Published in final edited form as:

Brain Res. 2009 June 18; 1276: 131–139. doi:10.1016/j.brainres.2009.04.033.

Control of vocalization at utterance onset and mid-utterance: different mechanisms for different goals

Colin S. Hawco and Jeffery A. Jones

Centre for Cognitive Neuroscience and Department of Psychology, Wilfrid Laurier University, Waterloo, Ontario, N2L 3C5, Canada

Abstract

A large body of evidence suggests that the motor system maintains a forward model that predicts the sensory outcome of movements. When sensory feedback does not match the predicted consequences, a compensatory response corrects for the motor error and the forward model is updated to prevent future errors. Like other motor behaviours, vocalization relies on sensory feedback for the maintenance of forward models. In this study, we used a frequency altered feedback (FAF) paradigm to study the role of auditory feedback in the control of vocal pitch (F0). We adapted subjects to a one semitone shift and induced a perturbation by briefly removing the altered feedback. This was compared to a control block in which a 1 semitone perturbation was introduced into an unshifted trial, or trials were randomly shifted up 1 semitone, and a perturbation was introduced by removing the feedback alteration. The compensation response to mid-utterance perturbations was identical in all conditions, and was always smaller than the compensation to a shift at utterance onset. These results are explained by a change in the control strategy at utterance onset and mid-utterance. At utterance onset, auditory feedback is compared to feedback predicted by a forward model to ensure the pitch goal is achieved. However, after utterance onset, the control strategy switches and stabilization is maintained by comparing feedback to previous F0 production.

1. Introduction

During speech, accurate motor control of the vocal articulators is critical for producing an intelligible utterance. In some models, initial motor control is believed to be driven by paired forward and inverse motor models (Wolpert and Kawato, 1998). Forward models predict the next state (position and velocity) of the system based on its current state and active motor command. In contrast, inverse models provide the motor command that will cause the desired change in state. Forward models are modified when differences between the predicted outcome of an action and the actual outcome of that action are detected using available feedback mechanisms. During vocalization, feedback takes the form of kinesthetic awareness of the position of the articulators and auditory feedback from the utterance. The forward model includes an “efference” comparator, a copy of the motor command sent to the sensory cortex to allow a prediction of the sensory consequences (Nowak et al. 2007). When feedback does not match the efference copy during a vocal production, a compensation response occurs to correct for the error, and the forward model is modified to correct future utterances.

*Author to whom correspondence should be addressed: Jeffery A. Jones, Department of Psychology, Wilfrid Laurier University, Waterloo, Ontario, N2L 3C5, Canada, E-mail: E-mail: jjones@wlu.ca, Tel: 519-884-0710 ext: 2992, Fax: 519-746-7605.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The plasticity of the forward model can be investigated by exposing speakers to altered auditory feedback, and tracking changes in vocalization. When a feedback alteration is encountered, a compensation response corrects for this feedback error by adjusting production in the direction opposite to the alteration. Adaptation occurs when speakers are exposed to a feedback alteration for a prolonged period. Evidence that adaptation has occurred is obtained when the alteration is removed and aftereffects persist for some period of time. Adaptation during vocalizations has been observed in several studies. For example, Houde and Jordan (1998, 2002) examined sensorimotor representations for formants by shifting F1 and F2 for the vowel /ε/ along the /i/ - /D/ axis. Subjects compensated for the feedback alterations by modifying their formant production. These modifications persisted when auditory feedback was removed, demonstrating adaptation occurred within the motor system. Similarly, Purcell and Munhall (2006) gradually shifted F1 during vowel production, and found a gradual return to baseline (de-adaptation) when feedback was abruptly returned to normal. The time of this de-adaptation response was not related to amount of time that the maximal feedback alteration was maintained.

Manipulations of auditory feedback for fundamental frequency (F0), or vocal pitch, have also been used to study sensorimotor control during vocalizations. F0 is distinct from formant frequencies in that formants must be controlled within each vowel while F0 appears to be controlled suprasegmentally, at least for non-tone languages (Natke and Kalveram, 2001, Perkell et al 1997). Jones and Munhall (2000, 2002) slowly shifted F0 over many trials and observed a compensation response in the direction opposite the shift. When the feedback alteration was removed, they found prominent aftereffects that suggested a remapping in the motor system for F0 control. Examining differences in singers and non-singers, Jones and Keough (2008) introduced an abrupt (as opposed to gradual) feedback change. They found non-singers adjusted to the feedback alteration almost immediately, while singers, who possess superior F0 control, were slower to modify their productions. Unlike Purcell and Munhall (2006), they did not observe a gradual de-adaptation in non-singers, who returned to baseline within a few trials. In contrast, singers were slower to adapt and slower to de-adapt than non-singers, suggesting that their internal models were more entrenched than those of non-singers.

