Compensations of F0 and formant frequencies in a real-time pitch-perturbation paradigm

Andreas Eckey[1], Ewen N. MacDonald[2]

[1] Institut für Hörtechnik und Audiologie, Jade Hochschule Oldenburg, Email: an.ec@icloud.com

[2] Hearing Systems Group, Technical University of Denmark, Email: emcd@elektro.dtu.dk

Abstract

While producing speech, talkers monitor both somatosensory and auditory feedback. Many studies have demonstrated that if auditory feedback is manipulated in real-time (e.g., using an effects processor to shift the frequency spectrum), subjects compensate by modifying their F0 in the direction opposite to the perturbation. However, shifting the entire frequency spectrum alters both F0 and formant frequencies. While compensations for real-time formant perturbations have been previously observed, these studies have used a paradigm that is very different from that of traditional pitch-perturbation experiments. In the present study, compensations in both F0 and formant frequencies were observed for perturbations of sustained vowels using a traditional pitch-perturbation paradigm. Within a sustained utterance, the auditory feedback was increased or decreased by 100 cents for a short duration. Previous studies have suggested that the large variability in compensation across individuals may be due to individual differences in weighting somatosensory and auditory feedback. Following this hypothesis, individuals' compensations in F0 and formant frequency should be correlated. However, only a weak correlation was observed.

Introduction

Previous studies have demonstrated that talkers use auditory feedback to monitor and control the pitch of their voice [1, 2]. In these studies, an effects processor was used to shift the entire frequency spectrum for a short period during a prolonged utterance of a vowel. As the entire spectrum was shifted, in addition to F0, the formant frequencies were also altered. Previous work has also demonstrated that talkers use auditory feedback to control formant frequencies [3, 4, 5]. However, in these studies, only formant frequencies were manipulated. Thus, when using an effects processor to alter auditory feedback, one would expect talkers to compensate in both fundamental and formant frequencies. While formant compensations in response to pitch shifting were observed in [6], these were elicited using a continuously shifted feedback and, thus, investigated the adaptive feedforward control system. In contrast, typical pitch shifting studies employ short, unpredictable manipulations to investigate the feedback control system.

A large range of individual differences are often observed in altered auditory feedback experiments [1, 5]. A source of this variability may be individual differences in the gains applied to auditory vs. somatosensory feedback [7]. If this hypothesis is true, then one would predict that individual compensations in F0 and formant frequencies should be correlated. Thus, talkers that rely on somatosensory feedback should not compensate in either F0 or formant frequencies, whereas talkers that rely on auditory feedback should exhibit large pitch and formant compensations.

There is a tendency across all languages that high vowels (e.g., /i/ and /u/) have higher F0 than low vowels (e.g., /a/) [8]. This difference in intrinsic F0 across vowels implies that the average tension on the vocal folds varies across vowels. This may influence the salience of somatosensory feedback and, as a consequence, modulate compensation.

In the present study, talkers produced prolonged utterances of three different vowels. During each utterance, a 100 cent pitch shift with random direction and onset was applied for 500 ms. This experiment was conducted for three main reasons. First, to confirm that when the entire spectrum is shifted, talkers compensate for the altered formant feedback. Second, to explore whether compensations in pitch and formant frequencies are correlated. Third, to explore whether pitch compensations vary across vowels.

Methods

The participants in this study were 9 young adults (w: 3, m: 6) with a mean age of 27 years. All talkers were native German speakers and reported no history of hearing impairment or cognitive deficit. Further, no talker reported having had significant singing training.

All recordings were conducted in a sound treated room at the Technical University of Denmark. Talkers were recorded using a Shure WH20TQG Dynamic Headset and an RME Fireface UCX audio interface. Pitch shifting was conducted using an Eventide Eclipse effects processor. The speech signal was mixed with pink noise and presented diotically via Sennheiser HDA 200 headphones. The pink noise was presented at 70 dB SPL. The gain on the microphone was adjusted such that when a talker produced the sustained vowel /a/, it was presented at 75 dB SPL.

