A Preliminary Study of Individual Responses to Real-Time Pitch and Formant Perturbations

Ewen N. MacDonald[1], *Kevin G. Munhall*[2]

[1]Centre for Applied Hearing Research, Technical University of Denmark, Denmark
[2]Department of Psychology, Queen's University, Canada

emcd@elektro.dtu.dk, kevin.munhall@queensu.ca

Abstract

Previous studies have demonstrated a wide range in individuals' compensations in response to real-time alterations of the auditory feedback of both pitch and formant frequencies. One potential source of this variability may be individual differences in the relative weighting of auditory and somatosensory feedback. The present study examined this variability by comparing individuals' compensations during two perturbation conditions: a pitch shift (+200 cents) and a formant shift (F1 +200 Hz, F2 -250 Hz). While no significant correlation was found between the two perturbation conditions, a modest correlation between compensations in pitch and formant frequency was observed within the pitch perturbation condition.

1. Introduction

When we talk, we monitor the sounds we produce to aid us in controlling speech production. Traditionally, this use of auditory feedback has been studied using perturbation techniques in which characteristics of the acoustic signal such as pitch [1] or formants [2] are altered in real-time. On average, talkers compensate by adjusting the acoustics of their voice in the direction opposite to that of the perturbation. However, the compensatory response has been found to vary significantly across individuals with some "following" rather than opposing the perturbation [3, 4]. For example, in a recent study that examined the compensation of 116 female talkers in response to the same formant shift (+200 Hz in F1, -250 Hz in F2), the average compensation (53 and 58 Hz in F1 and F2) was similar to the standard deviation of the compensations (44 and 69 Hz respectively) [5].

Somatosensory feedback also plays a role in speech-motor control [6]. Adapting speech to completely compensate for acoustic perturbations may result in somatosensory feedback that is incongruous. Evidence of increased compensation to pitch shifts when a local anesthetic was administered to the vocal folds supports this tradeoff between auditory and somatosensory feedback [7]. Thus, a possible explanation for the large variability in talkers' compensation may be individual differences in the relative weighting of auditory vs. somatosensory feedback in speech-motor control.

The approach taken in the present study was to investigate this relative weighting hypothesis by comparing the compensatory responses to pitch and formant shifts. A talker who relies mostly on auditory feedback should exhibit large compensations for both pitch and formant perturbations but a talker who relies mostly on somatosensory feedback should exhibit little compensation.

While pitch and formant perturbation experiments are similar conceptually, the experimental paradigms are often quite different. In a traditional pitch perturbation experiment, talkers are asked to produce sustained vowels, often with durations greater than 5 s. Over the course of a sustained vowel utterance, one or more short perturbations, with durations ranging from 100–500 ms, are randomly introduced. In contrast, in a typical formant perturbation experiment, talkers produce single-syllable utterances and when a perturbation is applied, it is applied over an entire utterance. In the present experiment, the paradigm typical of formant perturbation experiments was used for both the pitch and formant perturbation conditions. The main advantage in using this paradigm is that it allows us to examine if talkers exhibit pitch and formant compensation in the same set of utterances.

2. Method

2.1. Participants

The participants in this study were 22 undergraduate female students at Queen's University. All were native English speakers and reported no history of hearing or language disorders. All were found to have normal hearing thresholds between 500 and 4000 Hz (i.e., < 25 dB HL).

2.2. Equipment

The equipment used to conduct the real-time formant shifting was identical to that used by MacDonald et al. [8]. Testing was conducted in an Industrial Acoustics Co. (IAC) sound booth. Talkers spoke into a headset microphone (Shure WH20). Signal conditioning was performed using amplification (Tucker-Davis Technologies MA3 microphone amplifier), and low-pass filtering (cut-off frequency of 4500 Hz, Krohn-Hite 3384 filter).

For the condition where formants where shifted, the conditioned signal was digitized with a sampling rate of 10 kHz and filtered in real-time using custom software running on a National Instruments PXI-8106 controller. Formants were estimated every 900 µs using an iterative Burg algorithm with a model order that varied from 8 to 12 across individuals. IIR filter coefficients were computed based on these estimates such that a pair of spectral zeroes was placed at the location of the existing formant frequency and a pair of spectral poles was placed at the desired frequency of the new formant.

For the condition where the pitch was shifted, the conditioned signal was processed using an Eventide Harmonizer H3000 employing a proprietary algorithm.

The processed output was amplified and mixed with noise (Madsen Midimate 622 audiometer) and presented over headphones (Sennheiser HD 265) such that the speech and noise were presented at approximately 80 and 50 dBA, respectively.