While adaptation studies involve prolonged feedback alterations across an entire utterance, studies of vocal control have also examined brief perturbations within a prolonged utterance (Elman, 1981; Larson, 1998; Burnett et al., 1998; Natke and Kalveram, 2001). Burnett et al. (1998) shifted F0 by varying magnitudes (25, 50, 100, 150, 200, 250, or 300 cents) for a period of 100 to 500 ms within a prolonged utterance. They found a transient compensation response, with the magnitude and duration of the response being modulated by the perturbation. This compensation response has been termed the pitch-shift reflex. The pitch-shift reflex is distinct from the adaptation response described above, as the pitch-shift reflex reflects an online compensation process using ongoing auditory feedback when an unexpected feedback error is encountered, while adaptation represents an updating of the forward model in response to a predictable constant feedback alteration. The results of Burnett et al. (1998) suggested that the pitch-shift reflex is made up of two components, an early and a late response. In a follow up study, Hain et al. (2000) used a 500 ms perturbation, and asked subjects to compensate, not to compensate, or actively follow the pitch-shifted feedback when their voice was perturbed. They found that the early component of the pitch-shift reflex was automatic and not affected by task instructions while the late component was under volitional control. Munhall et al. (2009) conducted a study in which they shifted formants during production of a monosyllabic word ("head"). They found that participants compensated even when the manipulation was described to participants prior to the experiment, and participants were explicitly instructed not to compensate. These results suggest automatic (non-volitional) control of formants in short utterances, similar to a short perturbation during vowel production.

One important aspect of the pitch-shift reflex is that it is generally smaller than the perturbation, except for very small feedback alterations (Burnett et al., 1998; Larson et al. 2001; Liu and Larson, 2007). In contrast, responses to whole-utterance shifts are often close or equal in magnitude to the feedback alteration. Jones and Keough (2008) had a series of baseline trials, after which they exposed participants to a 100 cent (one semitone) feedback alteration. On the first trial after feedback alteration, both singers and non-singers adjusted their F0 by more than 50 cents; in fact non-singers adjusted their F0 by approximately 70 cents. Moreover, non-singers achieved full compensation to the feedback alteration within 5 or 6 trials. This approximately 70 cent response is much larger than that typically observed in the pitch-shift reflex, where a 100 cent perturbation generally results in compensation responses of less than 50 cents, and sometimes as little as 9 cents (Liu and Larson, 2007).

It has been theorized that an efference copy comparator is used to monitor auditory feedback, and that this comparator plays an important role in motor learning (Nowak et al. 2007). The efference copy is sent from the motor system, by the forward model, to the auditory system so that sensory feedback can be predicted. That prediction is compared to the perceived auditory feedback regarding the vocalization; when feedback does not match the prediction, an error is registered. This error is corrected online, and the forward model is subsequently modified. Some evidence for the use of an efference copy in monitoring auditory feedback has come from studies using event-related potentials (ERPs). Hienks-Maldonado et al. (2005) found the N100, an ERP component associated with sensory processing of a stimulus, was attenuated during vocalization. When feedback was altered, the N100 attenuation was reduced. This attenuation is believed to be driven by the efference copy. The auditory cortex is maximally attenuated when feedback matches the efference copy, allowing speakers to determine that what they are hearing is their own voice, and that no errors occurred during vocal production. Importantly, the N100 is only observed at utterance onset.

Larson et al. (2001) conducted a study in which they either altered feedback in mid-utterance, or altered feedback at utterance onset, and then removed the feedback alteration in mid-utterance. They found an identical mid-utterance compensation response in both conditions. That is, the compensation response always stabilized production so that the current F0 value matched the F0 prior to the alteration, independent of whether the change detected was initiation of a feedback alteration or its removal. The authors suggest that this finding rules out the use of a fixed efference copy as the reference used in voice stabilization, as such a system should not stabilize the system to artificially altered feedback. However, subjects were instructed to produce at a habitual F0, and the baseline F0 was not known. Thus, it is possible that subjects responded to the initial feedback alteration with a compensation that was equal in magnitude to the compensation to the mid-utterance feedback alteration. When the feedback alteration was removed, they then simply stopped compensating and their F0s returned to their baseline value, resulting in a response equal to that observed when a perturbation was initiated mid-utterance. We have some evidence that responses to initial F0 shifts may be larger than those typically observed for mid-utterance F0 shifts (Jones and Keough, 2008), but that study used a fixed external reference in the form of a target note. Speakers attempting to match an external reference might produce larger compensation responses than speakers who are not required to achieve a specific pitch target.

Hawco et al. (2009) examined the response to a mid-utterance perturbation using ERPs. Rather than observing an early effect that replicated the N100 results of Hienks-Maldonado et al. (2005), they found a mismatch negativity (MMN), a later ERP component associated with a violation of a sensory memory trace. The MMN is generally observed when a memory trace is formed for a stimulus, or a stream of stimuli, and a stimulus is encountered which violates that memory trace (see Näätänen, et al., 2007, for review). Kudo et al. (2004) found that the N100 response to a tone was reduced during vocalization, while the MMN was unaffected,

suggesting that the efference copy is related to the N100 but not to the MMN. The MMN observed by Hawco et al. (2009) suggests that the error detected in a mid-utterance perturbation was compared to some echoic memory representation, such as the unshifted baseline within the utterance, rather than an efference copy.