Participants were prompted to produce the vowel that appeared on a computer screen. Three different vowels were tested: /a/, /e/, /i/. When producing the vowel, talkers were asked to sustain the production of a vowel as long as the visual prompt remained on the screen (ap-

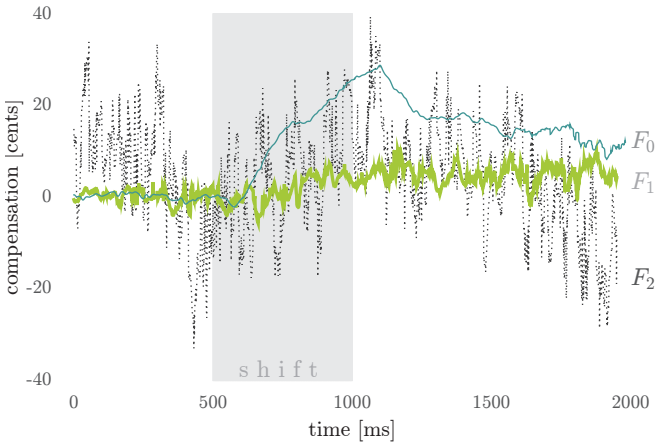

Figure 1: Average compensation in response to a 100 cent pitch shift applied for 500 ms (grey area). The results for F0 (thin), F1 (dotted), and F2 (thick) have been averaged across talkers, vowels and shift direction.

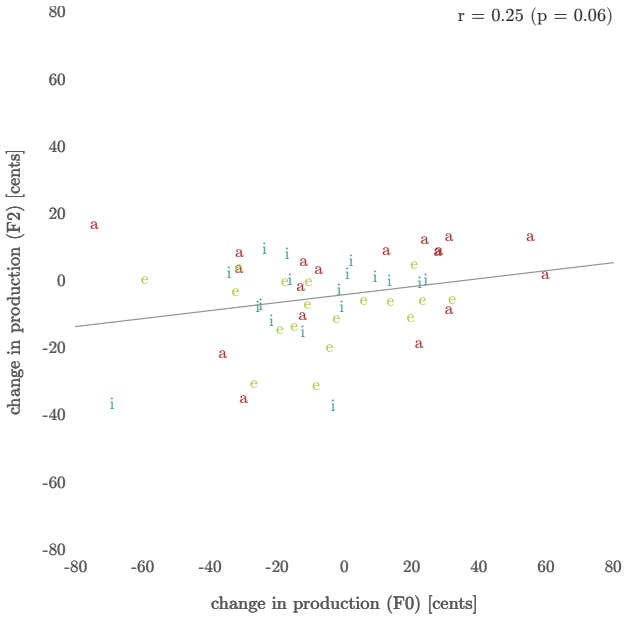

Figure 2: Scatterplot of changes in production of F0 vs. F2 for prolonged utterances of vowels /i/, /e/, and /a/.

proximately 5 s). Two training trials were run to ensure talkers were familiar with the procedure.

Within each utterance, a pitch shift was introduced at a random onset and sustained for 500 ms. Feedback was returned to normal following the shift. The magnitude of the pitch shift was always 100 cents. However, the direction of the pitch shift (either up or down) was randomized across utterances. For each vowel, a total of 30 utterances were collected (15 up and 15 down).

Results

For each vowel, utterances were trimmed and time-aligned such that the perturbation onset occurred at 1000 ms. A handful of recordings had to be discarded due to the fact that the perturbation was outside or too near the end of the utterance.

Fundamental and formant frequencies were estimated using Praat [9] and converted from Hz to cents using Equation (1). Here, f_{base} was the average frequency during the 200 ms interval prior to the perturbation onset.

$$cents = 1200 \cdot log_2(\frac{f}{f_{base}}) \qquad (1)$$

For each shift condition and vowel, the change in production of F0, F1, and F2, were measured based on the individual's average production over the interval from 500–700 ms after the pitch perturbation onset (i.e., the first 200 ms immediately after the perturbation was removed).

To verify that talkers changed production in response to the altered feedback, the changes in production in response to an upward vs. downward shift were compared. For each of F0, F1, and F2, a repeated measures ANOVA was conducted with shift direction and vowel as within-subject factors. For F0, a significant main effect of direction $[F(1,8) = 22.549,\ p = 0.001]$ and interaction between direction and vowel $[F(2,16) = 4.143,\ p = 0.035]$ were observed. As well, there was a trend for a main effect of vowel $[F(2,16) = 3.401,\ p = 0.059]$. For F1, no significant main effects of direction $[F(1,8) = 0.574,\ p = 0.47]$, vowel $[F(2,16) = 0.319,\ p = 0.73]$, or interaction $[F(2,16) = 0.706,\ p = 0.508]$ were observed. For F2, a significant main effect of direction $[F(1,8) = 31.321,\ p = 0.001]$ was observed but neither the main effect of vowel $[F(2,16) = 2.028,\ p = 0.164]$ nor interaction $[F(2,16) = 0.419,\ p = 0.665]$ was significant.