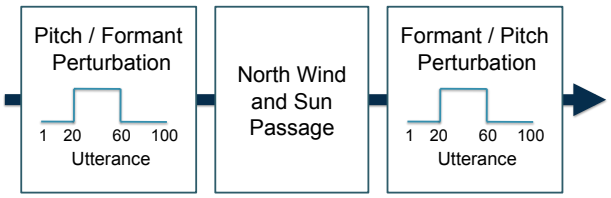

Figure 1: *Schematic of the procedure.*

2.3. Procedure

After collecting pure-tone hearing thresholds, talkers produced six utterances of seven English vowels in an /hVd/ context ("heed," "hid," "hayed," "head." "had," "hawed," and "who'd"). Talkers were instructed to say words that appeared on a computer monitor at a natural rate and speaking level. Each word prompt lasted 2.5 s and the inter-trial interval was approximately 1.5 s. These utterances were analyzed to select the best model order for each individual, using a heuristic based on minimum variance in formant frequency over a 25 ms segment midway through the vowel.

Each talker participated in a Pitch Perturbation condition and a Formant Perturbation condition. The order in which talkers completed the conditions was counterbalanced. Between the conditions, talkers read aloud "The North Wind and the Sun" passage [9]. An overall schematic of the experiment is illustrated in Figure 1.

In each of the perturbation conditions, talkers produced a total of 100 utterances of the word "head." For the first 20 utterances, the Baseline phase, talkers received normal auditory feedback (i.e., amplified and mixed with noise, but with no pitch or formant shift). For utterances 21–60, the Shift phase, talkers received altered auditory feedback. In the Pitch Perturbation condition, the auditory feedback was increased by 200 cents. In the Formant Perturbation condition, F1 was increased by 200 Hz and F2 was decreased by 250 Hz. For utterances 61–100, the Return phase, auditory feedback was returned to normal.

The procedure used for offline analysis was similar to that used by MacDonald et al. [8]. The boundaries of the vowel segment in each utterance were estimated using an automated process based on the harmonicity of the power spectrum. These boundaries were then inspected by hand and corrected, if required.

For each vowel segment, F0 estimates were calculated using Praat software (www.praat.org). A single "steady-state" value was calculated from the median of the estimates from 40% to 80% of the way through the vowel.

The first three formant frequencies were estimated offline from the first 25 ms of a vowel segment with the same algorithm used in the online shifting. The formants were estimated again after shifting the window 1 ms and repeated until the end of the vowel segment was reached. For each vowel segment, a single steady-state value for each formant was calculated by averaging the estimates for that formant from 40% to 80% of the way through the vowel. While using the best model order reduced gross errors in formant tracking, occasionally one of the formants was incorrectly categorized as another (e.g., F2 being misinterpreted as F1, etc.). These incorrectly categorized estimates were found and corrected by examining a plot with all of the steady-state F1, F2, and F3 estimates for each individual.

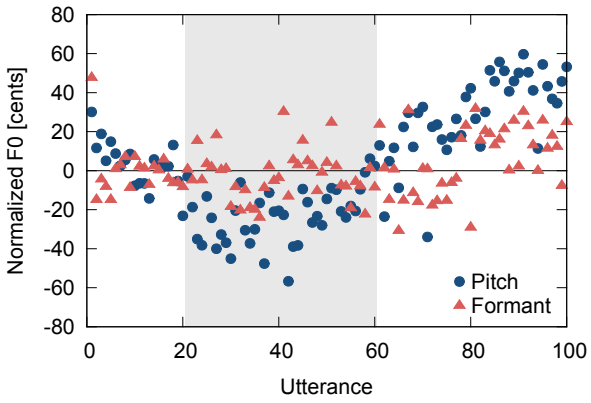

Figure 2: *Average F0 for each utterance in the Pitch (circles) and Formant (triangles) Perturbation conditions. Individuals' F0 results were converted to cents using the average of the last 15 utterances of the Baseline phase as a reference for each individual. The shaded region indicates when the feedback was perturbed.*

3. Results

3.1. Pitch Compensation

Using the F0 results from the Pitch Perturbation condition, a baseline average F0 was calculated for each individual from the last 15 utterances of the Baseline phase (i.e., utterances 6–20). Using their baseline average, individuals' F0 results were then normalized by converting from Hz to cents. The F0 results from the Formant Perturbation condition were analyzed in a similar manner. The normalized results for each utterance, averaged across talkers, can be seen in Figure 2.

