5.0 corresponding to [i], 4.0 to [ɩ], 3.0 to [ɛ], 2.0 to [æ], and 1.0 to [ɑ]). Reversing the numbering converts the +2.0 subjects' compensation responses from path projection decreases to path projection increases. This makes all subjects' results comparable in the plots: for all subjects, an increase in path projection indicates compensation.

These plots display several striking features. First consider Figure 6(a), which shows baseline mean path projections of subjects' compensation responses. For all subjects but TY, VS, and AH (the poorest adaptors), mean path projections are higher in the control experiment than in the adaptation experiment. Of the remaining subjects, all but SR show this same difference in their baseline adaptation responses (Figure 6(d)). ANOVA tests show that this baseline difference is significant across all subjects for both responses ($p = 0.014$ for compensation responses; $p = 0.043$ for adaptation responses).

The reverse of this situation is true for the testing phase: in this case, path projection seen in the adaptation experiment is generally higher than that seen in the control experiment. The observed difference is significant for compensation responses ($p = 0.020$) and marginally insignificant for adaptation responses ($p = 0.080$). However, if subjects VS and AH (the poorest adaptors) are excluded, the difference is also significant for adaptation responses.

These differences are summarized in Figure 7, which shows the path projections of Figure 6 averaged across subjects. As with Figure 6, increasing path projection values represent path projection changes in the compensating direction.

The figure shows subjects' compensation and adaptation in the adaptation experiment. The solid line in Figure 7(a) shows subjects' compensation responses in the adaptation experiment. These responses have an average baseline path projection close to 3.0 (the path position of [ε]) However, by the test phase, this average has shifted about 1.0 vowel unit in the compensating direction. This shift is highly significant ($F(1,7) = 22.325, p = 0.002$). The solid line in Figure 7(b) shows subjects' adaptation responses in the adaptation experiment. These responses have an average baseline path projection that is slightly higher than that of the baseline compensation responses. By the test phase, this average has shifted in the compensating direction, though not by as much as the shift seen in subjects' compensation responses. This shift is also significant ($F(1,7) = 11.590, p = 0.011$).

Figure 7 also shows that, although no appreciable production changes occur in the control experiment, baseline responses are shifted in the compensating direction. The dotted line in Figure 7(a) shows subjects' compensation responses in the control experiment. These responses have an average baseline path projection close to 3.4. Relative to the same responses in the adaptation experiment, this represents a significant shift in the compensating direction ($p = 0.014$). The figure also shows an insignificant shift in average path projection from the baseline to the test phase ($F(1,7) = 0.004, p = 0.951$). The dotted line in Figure 7(b) shows subjects' adaptation responses in the control experiment. These responses have an average baseline path projection that is slightly lower than that of the baseline compensation responses. By the test phase, however, average path projection has increased to equal the compensation response value. This increase was marginally significant ($F(1,7) = 5.819, p = 0.047$).

Discussion

From the mean compensation, adaptation, and path deviation analyses, we conclude that subjects changed vowel productions specifically to compensate for alterations of their feedback. We also conclude that these production changes are strong enough to be partly retained when feedback is blocked by noise. Thus, speech, like reaching, appears to exhibit sensorimotor adaptation, and may be amenable to the same models of motor learning that have been developed to explain reaching behavior. In addition to this main conclusion, our experiments also revealed several other characteristics of this speech adaptation effect.

For all subjects, compensation was greater than (or, occasionally, equal to) adaptation. One possible explanation for this difference is that some portion of each subject's compensation was accomplished by some temporary correction mechanism, active only in the presence of the altered feedback. In other words, vowel production could be partly under immediate auditory feedback control. This would seem to be a departure from the dichotomy, suggested by Perkell and Lane, of postural parameters, like pitch, being under feedback control and phonetic parameters, which determine phonetic identity, being insensitive to feedback control. However, like pitch, a steady state vowel is not a rapidly changing speech event, and could therefore be controlled directly by auditory feedback, in spite of the inherent neural transduction delays. In fact, in some theories of speech production, vowels do play the role of a kind of postural parameter: it has been proposed that speech is produced by adding consonant articulations to a separate, slower changing stream of vowels (Carre & Chennoukh, 1995). The issue of whether vowel production is partly controlled by direct auditory feedback must be examined in future experiments. In particular, an experiment similar to pitch perturbation could be done, looking at whether speakers produce immediate compensatory responses in their production of steady state vowels to sudden perturbations of the formants of their auditory feedback.

The amount of compensation seen varied widely across subjects, with some showing near-complete compensation and others showing little or no compensation. And yet, in the post-experiment interviews no subject reported being aware of either the altering of his feedback or his own compensatory responses to it. The interview results suggest, first of all, that the compensations and adaptations produced by subjects were not the result of conscious strategies. Moreover, these interview results also raise interesting questions about the poor compensators: (1) why did they not compensate more, and (2) why did they not report noticing the altered feedback?

Several explanations are possible. First, the post-experiment interview could have been an unreliable assessment of whether subjects were consciously aware of the altered feedback. Second, the fidelity of the feedback transform may have been imperfect for the poorly compensating subjects. Indeed, as discussed in (Houde, 1997a), there were challenging technical resolution and stability issues inherent in the method of generating the feedback transformations. The magnitude of these effects depends on the geometry of a subject's path vowels in formant space. Perhaps the poorly compensating subjects' path vowels were so arranged as to exacerbate these effects. Third, it may be that there are differences accross subjects as to the degree to which they rely on auditory feedback in their speech – either because of a subject's natural tendencies, or as a reflexive "gating out" of auditory input because of the unusual whispering conditions of the experiment. It may be that the poor compensators were ignoring their auditory feedback, and where thus insensitive to the whispered vowel sound differences created by the altered feedback. Finally, it is possible that the altered feedback may have induced adaptation of speech *perception* as well as speech production. That speech perception can exhibit adaptation has been shown in selective adaptation experiments (Cooper, 1979). Perhaps the degree of adaptation of perception versus production varies across subjects, with the poor compensators exhibiting mostly perceptual adaptation.