In the present study, we compared the compensation response at utterance onset to that at mid-utterance, to determine if these responses were equivalent. Fifteen women heard a target female voice vocalizing the vowel /a/ at a specific frequency (D4, 296.33 Hz) and were asked to produce the vowel at the same pitch. In two blocks, speakers' auditory feedback was either shifted up by 1 semitone (100 cents) *mid-utterance*, or shifted up 1 semitone at *utterance onset*, and then perturbed by removing this feedback alteration. Specifically, in the control block, feedback was either shifted upward 100 cents in pitch mid-utterance (Control Perturb trials), or randomly shifted upward 100 cents at utterance onset with the feedback alteration either maintained throughout the entire utterance (Onset No-perturb trials) or perturbed by removing the feedback alteration for 500 ms (Onset Perturb trials). In a separate block of trials, we adapted participants to a feedback alteration by introducing a constant and predictable feedback alteration prior to utterance onset (100 cent shift up), and perturbed their voice by removing the feedback alteration during some trials (Adapt Perturb trials). The Onset Perturb and Adapt Perturb trials had identical feedback alterations (see Figure 1), although the Adapt trials were predictably shifted at utterance onset, while the Onset trials were randomly shifted within the control block. If the pitch-shift reflex is a pure maintenance response, and uses the current F0 as its reference (even when an absolute referent is available in the form of a target F0), we should observe a response of similar magnitude regardless whether the perturbation is the sudden onset of a feedback alteration in mid-utterance, or the removal of a feedback alteration that was introduced at utterance onset. If, on the other hand, an efference comparator is used, the size of the pitch-shift response to a feedback alteration removal should match the size of a compensation response to the perturbation at utterance onset.

2. Results

Mean F0 across the entire vocalization, and median F0 for the initial 50 ms of the adaptation and control blocks are shown in Figure 2. Three subjects were excluded from the analysis: two because they showed no evidence of adaptation effects (subjects 11 and 14) and one subject was removed because they showed a following response, increasing their F0 throughout much of the adaptation block (subject 15). Adaptation was indexed by examining the F0 during the first 50 ms in each trial. The initial 50 ms of vocalization is driven by purely feed-forward controllers, as feedback is not available for at least 100 ms (Burnett et al., 1998). F0 data during later parts of the utterance can be influenced by both feed-forward controllers (because they control the initial F0 of the utterance) and feedback controllers, as F0 can be modified using auditory feedback during the later portions of vocalization (after 100 to 150 ms after utterance onset). If adaptation occurred, we expected to see a systematic shift in F0 at utterance initiation (the first 50 ms) across trials. If adaptation did not occur the initial F0 of each utterance would be equivalent.

To test for adaptation, we compared the median 50 ms data for trials 11 to 20 (the last 10 trials prior to F0 feedback alteration in the adaptation block; see Figure 1) to trials 51 to 60 (the 10 utterances prior to the onset of perturbations). These trials were compared across the adaptation and control blocks. A block (adaptation or control) by time (trials 11 to 20 or trials 51 to 60) repeated-measures ANOVA showed a main effect of block, $F(1,11) = 9.7$, $p = 0.0097$, and a block by time interaction, $F(1,11) = 4.8$, $p = 0.049$. Tukey's HSD post-hoc analysis of the 2-way interaction showed that trials 51 to 60 of the adaptation block differed from the other tested conditions, and no other conditions differed among themselves. This pattern of results indicates

that participants modified the F0 of the onset of their utterances due to the brief exposure to FAF during trials 21–50 of the adaptation block (see Figure 2b).

Mean compensation responses (in cents) for the Control Perturb, Onset Perturb and Adapt Perturb trials are shown in Figure 3. F0 values have been converted into cents relative to the target note (296.33 Hz). The pre-stimulus baseline in Figure 3 shows the baseline F0 prior to the mid-utterance perturbation. To measure the size of the compensation responses, area under the curve of the normalized compensation response (where the baseline was normalized to zero, as described below) was calculated. Because we were interested in comparing compensation responses in the presence of adaptation to compensation responses in the control and onset conditions, the three participants excluded from the adaptation analysis were also excluded from the compensation analysis. The compensation response was divided into two time windows, 100 to 250 ms post-perturbation, and 250–600 ms post perturbation. This windowing procedure was done to separate the earlier, automatic component of the pitch-shift reflex from the later, volitionally controlled component (Hain et al., 2000). For the first time window (100 ms to 250 ms post-perturbation, evaluating the pitch-shift reflex) a perturbation (Perturb or No-perturb) by condition (Adapt, Control, or Onset) repeated-measures ANOVA was conducted. An effect of perturbation was found, $F(1,11) = 18.2$, $p = 0.0013$, indicating that participants responded to the feedback alteration, but no main effect of condition was observed, $F(2,22) = 1.2$, $p = 0.30$. As well, no significant interaction was observed indicating that the magnitude of the compensation responses was similar across all conditions. The second time window was from 250 to 600 ms. Again, a main effect of perturbation was observed, $F(1,11) = 27.1$, $p = 0.00028$. In addition, a main effect of condition, $F(2,22) = 4.8$, $p = 0.018$, and an interaction existed, $F(2,22) = 3.4$, $p = 0.048$, indicating that the late compensation response differed across conditions. Tukey's HSD post-hoc analysis of the interaction indicated that when perturbations were present, the adaptation and control conditions differed ($p = 0.021$), but that the onset condition differed from neither the control ($p = 0.89$) nor the adaptation condition ($p = 0.18$).