From Figure 2, it appears that talkers did not alter F0 during the Formant Perturbation condition, but did alter F0 during the Shift phase of the Pitch Perturbation condition. To quantify this, the compensation of each talker was calculated. Here, the magnitude of compensation was defined as the average normalized F0 (in cents) of the last 15 utterances of the Shift phase. The sign of the compensation was defined as positive if it opposed the perturbation and negative if it followed the perturbation. For the Pitch Perturbation condition, a single sample t-test of talkers' compensations was not significantly different from 0 [$t(21) = 0.945$, $p > 0.35$]. A closer examination of individual results revealed a wide range of compensation. While 15 talkers compensated (i.e., altered production in the direction opposite the perturbation), 7 of the talkers followed (i.e., altered production in the same direction as the perturbation). Thus, the lack of statistical significance is likely due to the mix of compensators and followers.

3.2. Formant Compensation

Formant compensations were examined in two contexts: the response to a direct formant perturbation (Formant Perturbation session) and the response to an inadvertent shift of the formant when the pitch is shifted with an effects processor (Pitch Perturbation session).

From the formant results from the Formant Perturbation session, a baseline average for F1 and F2 was calculated for each individual from the last 15 utterances of the Baseline phase (i.e., utterances 6–20). Each individual's F1 and F2 results were

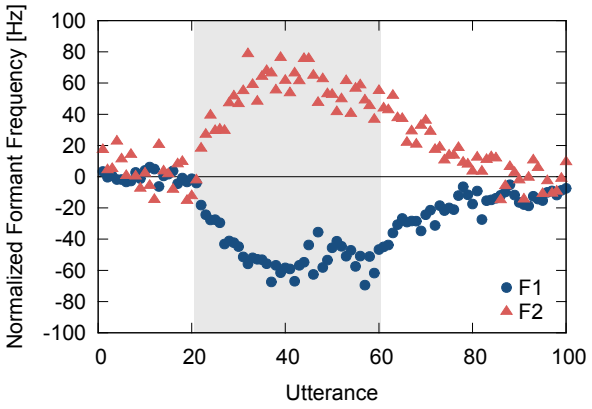

Figure 3: *Average normalized F1 (circles) and F2 (triangles) frequencies for each utterance in the Formant Perturbation condition. The shaded region indicates when the feedback was perturbed.*

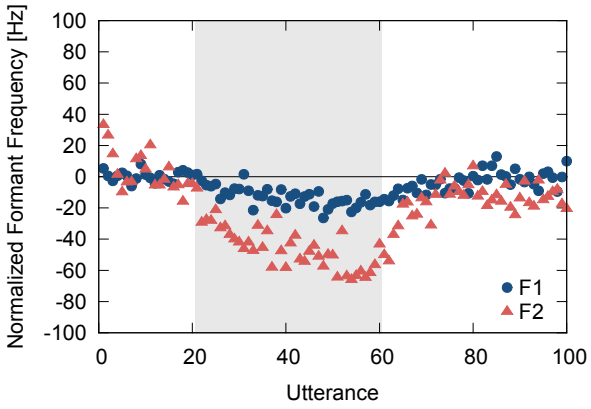

Figure 4: *Average normalized F1 (circles) and F2 (triangles) frequencies for each utterance in the Pitch Perturbation condition. The shaded region indicates when the feedback was perturbed.*

then normalized by subtracting that individual's baseline average. The normalized results for each utterance, averaged across talkers, can be seen in Figure 3.

The pitch shifting algorithm used by the effects processor shifts the entire spectrum. The result is that, along with the pitch, the formant frequencies are also perturbed. Over the last 15 utterances of the Baseline phase of the Pitch Perturbation condition, the average formant frequency produced by talkers was 736.5 and 2048.9 Hz for F1 and F2 respectively. During the Shift phase, the auditory feedback was increased by 200 cents. Thus, on average, the frequencies of F1 and F2 were shifted by 90.2 and 251.0 Hz respectively. For F2, this resulted in a formant shift that was almost identical in magnitude, but opposite in direction, to that applied in the Formant Perturbation condition.

To examine if talkers altered their formants in the Pitch Perturbation condition, a similar normalization process was conducted on the formant results. Again, a baseline average for F1 and F2 was calculated for each individual from the last 15 utterances of the Baseline phase and used to normalize each individual's F1 and F2 results. The normalized results for each utterance, averaged across talkers, can be seen in Figure 4.