Two other surprising characteristics of the speech adaptation effect were revealed in the control experiment. First, the formants of the vowels produced by subjects at the beginning of the control experiment were not the same as the vowel formants produced at the beginning of the adaptation experiment, even though feedback was unaltered at this point in both experiments. Formants at the beginning of the control experiment were shifted in the direction corresponding to compensation in the adaptation experiment. Although this result is theoretically possible, since all subjects did the control experiment after the adaptation experiment, it is nevertheless surprising, given that the interval between adaptation and control experiments was about one month for most subjects. Why didn't this month of hearing unaltered feedback of their vowels reset the subjects' productions of [ε]?

One explanation is that the speech SA experimental conditions are sufficiently novel that subjects develop representations of their vowel productions that are specific to the experiment. Such priming effects of context on implicit memory have been well documented in studies of implicit memory (Schacter, 1995). For example, if subjects are presented with unrelated word pairs (e.g. Book-Forest), and subsequently tested for recall in either a same-context condition (e.g. Book-For___) or a different-context condition (e.g. Pearl-For___), recall will be primed more in the same-context condition (Graf & Schacter, 1985). However, the study results could also be explained if the control of their whispered vowels was somewhat independent of the control of voiced vowels. To some extent this must be true: control of the glottis for whispered speech is necessarily different from glottal control for voiced speech (O'Shaughnessy, 1987; Titze, 1994). In addition, as discussed above, the formants of whispered vowels are somewhat different from those of voice vowels. Thus, it may also be that control of the supraglottal articulators is also represented separately for voiced and whispered vowels. As discussed above, . If this were the case, whispered vowel representations would not be completely affected by a month of voiced vowel feedback, since whispering is an infrequent mode of speaking.

Given that subjects generally began the control experiment with their vowel formants shifted from their original (pre-adaptation experiment) values, it is also remarkable that they didn't reset their formants over the course of the control experiment (see Figure 7). It is possible that the subjects were insensitive to the amount by which their formants were shifted from their original values. However, these shifts were relatively large (almost half the distance between vowel categories in formant space), and we don't see this much variation in their production of a steady state vowel. Another explanation for the results is that subjects don't retain long-term memories of their whispered vowel sounds. Without such memories, they would not have an absolute reference from which to judge correctness of the sounds of their vowel productions. In this case, subjects' initial articulations of [ε] would set their reference memory of what [ε] should sound like. This memory would be used for the rest of the experiment to judge sound correctness of subsequent articulations of [ε]. If, as during the adaptation experiment, a feedback transform made the perceived sound of [ε] differ from the reference memory, compensating productions would be induced. However, during the control experiment, the sounds of articulations of [ε] were not altered. In this case, these [ε] sounds would not differ from the reference memory, and no production alterations would be induced.

It has been important to show, as we have done in this study, that whispered speech exhibits a type of sensorimotor adaptation analogous to the kind seen in visually guided reaches. As we have noted above, this allows the same kinds of models of motor control that describe reaching to be applied to speech motor control. However, since the common mode of speaking is voiced speech, and, indeed, since some of our study results suggest that control of whispered and voiced speech might be somewhat independent, it will be important to confirm that voiced speech exhibits the sensorimotor adaptation effects we have seen in whispered speech.

References

Borden, G. J., Harris, K. S., & Raphael, L. J. (1994). *Speech science primer : physiology, acoustics, and perception of speech* (3rd ed.). Baltimore: Williams and Wilkins.

Carre, R., & Chennoukh, S. (1995). Vowel-consonant-vowel modeling by superposition of consonant closure on vowel-to-vowel gestures. *Journal of Phonetics*, *23*(1-2), 231-241.

Cooper, W. E. (1979). *Speech perception and production: Studies in selective adaptation.* Norwood, NJ: Ablex Publishing Corp.

Cowie, R., & Douglas-Cowie, E. (1983). Speech production in profound post-lingual deafness. In M. Lutman & M. Haggard (Eds.), *Hearing science and hearing disorders* (pp. 183–231). New York: Academic Press.

Fairbanks, G. (1954). Systematic research in experimental phonetics: 1. a theory of the speech mechanism as a servosystem. *Journal of Speech and Hearing Disorders*, *19*, 133–139.

Gracco, V. L., Ross, D., Kalinowski, J., & Stuart, A. (1994). Articulatory changes following spectral and temporal modifications in auditory feedback. *Journal of the Acoustical Society of America*, *95*(5 Part 2), 2821. (Proceedings of the 127th Meeting, MIT, Cambridge, MA. June 6-10, 1994)

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *J Exp Psychol Learn Mem Cogn*, *11*(3), 501-18.

Hein, A. V., & Held, R. (1962). A neural model for labile sensorimotor coordination. In E. Bernard & M. Hare (Eds.), *Biological prototypes and synthetic systems* (Vol. 1). New York: Plenum Press.

Held, R., & Hein, A. V. (1958). Adaptation of disarranged hand-eye coordination contingent upon re-afferent stimulation. *Perceptual and Motor Skills*, *8*, 87–90.