Visual inspection of the data in Figure 3 suggests that the F0 change in response to the feedback alteration at utterance onset (in the Onset Perturb trials) is much larger than the compensation response to the mid-utterance perturbation. Table 1 compares the mean F0 data (in cents) of the pre-shift baseline period (the 150 ms preceding the perturbation) to the first time window (100 ms to 250 ms post-perturbation), and clearly shows that the post-perturbation response does not approach the level of the control baseline. If the mid-utterance compensation responses were equal to the response to the onset, it would suggest that participants were returning their voice to pre-feedback shift baseline, and that the compensation response to the onset was the same as the response to the perturbation. If subjects were returning to their baseline F0 when we removed the feedback alteration in the onset trials (the perturbation), then the compensation response should reach a value equal to the control trials with no perturbations. Figure 3 shows that the compensation responses in the onset and adaptation trials were much smaller than the difference between the baseline, pre-perturbation values for Control and Onset or Adapt trials, indicating that participants shifted their voices more for the initial feedback alteration (present before they began their utterance) than when the feedback alteration was removed. Note that the pre-perturbation baseline shown within Figure 3 and Table 1 represents F0 values within the middle of the utterance, not at utterance onset.

We examined F0 when both predictive initial F0 changes due to adaptation effects and compensatory changes to F0 using auditory feedback after utterance onset have occurred by comparing the mean F0 values from 100 ms to 600 ms post-perturbation for the Onset Perturb and Adapt Perturb trials to the Control No-Perturb trials, which represents the baseline F0 value from which the other two conditions should differ. We found a significant difference, $F(2,22) = 15.1$, $p = 0.000072$. Tukey's HSD post-hoc showed that the control data were significantly

different than the adaptation ($p = 0.00022$) and onset ($p = 0.00096$) compensation responses. The adaptation and onset compensation responses were not found to be different. This demonstrates that the compensation to a feedback shift at utterance onset is larger than the compensation to a mid-utterance perturbation.

It is possible that Onset trials may have an effect on subsequent trials. In order to assess single trial adaptation effects, we examined the first 50 ms of utterances of Control trials that *followed* either Onset trials or Control trials in the control block. Because it was possible to have 2 or 3 consecutive onset trials, we only examined Control trials that followed a single onset trial (i.e. onset trials that were preceded by a control trial) to avoid the inclusion of adaptation effects that resulted from consecutive presentation of Onset trials (e.g., Jones & Munhall, 2000). Likewise, we only examined trials following control trials that were preceded by an Onset trial. Significant differences were observed between trials following onset trials and trials following control trials, $F(1,11) = 7.7$, $p = 0.017$, with the F0 of the initial 50 ms being lower in trials following Onset trials. This result shows that the onset trials had an effect on the following trial, suggesting some rapid, single-trial adaptation occurred.

3. Discussion

The results of this study demonstrate that a mid-utterance perturbation results in an identical compensation response when the perturbation is the introduction of a feedback alteration, the removal of a randomly occurring feedback alteration, or the removal of a feedback alteration after adaptation has occurred. This is similar to the results of Larson et al., (2001), who found identical magnitudes of compensation when a feedback alteration was induced, or when an existing feedback alteration was removed. More importantly, it was found that the compensation response to a perturbation at utterance onset is much larger than the compensation response to a feedback alteration within an ongoing utterance. Overall, this pattern of results suggests differences in the mechanism used to evaluate F0 feedback at utterance onset and mid-utterance.

3.1 Mid-utterance compensation responses

We found similar compensation responses in all conditions for the early, automatic part of the mid-utterance compensation. While differences were found in the later phase of the mid-utterance compensation, these are less interesting as this part of the compensation response is subject to volitional control, and it is difficult to determine what factors might contribute to the observed differences. Of particular interest is the Onset Perturb condition. In this condition, we initiated a feedback alteration at utterance onset, and removed it in mid-utterance for a 500 ms period. In this case, as in the adaptation trials, subjects treated their unaltered feedback as an error and responded to it by compensating in the opposite direction. It is clear that this mid-utterance compensation response is not a 'switching-off' of the compensation to the perturbation at utterance onset, because if that were the case, we would have observed a much larger compensation in the Onset than Control conditions as F0 returned to the control-baseline values.

The mid-utterance compensation response may be driven by one of two possible mechanisms: comparison to a relative reference where the current F0 at the time of feedback alteration represents the goal, or to an absolute reference in which there is a specific, fixed F0 that represents the pitch goal. An efferent comparator is a form of internal absolute reference, in which feedback is compared to an efference copy of the motor command for a specific F0. In the current study, we introduced an absolute external reference by asking subjects to match a specific pitch value. If subjects were using an absolute referent they should show a larger response in the Onset condition (relative to the control) as they re-adjust their F0 back to the control baseline, removing their large compensation response to the perturbation at utterance

onset. Larson et al. (2001) found that when speakers began an utterance under altered feedback and heard the alteration removed, they produced the same compensation response as when they heard their feedback altered part way through their utterance. The authors suggest that subjects used an internal, variable reference when no absolute reference was available. We found no differences in compensation responses even when we added an absolute external reference. This suggests that, when maintaining a steady F0, the reference is always internally based on the current, pre-shift F0. According to this hypothesis, the purpose of the compensation response is not to attain a specific pitch goal, but to adjust for unintentional fluctuations within F0 during an utterance.