From Figures 3 and 4, it is clear that, on average, talkers altered formant production during the Shift phase of both perturbation conditions. To confirm this, repeated measures ANOVAs were conducted with phase (the average formant frequency of the last 15 utterances in the Baseline vs. Shift) as within- and order of the perturbation conditions as between-subject factors. For the results from the Formant Perturbation condition, a significant effect of phase was found for both F1 [$F(1, 20) = 35.566$, $p < 0.001$] and F2 [$F(1, 20) = 15.67$, $p = 0.001$] but no significant effect of order was found for either F1 [$F(1, 20) = 3.77$, $p = 0.07$] or F2 [$F(1, 20) = 1.624$, $p = 0.22$]. Similarly, for the results from the Pitch Perturbation condition, a significant effect of phase was found for both F1 [$F(1, 20) = 9.643$, $p = 0.006$] and F2 [$F(1, 20) = 26.881$, $p < 0.001$] but no significant effect of order was found for either F1 [$F(1, 20) = 2.233$, $p = 0.15$] or F2 [$F(1, 20) = 1.508$, $p = 0.23$].

For each perturbation condition, the compensation in F1 and F2 of each talker was calculated. Here, the magnitude of compensation was defined as the difference between the average formant frequencies of last 15 utterances of the Baseline and Shift phases. Again, the sign of the compensation was defined as positive if it opposed the perturbation and negative if it followed the perturbation. While most compensated, a few talkers followed rather than opposed the formant perturbation; one talker followed in F1, and two in F2, but no talker followed in both F1 and F2.

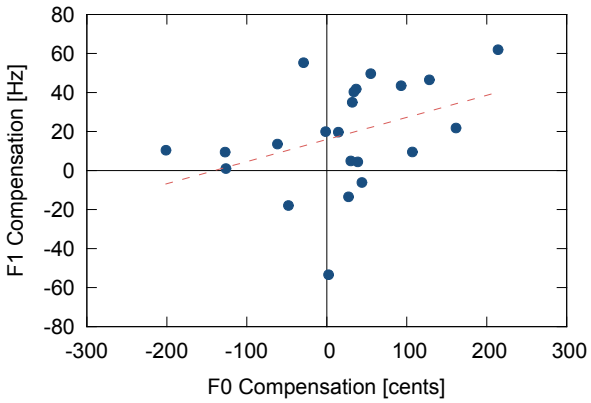

Figure 5: *Scatterplot of talkers' compensation in F0 and F1 in the Pitch Perturbation condition.*

3.3. Comparison of Pitch and Formant Compensations

Talkers' compensations to formant and pitch perturbations in the different conditions were compared and correlations were computed. A modest correlation was observed between F0 and F1 compensations within the Pitch Perturbation condition [$r(22) = 0.392$, $p = 0.04$, one-tailed; see Fig. 5] but not between F0 compensation in the Perturbation condition and F1 compensation in the Formant Perturbation condition [$r(22) = -0.124$, $p > 0.29$, one-tailed].

A trend for a modest correlation was observed between F1 compensations between the Pitch and Formant Perturbation conditions [$r(22) = 0.300$, $p = 0.09$, one-tailed] but not for F2 compensations between the Pitch and Formant Perturbation conditions [$r(22) = -0.230$, $p = 0.15$, one-tailed].

Further, no significant correlation was observed between compensations in F1 and F2 within the Formant Perturbation condition [$r(22) = 0.159$, $p = 0.24$, one-tailed] or between F1 and F2 within the Pitch Perturbation condition [$r(22) = 0.082$, $p = 0.36$, one-tailed].

4. Discussion

In this study, talkers repeatedly produced utterances of the word "head" while receiving auditory feedback in which the pitch or formant frequencies had been perturbed in real-time. As in previous studies, partial compensations to both pitch and formant perturbations were observed and the magnitude of compensation varied widely across talkers.

One potential explanation for the large variability in compensations observed across talkers may be individual differences in the relative weighting of auditory vs. somatosensory feedback in speech-motor control. While no significant correlation was found between pitch compensations in the Pitch Perturbation condition and formant compensations in the Formant Perturbation condition, a modest correlation was observed between compensations in pitch and formant frequencies within the Pitch Perturbation condition. The observation of a modest correlation supports this relative weighting hypothesis. However, the lack of significant correlation across perturbation conditions suggests that other sources of variability are also involved.

The paradigm used in this experiment is different from that used in a typical pitch perturbation study. In the present study, pitch shifts were applied to an entire utterance rather than being restricted to a short interval midway through the vowel. This provides two advantages. First, it allowed talkers to speak normally rather than prolonging their vowels. Second, it allowed us to examine pitch and formant compensations within the same set of utterances.