Helmholtz, H. von. (1962). The direction of vison. In J. P. Southall (Ed.), *Helmholtz's treatise on physiological optics: Translated from the third german edition* (Vol. 3, pp. 246–247). New York, NY: Dover. (This is a republication of the English translation first published by the Optical Society of America in 1925)

Houde, J., & Jordan, M. (1998). Sensorimotor adaptation in speech production. *Science*, *279*(5354), 1213–1216.

Houde, J. F. (1997a). *Sensorimotor adaptation in speech production.* Unpublished doctoral dissertation, MIT.

Houde, J. F. (1997b). *Sensorimotor adaptation in speech production.* Unpublished doctoral dissertation, MIT.

Jordan, M. I., & Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive Science*, *16*(3), 307–354.

Kallail, K. J., & Emanuel, F. W. (1984). Formant-frequency differences between isolated whispered and phonated vowel samples produced by adult female subjects. *J Speech Hear Res*, *27*(2), 245-51.

Kawahara, H. (1993). Transformed auditory feedback: Effects of fundamental frequency perturbation. *Journal of the Acoustical Society of America*, *94*(3 Part 2), 1883. (Proceedings of the 126th Meeting, Denver, CO. Oct 4-8, 1993)

Lane, H., & Tranel, B. (1971). The lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, *14*, 677–709.

Lane, H., & Webster, J. (1991). Speech deterioration in postlingually deafened adults. *Journal of the Acoustical Society of America*, *89*, 859–856.

Lane, H., Wozniak, J., Matthies, M., Svirsky, M., Perkell, J., O'Connell, M., & Manzella, J. (1997). Changes in sound pressure and fundamental frequency contours following changes in hearing status. *J Acoust Soc Am*, *101*(4), 2244-52.

Lee, B. S. (1950). Some effects of side-tone delay. *Journal of the Acoustical Society of America*, *22*(5), 639–640.

Levitt, H., Stromberg, H., Smith, C., & Gold, T. (1980). The structure of segmental errors in the speech of deaf children. *J Commun Disord*, *13*(6), 419-41.

Lombard, E. (1911). Le signe de lelevation de la voix. *Ann. maladies oreille larynx nez pharynx*, *37*, 101–119.

Matthies, M. L., Svirsky, M., Perkell, J., & Lane, H. (1996). Acoustic and articulatory measures of sibilant production with and without auditory feedback from a cochlear implant. *J Speech Hear Res*, *39*(5), 936-46.

Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development*, *59*(2), 441-449.

Osberger, M. J., & McGarr, N. S. (1982). Speech production characteristics of the hearing-impaired. In N. J. Lass (Ed.), *Speech and language: advances in basic research and practice* (p. 221-284). New York: Academic Press.

O'Shaughnessy, D. (1987). *Speech communication: Human and machine.* Reading, MA: Addison-Wesley.

Perkell, J. S. (1996). Articulatory processes. In *The handbook of phonetic sciences* (pp. 333–370). London: Blackwell.

Polit, A., & Bizzi, E. (1979). Characteristics of motor programs underlying arm movements in monkeys. *J Neurophysiol*, *42*(1 Pt 1), 183-94.

Ross, M., & Giolas, T. G. (1978). *Auditory management of hearing-impaired children : principles and prerequisites for intervention.* Baltimore: University Park Press.

Sainburg, R. L., Ghilardi, M. F., Poizner, H., & Ghez, C. (1995). Control of limb dynamics in normal subjects and patients without proprioception. *J Neurophysiol*, *73*(2), 820-35.

Schacter, D. L. (1995). Implicit memory: A new frontier for cognitive neuroscience. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 815–824). Cambridge, MA: MIT Press.

Smith, C. R. (1975). Residual hearing and speech production in deaf children. *J Speech Hear Res*, *18*(4), 795-811.

Stevens, K. N. (1999). *Acoustic phonetics.* Cambridge, MA: MIT Press.

Titze, I. R. (1994). *Principles of voice production.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Welch, R. B. (1978). *Perceptual modification: Adapting to altered sensory environments.* New York: Academic Press.

Welch, R. B. (1986). Adaptation of space perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1: Sensory Processes and Perception, pp. 24-1 – 24-45). New York, NY: John Wiley and Sons.

Yates, A. J. (1963). Delayed auditory feedback. *Psychological Bulletin*, *60*(3), 213–232.

Figure Captions

Figure 1: Overview of the experimental apparatus: Words are presented on a PC video monitor for the subject to pronounce. The subject's whispered speech is transduced by the microphone and fed as input to a digital signal processing (DSP) board inside the PC. The DSP repeatedly (a) Captures from the microphone an 8ms frame of the subject's whispered speech (64 time samples at an 8KHz sampling rate); (b) Performs a 64-channel spectral analysis of this frame, retaining only a smoothed magnitude spectrum of it; (c) Estimates the first four formants from the magnitude spectrum; (d) Alters the frequencies of the three lowest formants; (e) Resynthesizes a new 8ms frame of whispered speech from the altered formants, and returns this altered feedback to the subject via the insert earphones. It also records the formants of subject's utterances (Houde & Jordan, 1998).