3.2 Differences in comparators at onset or mid-utterance

Figure 3 clearly shows that the compensation response to the feedback alteration at the beginning of the Onset and Adapt trials was larger than the compensation to the mid-utterance perturbation. It should be noted that the pre-perturbation baseline for the Adapt trials may include changes in F0 from both feedforward adaptation and feedback based compensation, while the pre-perturbation baseline to the Onset trials represents a purely compensation based response, as the Onset trials are randomized and adaptation should not occur. The compensation response at utterance onset is much larger than the mid-utterance compensation. As discussed above, we have no evidence that an efference comparator is being used in the mid-utterance perturbations. However, several studies have suggested that efference copy plays a role at utterance onset (Curio et al., 2000, Houde et al., 2002, Hienks-Maldonado et al., 2005). For example, Hienks-Maldonado et al. (2005) found that N100 attenuation related to the efference copy was reduced when auditory feedback was altered, suggesting that the efference copy may be used as a means to detect errors within one's voice. This is in contrast to the results of Hawco et al. (2009), who found an MMN, rather than an N100, to mid-utterance perturbations. The fact that Hawco et al. (2009) observed an MMN, and no early activity similar to the N100 effects from Hienks-Maldonado et al. (2005) suggests that feedback at utterance onset and during mid-utterance may be monitored using different mechanisms. However, it should be noted that Heinks-Maldonado et al. (2005) were only interested in observing the perception of the speaker's own voice, and not in linking that perception to vocal control.

At utterance onset, the goal is to match a specific and pre-planned F0, adjusting onset F0 to match a specific goal (e.g., a target note). Feedforward mechanisms must be used at utterance onset to hit an F0 target, or we would not observe a change in initial F0 following trials with altered feedback. In the present study, subjects' initial utterance F0s were modified by repeated exposure to the feedback alteration indicating that the subjects' feedforward plans were modified. However, this feedforward system was not highly accurate at attaining the specific target immediately at utterance onset. Thus, feedback was also used to adjust for errors after utterance onset (initial F0 was lower than final F0, indicating a searching strategy was used). When feedback is used during utterance onset to reach a desired F0, an absolute reference must be used, as a variable reference would result in large and unpredictable errors. Indeed, it is difficult to imagine from where a variable reference would be derived at utterance onset, when no auditory feedback is available. Two possible absolute references exist at utterance onset during the task in the present experiments. The first is the efference copy, which is known to respond to changes in auditory feedback (Heinks-Maldonado et al., 2005). The second is a memory trace of the target note. It is difficult in the context of the current study to rule out the second possibility, though we believe that an efference comparator is more likely as it should exist during all vocalizations, including during normal speech, where an external target F0 does not generally exist. One way to test this hypothesis may be to compare the compensation response for a feedback alteration at utterance onset with and without a target note. If a memory trace of the target note were used as the comparator, we would expect to see a larger compensation at utterance onset when a target is present than when one is not. It is also possible

that both an efference comparator and a memory trace of the target note are being used in concert, or that individuals may vary in their reliance on either of these possible comparators.

When the target F0 has been matched after utterance initiation, a goal switching takes place, in which the goal shifts from target matching (be it matching to an efference copy or to a memory trace of the target) to pitch maintenance. This maintenance mechanism drives the pitch-shift reflex, as described above. In other words, there is a switch from an efference or memory trace comparator to a current F0 comparator/voice stabilizer. When an alteration is encountered, the pitch-shift reflex serves to stabilize the voice to the current F0. We therefore arrive at a separation of the mechanisms used at utterance onset to correct for feedback errors and those used after onset to stabilize F0 during an utterance. Such a switch in mechanisms of F0 control might serve two purposes. Firstly, maintaining an absolute reference (efference or memory trace) requires the allocation of additional neural resources. Secondly, the pitch-shift reflex is a relatively small response, ideal for maintaining a stable F0 during production (Hain et al., 2000)

Arm reaching studies have also suggested possible differences for specific aspects of a motor command. Dizio and Lackner (1995) exposed participants to Coriolis forces by placing them in a slowly rotating room. They had subjects perform a reaching task in the dark (i.e. without visual feedback) using their dominant arm while the other arm remained stationary. When the Coriolis force was removed, the right arm, exposed to the Coriolis force, showed a curved trajectory that mirrored that of the Coriolis force. The non-exposed arm, in contrast, showed a linear trajectory, but had errors in their final position. This demonstrated a difference in end-point and trajectory adaptation and how it generalized between the limbs. Scheidt and Ghez (2007) conducted a study on differences in end-point and trajectory, and then modeled their results. Their simulation best matched their results when they added two sequential forward controllers; the first to initiate trajectory, and the second for control of final position. This difference in initial trajectory and final position may be somewhat analogous to the differences observed in the present study between utterance onset and F0 maintenance. In the contexts of the present study, we suggest that F0 maintenance uses distinct feedforward controllers from those used at utterance onset and during volitional changes in F0. Another possible analogue of the present study may be force exertion (such as pushing something), in which a fairly constant force or velocity is the goal. In such a case, differences might exist at the onset of the motor command (when the pushing movement is initiated and a desired velocity is reached), and during maintenance of the motor command.