The compensation results averaged across the three vowels and two shift directions can be seen in Figure 1. Here, a positive compensation is defined as a change in production in the direction opposite to the perturbation. On average, talkers compensated in F0, F1, and F2. However, only the changes in F0 and F2 were statistically significant.

A scatterplot of individuals' average change in production of F0 vs. F2 for each of the three vowels can be see in Figure 2. Overall, a trend for a weak correlation was observed $[R = 0.254,\ p = 0.064]$.

The F0 results for each vowel, averaged across all talkers and utterances, are plotted in Figure 3. From the figure, it is clear that talkers compensated for the pitch perturbation when producing all three vowels. While the magnitude of compensation was similar across vowels when the pitch was shifted up, the compensation varied across vowels when the pitch was shifted down. For the vowel /a/, the compensation for both shift directions was similar in magnitude. However, the for the vowels /e/ and /i/, the magnitude of compensation was much smaller in response to the downward shift.

A repeated measures ANOVA with shift direction and vowel as within-subject factors was conducted on the average compensations in F0. A statistically significant main effect of vowel $[F(2,16) = 4.143,\ p = 0.035]$ and a

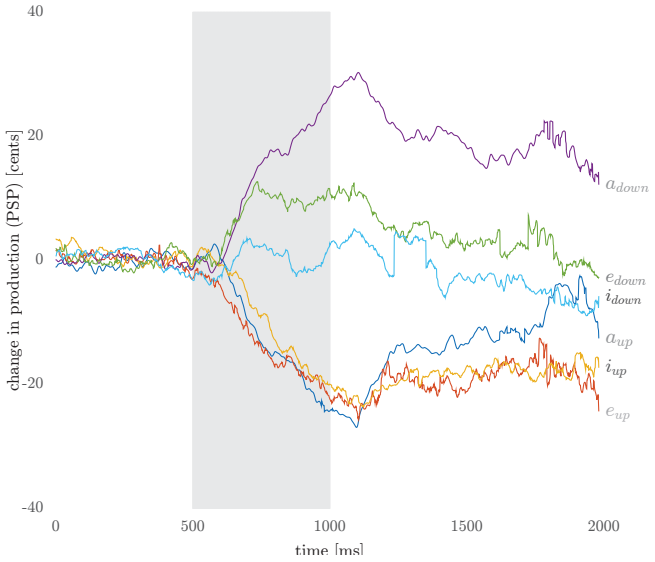

Figure 3: Average F0 response to upward and downward shifts in pitch feedback during the production of prolonged vowels /a/, /e/, /i/. The grey area indicates the interval during which the 100 cent pitch perturbation was present.

trend for an interaction between shift direction and vowel [$F(2, 16) = 3.401$, $p = 0.059$] was observed. For both shift directions, a separate repeated-measures ANOVA with vowel as a within-subject factor was then conducted. No significant difference across vowels was observed for the upward shift [$F(2, 16) = 0.023$, $p = 0.977$]. However, a significant difference was observed for the downward shift [$F(2, 16) = 5.509$, $p = 0.015$].

Discussion

In the present study, talkers received altered auditory feedback in which the entire spectrum was shifted by 100 cents for a period of 500 ms during a prolonged utterance of three different vowels. The direction of the shift (either upwards or downwards) and the onset were unpredictable. While previous work has reported compensations in F0 [1, 2], in the present study, statistically significant changes in the frequency of F2 were also observed.

To our knowledge, this is the first study to report formant compensations in response to short, unpredictable pitch shifts. A previous study [6] reported formant compensation in response to pitch perturbations. However, in that study, the pitch shift was applied continuously over the course of many regular utterances. Thus, in that study, the compensations were related to adaptive feedforward control rather than feedback control observed in the present study.

In [6], statistically significant compensations in response to a 200 cent increase in pitch were observed in both F1 and F2. However, in the present study, only compensation in F2 was significant. There are two possible reasons for this difference. First, in the present study, the magnitude of the pitch shift was 100 cents rather than the 200 cents used in the previous study. For the talkers in the present study, a 100 cent pitch shift resulted in a shift in F1 ranging from approximately 15 Hz for /i/ to 43 Hz for /a/. These shifts may have been too small to elicit a compensatory response in F1 as they are much smaller than the 100 Hz or larger shifts that have been applied in typical formant perturbation studies [4, 5, 6].