Figure 2: Construction of the feedback transformations used in the experiments: (a) Image shift by prisms: Gray arrows show the image-shifting action of prisms worn by subject. When the subject reaches to points P1 and P2, he instead sees shifted images of his hand at points P1' and P2'. With respect to the reference path (dashed line), this image shift preserves the how much the subject's hand deviates from the path, but subtracts 2.0 units from the hand's apparent distance along the path. The black arrow shows a typical compensation response: After several reaches for point P2, subject's hand actually arrives at point P2c – whose image P2c' is closer to the intended target P2. (b) Idealized audio analog of prism image shift: Gray arrows show an audio feedback transformation that shifts vowel sounds along the subject's [i]–[ɑ] path – the numbered path formed by the subject's productions of the vowels [i], [ɩ], [ɛ], [æ], and [ɑ](dashed line). When a subject produces vowel sound near [ɛ] (e.g., V1 or V2), the audio transformation subtracts 2.0 units from the vowel sound's distance along the [i]–[ɑ] path, but preserves its deviation from the path. As a result, the subject's audio feedback is a vowel sound near [i] (e.g., V1' or V2'). The black arrow shows a

possible compensation response: After several productions of vowel sound V2, subject shifts his production to vowel sound V2c. As a result, the subject's audio feedback V2c' is now closer to the intended vowel sound V2. (c) Feedback transformation used in the experiments: Dashed line shows subject OB's [i]–[ɑ] path in (F1,F2) space. This path is distorted from the idealized [i]–[ɑ] path of Figure 2(b): the path is not straight, and the distance between vowels on the path is variable. A vowel sound's distance along the path – called *path projection* – is the nearest point on the [i]–[ɑ] path to the vowel and this distance is normalized to be the same between all adjacent vowels on the path. In this figure, the gray arrows show the action of the -2.0 transformation – one of the two formant-shifting audio transformations used in the experiments. The black arrow again shows a possible compensation response, analogous to that shown in Figure 2(b).

Figure 3: Experimental design: PC video screen prompted the subject to whisper the printed word with a 300ms target duration (as indicated by the length of the bracket beneath the word). Dots appearing as the subject whispered indicated when he had reached this target duration. The word promptings came in groups of 10 called *epochs*. The first 6 prompted words came from a training word set ($\mathbf{W_{train}}$) and the last 4 came from a test word set ($\mathbf{W_{test}}$). Via the earphones, the subject heard feedback of his first 5 utterances but heard only masking noise after that. Thus the subject never heard his production of the test words, which were used to test how training word adaptation generalized. The experiment consisted of 422 epochs, organized into a sequence of five *phases* whose durations are shown in the figure: a **warmup phase** to acclimate subject to experiment; a **baseline phase** with feedback unaltered, for later comparison with test phase; a **ramp phase** in which the feedback transform was linearly ramped up to full strength; a **train phase** for extended exposure to the altered feedback; a **test phase** with feedback still altered, for comparison with baseline phase. This sequence of phases was exactly the same in the control experiment, with the only difference that no feedback transformation was ramped in. As the

figure shows, durring the adaptation experiment, half the subjects were exposed to the -2.0 feedback transformation and the rest were exposed to the +2.0 transformation.

Figure 4: Subject OB's responses in the control and adaptation experiments: Gray arrow running along OB's [i]–[ɑ] path shows the action of the -2.0 feedback transformation used in the adaptation experiment. The base of the arrow labeled "Baseline" represents the mean (F1,F2) values of OB's production of the vowel /eh/ in the baseline phase of the experiment. The arrow tip represents the effect of this transformation on OB's audio feedback of his baseline production of [ɛ] – it preserves the path deviation of 70 Hz, but reduces the path deviation to 1.2, resulting in a vowel sound close to [i]. The black arrow shows OB's compensation response to the feedback transformation: it shows OB altered his production of [ɛ] to oppose the action of the feedback transformation. The tip of the arrow labeled "Test" represents the mean (F1,F2) values of OB's production of [ɛ] by the end of the experiment, in the test phase. The white arrow shows OB's adaptation response to the feedback transformation: it shows how his production of [ɛ] without audio feedback (i.e., blocked by masking noise) changed from the baseline to the test phase of the experiment (Houde & Jordan, 1998).

Figure 5: Mean compensation and path deviation change for each subject: The left column shows plots of subjects' *compensation responses*: vowel production changes observed when subjects could hear feedback of their whispering. The right column shows plots of subjects' *adaptation responses*: vowel production changes that subjects retained when prevented from hearing their whispering by masking noise. Plots (a) and (b) show mean compensation and path deviation change for each subject's compensation response. Plots (c) and (d) show the same for each subject's adaptation response. In each plot, black dots linked by a solid line indicate adaptation experiment data, while white dots linked by a dotted line indicate control experiment data. Small bars around each dot indicate confidence intervals.

Figure 6: Path projections of each subject's vowel productions in the baseline and testing phases: Plots (a) and (b) show path projections for each subject's compensation response. Plots (c) and (d) show the same for each subject's adaptation response. In each plot, black dots linked by a solid line indicate adaptation experiment data, while white dots linked by a dotted line indicate control experiment data. Small bars around each dot indicate confidence intervals.

Figure 7: Path projections of Figure 6 averaged across subjects: Plot (a) shows average path projections for subjects' compensation responses. Plot (b) shows the same for subjects' adaptation responses. In each plot, the filled and open dots above the "base" label are the average path projections seen in the baseline phase of the adaptation and control experiments, respectively. The dots above the "test" label are the average path projections seen in the test phase. The solid and dotted lines connecting the dots highlight the average path projection changes seen in the adaptation and control experiments, respectively (Houde & Jordan, 1998). (Note: confidence intervals for each average are shown but are so small that the bars representing them are obscured by the dots.)