3.3 Single trial Adaptation effects

We found that the onset trials caused a shift in the initial F0 of the following trials, even when an onset trial was preceded by a control trial, indicating some level of single-trial adaptation. This finding is consistent with the results of Donath et al. (2002) who found that compensation responses carried over within and between trials when speaking a nonsense word. Such single trial adaptation has also been found in arm reaching, with perturbations within a trial affecting subsequent trials (Thoroughman et al., 2007). The benefit of the current method is that we were able to use our analysis of the first 50 ms of the utterance to demonstrate that the single trial adaptation effects modified the forward motor plan used in the next utterance. Without an examination of the very beginning of a motor command, it is possible that any aftereffects are caused by an early within-movement compensation for feedback alterations. In other words, we have demonstrated that any differences in trials following onset trials represents alterations of the feedforward motor plans.

A particularly important question to investigate is how much adaptation occurs between trials. While the forward model is modified by a single trial event, we do not yet know whether the magnitude of this modification is comparable to changes that occur during an adaptation trial.

That is, the cumulative effect of repeated exposure has not been quantified. Even within the adaptation block, the change in initial F0 was much smaller than the feedback alteration and the final change in F0 (as measured by median F0 during the entire utterance). It is not clear if this is because our observed adaptation effects represent only a partial remapping of the forward model along with some additional compensation within each utterance, or because of physiological limitations within the system preventing a radical change in initial F0.

3.4 Conclusion

In this study, we have shown that the compensation response at utterance onset is larger than the compensation response to a perturbation mid-utterance. We suggest that our results are best explained by a change in the comparator used at utterance onset and mid-utterance. At utterance onset, an absolute comparator is used to match auditory feedback to the intended F0. After utterance onset, a goal change occurs to a stabilization mechanism, which uses the current F0 as the pitch goal (variable referent). This suggests that F0 is not universally controlled by an efference mechanism, but that different mechanisms for F0 control and stabilization can be used to suit different goals within different contexts. This finding may have implications for theories of motor control and the universality of motor control mechanisms under different goals and contexts.

4. Experimental Procedures

4.1 Participants

Fifteen female participants (aged 19 to 24) were recruited for this study. All reported that they had never received formal singing training, and were not practicing singers (e.g. in a choir). No participant spoke any tonal languages. All participants read and signed an informed consent form, in accordance with the ethical policies of Wilfrid Laurier University.

4.2 Procedure

Participants heard a target female voice vocalizing the vowel /a/ at a specific frequency (D4, 296.33 Hz) for 1 s. They were instructed to begin vocalizing for 3 s when the target voice finished, matching the pitch of the target. A 1000 Hz tone indicated when they should cease vocalizing. A loudness monitor in front of the participants allowed them to maintain a specific volume. This was done to prevent an increase in vocal intensity that is generally seen in the presence of noise (the Lombard effect), as visual feedback has been shown to inhibit the Lombard effect (Pick, Siegel, Fox, Garber & Kearney, 1989). Participants were instructed to vocalize at a volume of approximately 75 dB. Auditory feedback was amplified and heard over the headphones at approximately 85 dB.

The experiment was divided into two blocks, with a filler task between them. Each block had 140 trials and lasted approximately 18 min. In the adaptation block, participants heard their feedback unaltered for 20 trials. From trials 21 to 40, the F0 of their feedback was gradually increased by 5 cents/trial, until it had reached 100 cents at trial 40. This 100 cents upward feedback shift was maintained for the rest of the block (Adapt trials). On half of trials 61 to 140, subjects' feedback was perturbed by removing the feedback alteration (i.e. returning their feedback to its unaltered state of 0 cents, Adapt Perturb trials). This effectively lowered the participant's feedback by 100 cents. The perturbation began between 1000 ms and 1800 ms after utterance onset and lasted for 500 ms.

During the control block, participants heard their unaltered feedback for the first 60 trials. In half of the trials from trial 61 to trial 140, a shift 100 cent upwards was pseudorandomly introduced *prior* to utterance onset (Onset trials). No more than three successive Control or Onset trials occurred during the block. The Onset shift was equivalent to the shift used at onset

in the adaptation block, though it was a randomly presented (i.e. unpredictable) as opposed to the predictable shift in the adaptation block. In half of the Control and Onset trials, a perturbation was induced between 1000 ms and 1800 ms after utterance onset for 500 ms. In the Onset condition, this perturbation was induced by removing the feedback alteration, while in the Control trials the perturbation consisted of a 100 cent downward shift. Therefore four conditions occurred during the control block: a Control No-perturb condition where subjects heard unaltered feedback, a Control Perturb condition where subjects heard their feedback suddenly shifted down mid-utterance, an Onset No-perturb condition where subjects heard their voice shifted 100 cents up during their entire utterance, and an Onset Perturb condition where subjects heard their voice shifted up 100 cents from the beginning of their utterance, but this alteration was turned off briefly mid-utterance. A schematic diagram of the F0 shifts in each condition in the control and adaptation block is shown in Figure 1.