The pitch shifting algorithm employed by the effects processor used in this study shifted the entire spectrum of the input signal. Thus, both pitch and formant frequencies were affected. The magnitude of the pitch perturbation used in the present study was 200 cents. This value was chosen because, for the word and talkers used in the study, F2 would be shifted by similar amounts in both the Pitch and Formant Perturbation conditions. On average, in both conditions, the talkers compensated equally. However, individual compensations in F2 were not correlated across conditions. While the formant shift of F1 was smaller in the Pitch Perturbation condition, talkers still compensated, and a trend for modest correlation between individuals' F1 compensations in the Pitch and Formant Perturbation conditions was observed. Thus, while the order of perturbation conditions was not found to have an effect on overall compensation, individuals' responses to formant perturbations varied between conditions. This variability may suggest that the compensatory response may not be as stable as previously thought.

The 200 cent pitch perturbation used in the present study is, in general, larger than most studies of pitch perturbation. Previous perturbation studies have observed that the percentage of compensation (i.e., the compensation divided by the magnitude of the perturbation) decreases for large perturbations [10]. Similarly, formant compensation has been found to be approximately linear for small perturbations, but non-linear, and proportionally smaller, for large perturbations [8, 11]. Thus, the magnitudes of the perturbations used in the present experiment may have resulted in a more linear response to the formant than the pitch perturbation.

Studies that have more closely examined the time course of adaptations to pitch perturbations have identified both volitional and reflexive components of the reponse [12]. In the present study, the measurement of pitch compensation did not explore the effects of these individual components. In contrast, a voluntary component of the response to formant perturbation has not been found [4]. Thus, there are some differences in mechanisms used for speech-motor control of pitch and formants. Future studies that can isolate the components of the pitch response and compare them to the formant response may better test the relative-weighting hypothesis.

Identifying the source of individual differences in compensatory responses remains a difficult task. The results of the present study suggest that the comparison of individual compensations to perturbations of different acoustical characteristics of speech is a promising method to explore this problem.

5. Acknowledgements

This research was supported by the (US) National Institute of Deafness and Communicative Disorders Grant DC-08092

6. References

[1] Kawahara H, "Hearing voice: Transformed auditory feedback effects on voice pitch control," in *Proceedings of the International Joint Conference on Artificial Intelligence: Workshop on Computational Auditory Scene Analysis*, Montreal, Canada, 1995, pp. 143–148.

[2] J. F. Houde and M. I. Jordan, "Sensorimotor adaptation in speech production." *Science*, vol. 279, no. 5354, pp. 1213–1216, 1998.

[3] T. A. Burnett, M. B. Freedland, C. R. Larson, and T. C. Hain, "Voice F0 responses to manipulations in pitch feedback," *Journal of the Acoustical Society of America*, vol. 103, no. 6, pp. 3153–3161, 1998.

[4] K. G. Munhall, E. N. MacDonald, S. K. Byrne, and I. Johnsrude, "Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate." *The Journal of the Acoustical Society of America*, vol. 125, no. 1, pp. 384–90, Jan. 2009.

[5] E. N. MacDonald, D. W. Purcell, and K. G. Munhall, "Probing the independence of formant control using altered auditory feedback." *The Journal of the Acoustical Society of America*, vol. 129, no. 2, pp. 955–65, Feb. 2011.

[6] S. Tremblay, D. M. Shiller, and D. J. Ostry, "Somatosensory basis of speech production." *Nature*, vol. 423, no. 6942, pp. 866–869, 2003.

[7] C. R. Larson, K. W. Altman, H. Liu, and T. C. Hain, "Interactions between auditory and somatosensory feedback for voice F0 control." *Experimental Brain Research*, vol. 187, no. 4, pp. 613–621, 2008.

[8] E. N. MacDonald, R. Goldberg, and K. G. Munhall, "Compensations in response to real-time formant perturbations of different magnitudes." *The Journal of the Acoustical Society of America*, vol. 127, no. 2, pp. 1059–68, Feb. 2010.

[9] IPA, *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press, 1999.

[10] H. Liu and C. R. Larson, "Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex." *Journal of the Acoustical Society of America*, vol. 122, no. 6, pp. 3671–3677, 2007.

[11] S. Katseff, J. Houde, and K. Johnson, "Partial Compensation for Altered Auditory Feedback: A Tradeoff with Somatosensory Feedback?" *Language and Speech*, in Press. [Online]. Available: http://dx.doi.org/10.1177/0023830911417802

[12] T. A. Burnett, K. E. McCurdy, and J. C. Bright, "Reflexive and volitional voice fundamental frequency responses to an anticipated feedback pitch error." *Experimental Brain Research*, vol. 191, no. 3, pp. 341–51, Nov. 2008.