Both pitch and formant perturbation studies have observed a large variability in the degree of compensation exhibited across talkers [1, 3, 4, 5]. A potential explanation for this is that the relative weighting of auditory vs. somatosensory feedback used in speech motor control varies across individuals [7]. Based on this individual weighting hypothesis, one would expect that talkers that rely on auditory rather than somatosensory feedback should exhibit large compensations in both F0 and formant frequencies. In the present study, a weak correlation ($R = 0.25$) was observed between compensation in F0 and F2. In [6], a modest correlation ($R = 0.39$) was observed in compensation in F0 and F1. While these correlations offer some support for the individual weighting hypothesis, they are relatively weak. Thus, there are likely other factors that contribute to the large variability across talkers.

With regards to vowel production, recent work has demonstrated that compensation is influenced by the relationship of the produced vowel to the vowel space of the talker [10, 11]. Thus, the adaptive feedforward system (and presumably the feedback control system as well) involves vowel targets where the error is processed in a multidimensional perceptual space. While they may share the same native language, the average acoustic characteristics of each vowel can vary substantially across talkers. Further, there may be local perceptual differences across individuals in the multidimensional control space. Together, these differences may also contribute to the variability in compensation observed in perturbation studies.

Talkers produced three different vowels In the present study and compensation in F0 was observed in each case. For the upward shift condition, average compensation was similar in magnitude across vowels. However, for the downward shift condition, a difference in compensation was observed across the three vowels. A potential explanation for this pattern of results could be that the difference in intrinsic F0 across the three vowels results in differences in average vocal fold tension, and as a consequence, a difference in the salience of the somatosensory feedback. However, it is not clear that this can explain the asymmetrical results (i.e., the difference in compensation was observed across vowels only when the shift is applied in one direction). A second potential explanation is related to the vowel targets used in speech motor control. Since these targets are controlled in a multidimensional space, the salience of different acoustic characteristics might vary across different vowels.

The majority of formant perturbation studies have investigated how auditory feedback is used by the adaptive feedforward speech motor control system. The results of the present study demonstrate that talkers also con-

trol formant frequencies using a feedback control system. The relationship between these two control systems is intriguing and further experiments should be conducted to investigate the commonalities and differences of these feedback and adaptive feedforward control systems.

References

[1] Burnett, T.A., Freedland, M.B., Larson, C.R., and Hain, T.C.: Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America* (1998) 103, 3153–3161

[2] Larson, C.R., Altman, K.W., Liu, H. and Hain, T.C.: Interactions between auditory and somatosensory feedback for voice F0 control. *Experimental Brain Research* (2008) 187(4):613–621.

[3] Houde, J.F. and Jordan, M.I.: Sensorimotor Adaptation in Speech Production. *Science* (1998) 279, 1213–1216.

[4] Purcell, D.W. and Munhall, K.G.: Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America* (2006) 119, 2288–2297.

[5] MacDonald, E.N., Purcell, D.W., and Munhall, K.G.: Probing the independence of formant control using altered auditory feedback. *The Journal of the Acoustical Society of America* (2011) 129, 955–965.

[6] MacDonald, E.N. and Munhall, K.G.: A Preliminary Study of Individual Responses to Real-Time Pitch and Formant Perturbations. *Proceedings of The Listening Talker Workshop* (2012), Edinburgh, Scotland, 32–35.

[7] Lametti, D.R., Nasir, S.M., and Ostry, D.J.: Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *The Journal of Neuroscience* (2012) 32, 27, 9351–9358

[8] Whalen, D.H. & Levitt, Andrea G.: The universality of intrinsic F0 of vowels. *Journal of Phonetics* (1995) 23, 349-366.

[9] Pratt, URL: `http://www.praat.org`

[10] Mitsuya, T., MacDonald, E.N., Purcell, D.W., and Munhall, K.G.: A cross-language study of compensation in response to real-time formant perturbation. *The Journal of the Acoustical Society of America* (2011) 130, 2978–2986.

[11] Mitsuya, T., Samson, F., Menard, L., and Munhall, K.G.: Language dependent vowel representation in speech production *The Journal of the Acoustical Society of America* (2013) 133, 2993–3003.