A filler task was performed between blocks. Participants were asked to read 100 sentences (taken from Kalikow et al. 1977) silently, speak the sentence in a monotone voice, and then repeat the last word. This took approximately 15–20 minutes. This filler task served to remove any carry-over effects from the adaptation block, if the adaptation block occurred before the control block. During the filler task, participants heard their unaltered feedback. Participants performed the filler task even when the control block was presented first.

4.3 Apparatus

Participants sat in a double-walled sound attenuated booth (Industrial Acoustic Company, Model 1601-01), and wore headphones (Sennheiser HD 280 Pro) and a headset microphone (Countryman E6 Omni). Vocalizations were sent from the microphone to a mixer (Mackie Oynx 1220, Loud Technologies), which passed the voice signal to a digital signal processor (DSP; VoiceOne, T.C. Hellicon). The DSP shifted the participant's voice and returned it to the mixer, where it was mixed with pink masking noise (70 dB) and returned to the participant as auditory feedback. The unaltered voice signal was digitally recorded (TASCAM HD-P2) at a sampling rate of 44.1 KHz.

4.4 Analysis

Each trial onset was manually segmented and saved into a separate WAV file. F0 for each utterance was calculated using an autocorrelation algorithm included in the Praat program (Boersma, 2001), with a sampling rate of 5 ms.

To measure adaptation effects, the F0 for the entire utterance was converted into cents using the formula:

$$\text{cents} = 100 * (39.86 * \log_{10}(\text{F0}/\text{baseline}))$$

where the baseline was 296.33 Hz (the target pitch participants were instructed to match).

The compensation response for the mid-utterance feedback alterations was also calculated. The F0 trajectories for each perturbation trial type were time aligned at the point of the perturbation and average waveforms were generated. The period between 500 ms before to 1000 ms after the perturbation was used to evaluate the compensation response. In order to compare the compensation responses across subjects, each subject's averaged waveform was converted to cents and normalized such that the baseline period (the 250 ms preceding the perturbation) had a mean of 0. The magnitude of the compensation response was determined by calculating the area under the curve using the trapezoidal rule, for 2 time periods. The first period, 100 to 250 ms post-perturbation was used to evaluate the automatic pitch-shift reflex, while the second time window, from 250–600 ms post perturbation, was used to evaluate later compensation

responses that are subject to volitional control (Hain et al., 2000). An alpha of 0.05 was used for all statistical analysis in this study.

Acknowledgments

This research was supported by the National Institute of Deafness and Communicative Disorders Grant DC-08092 and a grant from the Natural Sciences and Engineering Research Council of Canada. We thank Farina Pinnock and Danielle Culbert for their help collecting data.

References

- Boersma P. Praat, a system for doing phonetics by computer. *Glott International* 2001;5:341–345.
- Burnett TA, Freedland MB, Larson CR, Hain TC. Voice F0 responses to manipulations in pitch feedback. *J Acoust Soc Am* 1998;103:3153–3161. [PubMed: 9637026]
- Curio G, Neuloh G, Numminen J, Jousmäki V, Hari R. Speaking modifies voice-evoked activity in the human auditory cortex. *Hum Brain Mapp* 2000;9:183–191. [PubMed: 10770228]
- Dizio P, Lackner JR. Motor adaptation to Coriolis force perturbations of reaching movements: endpoint but not trajectory adaptation transfers to the nonexposed arm. *J Neurophysiol* 1995;74:1787–1792. [PubMed: 8989414]
- Donath TM, Natke U, Kalveram KT. Effects of frequency-shifted auditory feedback on voice F0 contours in syllables. *J Acoust Soc Am* 2002;111:357–366. [PubMed: 11831808]
- Elman JL. Effects of frequency-shifted feedback on the pitch of vocal productions. *J Acoust Soc Am* 1981;70:45–50. [PubMed: 7264071]
- Hain TC, Burnett TA, Kiran S, Larson CR, Singh S, Kenney MK. Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Exp Brain Res* 2000;130:133–141. [PubMed: 10672466]
- Hawco CS, Jones JA, Ferretti TR, Keough D. ERP correlates of online monitoring of auditory feedback during vocalization. *Psychophysiology*. 2009 in press
- Heinks-Maldonado TH, Mathalon DH, Gray M, Ford JM. Fine-tuning of auditory cortex during speech production. *Psychophysiology* 2005;42:180–190. [PubMed: 15787855]
- Houde JF, Jordan MI. Sensorimotor adaptation in speech production. *Science* 1998;279:1213–1216. [PubMed: 9469813]
- Houde JF, Jordan MI. Sensorimotor adaptation of speech I: Compensation and adaptation. *J Speech Lang Hear Res* 2002;45:295–310. [PubMed: 12003512]
- Houde JF, Nagarajan SS, Sekihara K, Merzenich MM. Modulation of the auditory cortex during speech: an MEG study. *J Cogn Neurosci* 2002;14:1125–1138. [PubMed: 12495520]
- Jones JA, Keough D. Auditory-motor mapping for pitch control in singers and nonsingers. *Exp Brain Res* 2008;190:279–287. [PubMed: 18592224]
- Jones JA, Munhall KG. Perceptual calibration of F0 production: evidence from feedback perturbation. *J Acoust Soc Am* 2000;108:1246–1251. [PubMed: 11008824]
- Jones JA, Munhall KG. The role of auditory feedback during phonation: Studies of Mandarin tone production. *J Phonetics* 2002;30:303–320.
- Kalikow DN, Stevens KN, Elliott LL. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J Acoust Soc Am* 1977;61:1337–1351. [PubMed: 881487]
- Kudo N, Nakagome K, Kasai K, Araki T, Fukuda M, Kato N, Iwanami A. Effects of corollary discharge on event-related potentials during selective attention task in healthy men and women. *Neurosci Res* 2004;48:59–64. [PubMed: 14687881]
- Larson CR, Burnett TA, Bauer JJ, Kiran S, Hain TC. Comparison of voice F0 responses to pitch-shift onset and offset conditions. *J Acoust Soc Am* 2001;110:2845–2848. [PubMed: 11785786]
- Larson CR. Cross-modality influences in speech motor control: the use of pitch shifting for the study of F0 control. *J Commun Disord* 1998;31:489–502. [PubMed: 9836138]
- Liu H, Larson CR. Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex. *J Acoust Soc Am* 2007;122:3671–3677. [PubMed: 18247774]

- Munhall KG, MacDonald EN, Byrne SK, Johnsrude I. Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *J Acoust Soc Am* 2009;125:384–390. [PubMed: 19173425]
- Näätänen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin Neurophysiol* 2007;118:2544–2590. [PubMed: 17931964]
- Natke U, Kalveram KT. Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables. *J Speech Lang Hear Res* 2001;44:577–584. [PubMed: 11407562]
- Nowak DA, Topka H, Timmann D, Boecker H, Hermsdörfer J. The role of the cerebellum for predictive control of grasping. *Cerebellum* 2007;6:7–17. [PubMed: 17366262]
- Perkell J, Matthies M, Lane H, Guenther F, Wilhelms-Tricarico R, Wozniak J, Guiod P. Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. *Speech Communication. Special Issue: Speech Production: Models and Data* 1997;22:227–250.
- Pick HL, Siegel GM, Fox PW, Garber SR, Kearney JK. Inhibiting the Lombard effect. *J Acoust Soc Am* 1989;85:894–900. [PubMed: 2926004]
- Purcell DW, Munhall KG. Compensation following real-time manipulation of formants in isolated vowels. *J Acoust Soc Am* 2006;119:2288–2297. [PubMed: 16642842]
- Scheidt RA, Ghez C. Separate adaptive mechanisms for controlling trajectory and final position in reaching. *J Neurophysiol* 2007;98:3600–3613. [PubMed: 17913996]
- Thoroughman KA, Fine MS, Taylor JA. Trial-by-trial motor adaptation: a window into elemental neural computation. *Prog Brain Res* 2007;165:373–382. [PubMed: 17925258]
- Wolpert DM, Kawato M. Multiple paired forward and inverse models for motor control. *Neural Netw* 1998;11:1317–1329. [PubMed: 12662752]

Figure 1.

Schematic representation of the four shifts presented in the different conditions. No mid-utterance perturbations were presented in trials 1–60 in the control and adaptation blocks. The onset trials were similar to the adaptation trials except that the onset trials were presented randomly in the same block as the control trials.

Figure 2.

A) Mean of the median F0 of each utterance across trials for the control and adapt blocks. By trial 40, when auditory feedback F0 in the adaptation block has been increased by 100 cents, subjects have reduced their F0 by approximately 70 to 80 cents to offset the feedback alteration.

B) Mean of the median F0 for the first 50 ms of vocalization, averaged over blocks of 5 trials. The F0 for utterance initiation is lower in the adaptation block, indicating that subjects adjusted their initial F0 in response to the change in feedback F0, suggesting adaptation occurred.

Subjects 11, 14, and 15 are excluded from these graphs. Note that the data for trials 60 to 140 in the control block is affected by the presence of the onset trials, which have effects on both

the mean F0 for the whole utterance, and the initial F0 of utterances following onset trials. Error bars represent 1 standard deviation.

Figure 3.

Compensation responses for the three perturbation conditions. Note that F0 in the onset condition was 50 cents below the F0 of the control condition before the onset of the mid-utterance perturbation, but the compensation to the mid-utterance perturbation in the onset condition is only approximately 17 cents. This shows that subjects had a larger response to the perturbation at trial onset than the perturbation in mid-utterance. Error bars represent 1 standard deviation. Time 0 indicated the time of perturbation onset, which was 1000 to 1800 ms after utterance onset. Error bars indicate 1 standard deviation from the mean.

Table 1

Mean (and standard deviation) F0 in cents prior to and after perturbation

	Adapt Perturb	Onset Perturb	Control Perturb
Baseline (150 ms prior to perturbation)	−64.68 (0.33)	−61.79 (0.35)	−12.15 (0.25)
Post-Shift (100 ms to 250 ms post-perturbation)	−51.28 (5.79)	−47.58 (6.44)	4.48 (7.76)