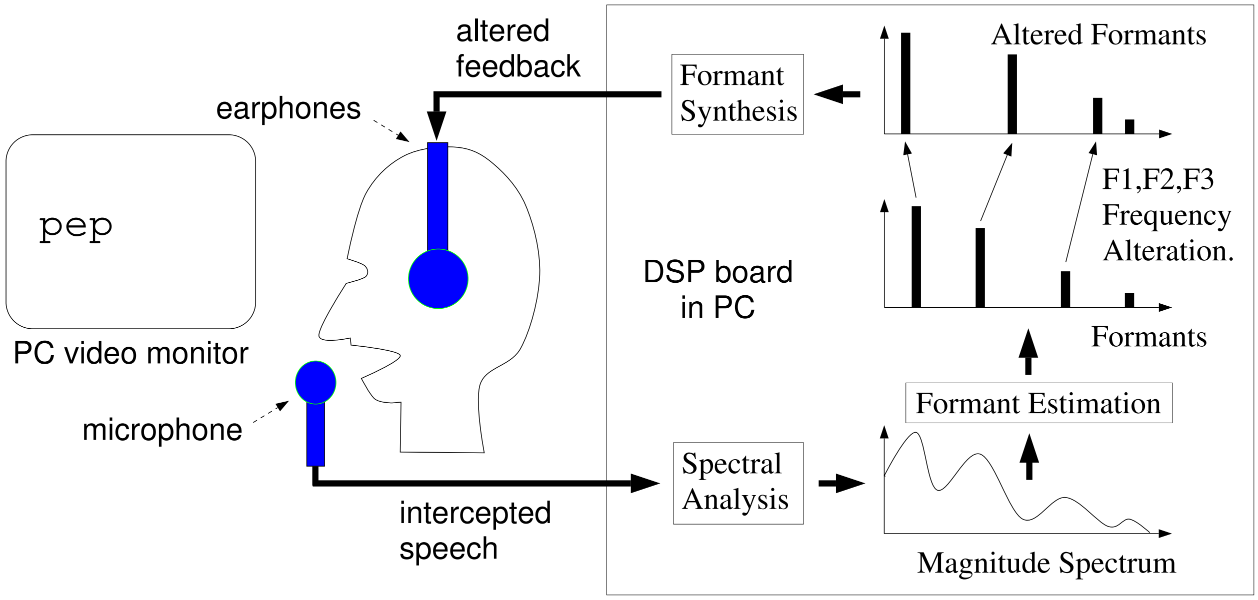
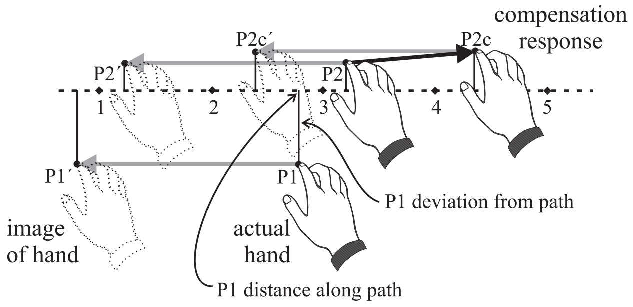
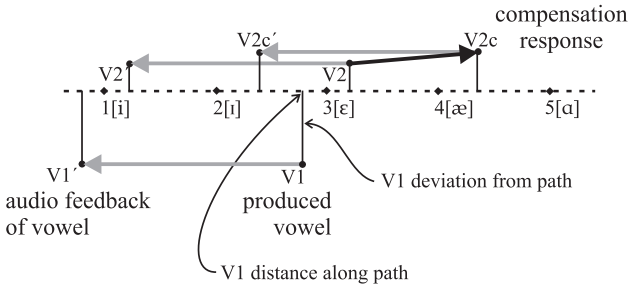

altered
feedback
earphones
pep
PC video monitor
microphone
intercepted
speech
Formant
Synthesis
Altered Formants
F1,F2,F3
Frequency
Alteration.
DSP board
in PC
Formants
Formant Estimation
Spectral
Analysis
Magnitude Spectrum

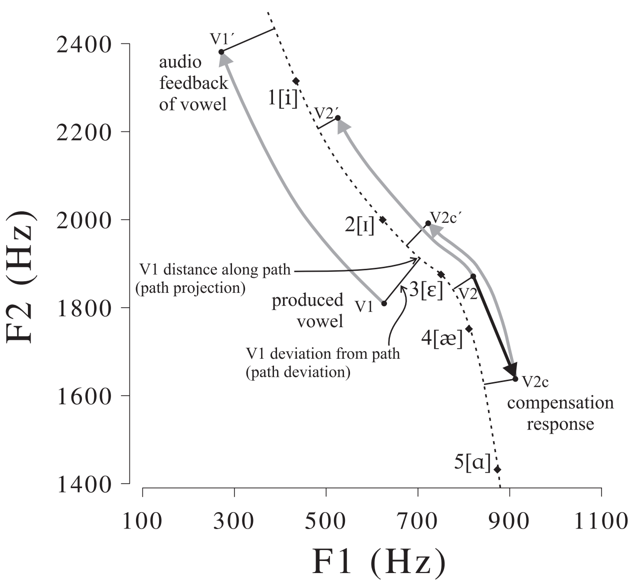
Figure 1:



(a) image shift by prisms



(b) audio analog of prism image shift



(c) feedback transformation used in the experiments

Figure 2:

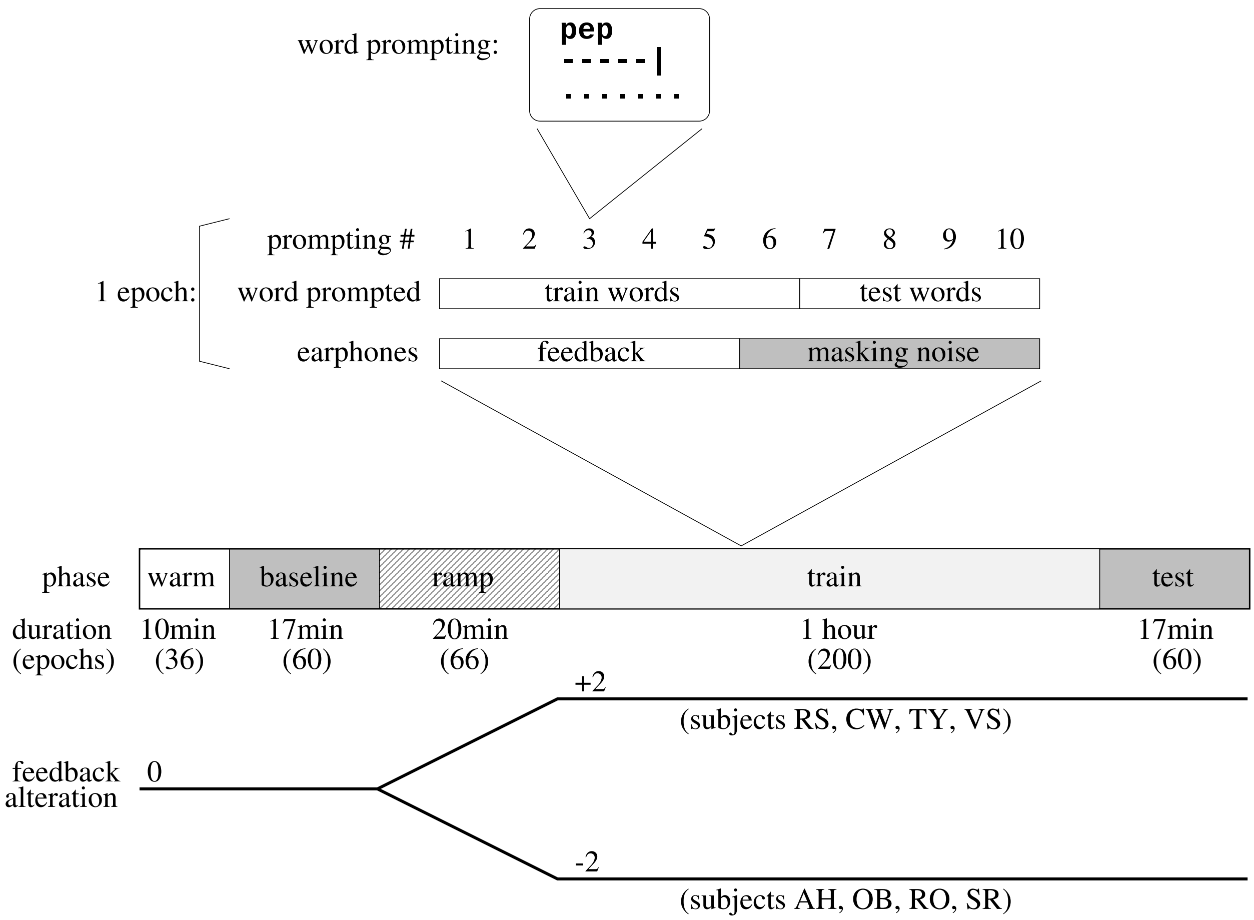

word prompting:
pep
- - - - - |
. . . .
prompting # 1 2 3 4 5 6 7 8 9 10
1 epoch:
word prompted
train words
test words
earphones
feedback
masking noise
phase
warm
baseline
ramp
train
test
duration (epochs)
10min (36)
17min (60)
20min (66)
1 hour (200)
17min (60)
feedback alteration
0
+2
(subjects RS, CW, TY, VS)
-2
(subjects AH, OB, RO, SR)

Figure 3:

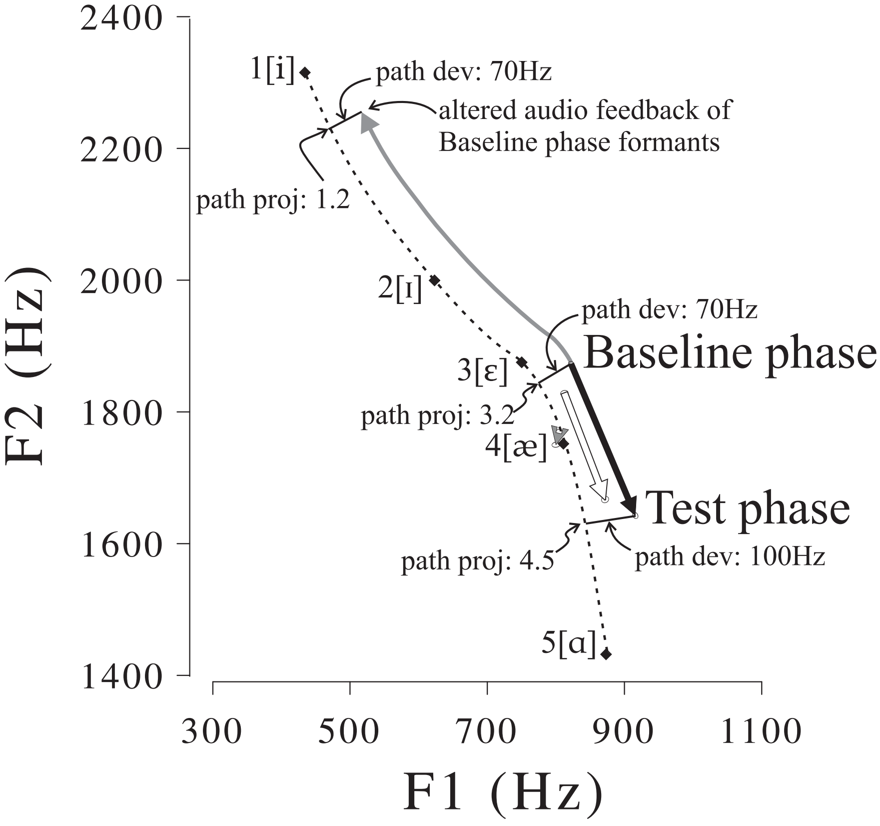
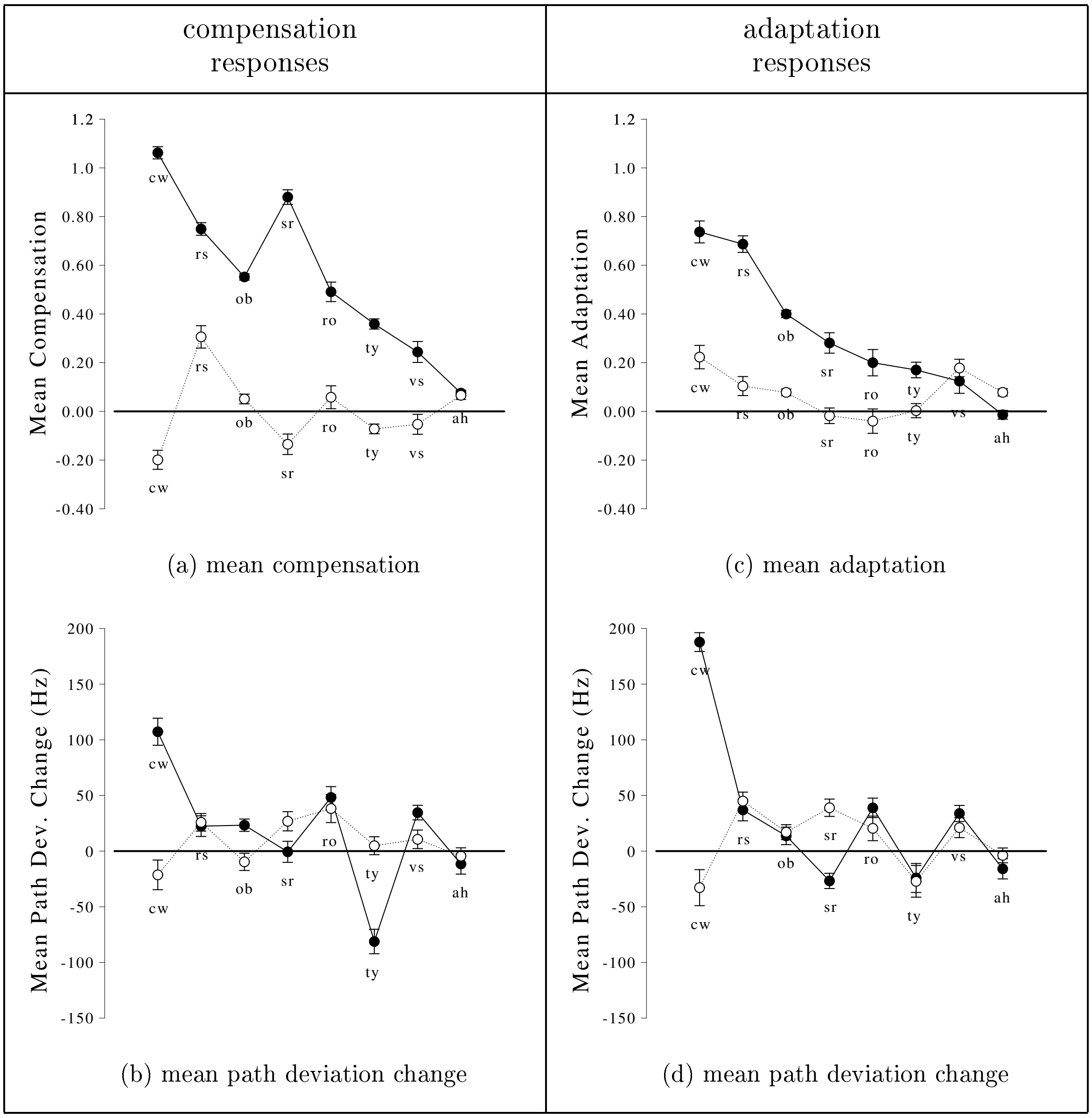

2400
2200
2000
1800
1600
1400
F2 (Hz)
300
500
700
900
1100
F1 (Hz)
1[i]
2[ɪ]
3[ɛ]
4[æ]
5[ɑ]
path dev: 70Hz
altered audio feedback of
Baseline phase formants
path proj: 1.2
path dev: 70Hz
Baseline phase
path proj: 3.2
Test phase
path proj: 4.5
path dev: 100Hz

Figure 4:



(a) mean compensation

(c) mean adaptation

(b) mean path deviation change

(d) mean path deviation change

Figure 5:

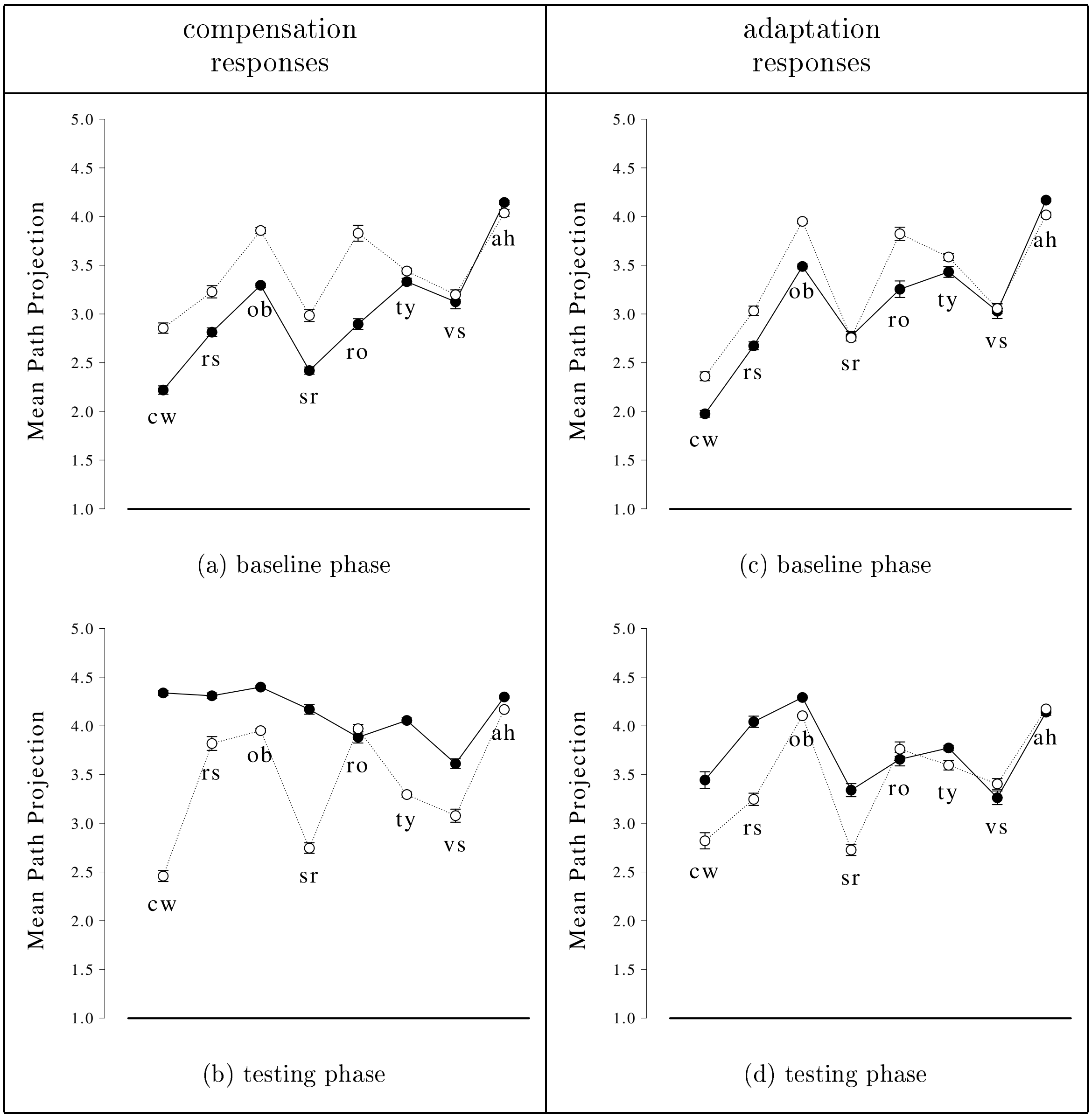
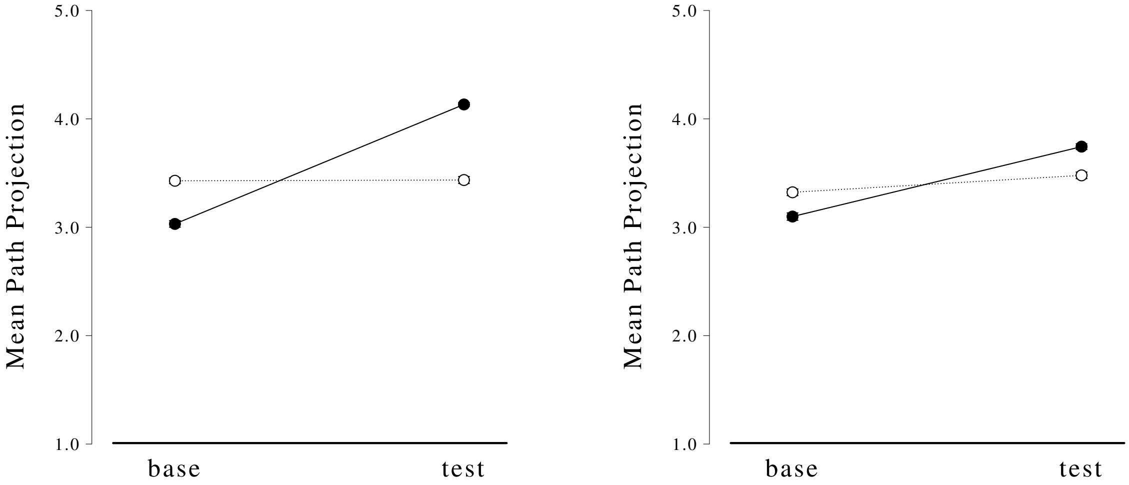

compensation responses
adaptation responses
Mean Path Projection
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
cw
rs
ob
sr
ro
ty
vs
ah
(a) baseline phase
(b) testing phase
(c) baseline phase
(d) testing phase

Figure 6:



(a) compensation responses

(b) adaptation responses

Figure 7: