Speech production in response to multiple perturbations of auditory feedback

Jinyu Li[1], Leonardo Lancia[1]
[1]Laboratoire de Phonétique et Phonologie (CNRS/Sorbonne-Nouvelle, Paris, France).

Motivation Previous research showed that speakers responded rapidly to unpredictable perturbations in the f0 of their auditory feedback. Compensatory responses were observed when speakers produced sustained vowels [2], one-syllable words [7], sentences [3], and in the presence of delayed auditory feedback (DAF) [6]. Speakers monitor their production by predicting constantly the results of their motor commands [8]. They compensate for perturbations of their auditory feedback because they consider the discrepancy between the intended and the perceived sounds as due to production errors [8]. However, some speakers sometimes follow the direction of the f0 shift instead of compensating for it. When this happens, the difference between the intended and the perceived values of f0 even increases. However, this happens more when the f0 shift is predictable [1], or when the shift range is larger [2]. Moreover, it has been shown that the direction of the speakers' response is related to fluctuations of the production system's behavior [4] and to the application of the perturbation across all trials of the experiment or only to some trials [5]. In sum, auditory feedback has a great impact on the stability of speech production and the processing of auditory feedback is condition dependent. In order to better understand the processing of auditory feedback system during speech production, this study investigates the interactive effects of the DAF, of the consistent f0 shift of auditory feedback, and of the complexity of the syllables composing the produced utterances on speech production.

Methodology The experiment started with a training session. 20 French female native speakers were asked to read a short French text with a DAF of 120ms. The speakers' manipulated voices were played back in real time through earphones. During the training session the volume of the auditory feedback coming through the earphones was adjusted in order to minimize the speaker's perception of their non-manipulated voice [2]. The speakers were then asked to repeat several times three different French sentences of five syllables in a different random order without interruptions. The sentences differed in syllabic complexity (being composed by mainly CV, CVC or CCV syllables), but had the same prosodic structure. The experimental trials (each consisting in one uninterrupted repetition of the three sentences) were organized in blocs of six. The trials composing the first bloc were considered as baseline trials, because no alteration of auditory feedback was present. During all trials of each following bloc, the f0 of the auditory feedback signal was consistently shifted by zero, one or two semitones. Each degree of f0 perturbation was applied to trials of five blocs randomly chosen. 10 speakers were exposed only to positive f0 shifts (group1) and the other 10 speakers were exposed to negative f0 shifts (group2). During each trial, the value of DAF was randomly chosen among 0, 60 and 120 ms. Therefore, each DAF value was applied two times in each bloc. All types of perturbation were constantly present from the onset to the offset of the trial.

Analysis and results The analysis of the speakers' responses to auditory feedback perturbations focused on the accented vowels of the three intonational phrases composing each sentence, because these vowels could be more sensitive to the effect of DAF or f0 shift (see [6,7]). Statistical analyses were conducted separately on the two groups of speakers by means of linear mixed models with maximal random effects structures. We first compared the f0 of the accentual vowels obtained in the baseline with that obtained with no DAF during the other blocks separately for speakers of the two groups. We tested the interaction between the effect of the f0 perturbation, that of the vowel duration and that of the syllabic complexity of the sentences on the f0 of the accentual vowels. In both groups we obtained a deviation from the baseline f0 going in the direction of the perturbation. This means that in average the speakers trended to follow the f0 shift. In order to further investigate the responses of the speakers who followed the f0 shift to DAF, we ran speaker specific linear regressions with the same model specification and then removed speakers for which the effect of the f0 shift went in the opposite direction with respect to the group (2 speakers from group1 and 3 speakers from group2). 6 speakers from group1 and 5 speakers from group2 significantly shifted their f0 in the direction of the perturbation. We then tested the effects of DAF, f0 shift, syllabic complexity and that of their triple interaction on accentual vowels' duration, on their average f0 and on the variability of spectral change patterns during the production of these vowels (as captured by the mean absolute deviation of the average squared change of Mel-Frequency Cepstral Coefficients trajectories). The lengthening of syllables is the most important effect of DAF on speakers' production (e.g. [6]). As expected, the degree of DAF was positively correlated with the duration of the accentual vowels for the two groups. The variability of the spectral change patterns can be an index of the stability of speakers' production. We found that a small delay in auditory feedback (60ms) reduced

significantly the variability of the spectral change patterns. On the other hand, this variability increased with higher syllabic complexity. The impact of syllabic complexity on spectral variability was weakened by positive f0 shift (group I) but it was increased when the f0 shift was increased by 2 semitones (group II). Syllabic complexity tended to decrease the f0 of the accentual vowels, which was significantly lower in the sentences mainly containing CVC (for group2) and CCV (for group1 and 2) syllables. However, for group2, this effect was reduced when DAF increased.

Discussion Contrary to most previous studies reporting that speakers produce compensatory motor commands in response to unexpected perturbations of auditory feedback, in this study, most subjects trended to follow the f0 shift perturbation. Therefore, when the auditory feedback was continuously perturbed in the same direction during connected speech production (sentence reading), speakers tended to adapt their motor commands to the auditory feedback in order to stabilize their f0. One the one hand, this may be partly due the consistency of the f0 perturbation in this study. On the other hand, this may also be due to the fact that the production of polysyllabic sentences demands a more dynamical control of f0 comparing to the single vowel or syllable production. Il was also shown that the f0 control was affected by the DAF and by the syllabic complexity of the utterances. The reduction of the variability in the production of segmental material due to DAF (as revealed by the analysis of spectral change patterns) could be explained as due to the lengthening of accentual vowels. To some extent, the effect of syllabic complexity on the stability of production is predictable. However, the observed effect of the f0 shift on spectral variability and its dependency on the direction of the f0 perturbation require further investigation.

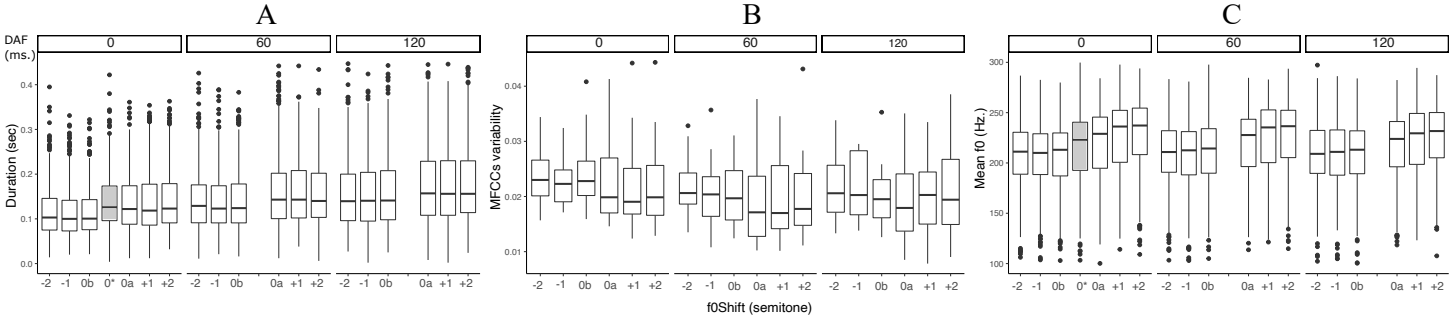

Figure 1 Vowel duration (A), Mel-Frequency Cepstral Coefficients variability (B) and Mean f0 (C) of the accentual vowels with the DAF of 0, 60 and 120 ms (from left bloc to right bloc in each image), with f0 shift of 0, 1 and 2 semitones (0, 0a and 0b correspond respectively to blocs of baseline, blocs without f0 shift in group1 and group2)*

Reference

[1] Behroozmand, R., Korzyukov, O., Sattler, L., & Larson, C. R. (2012). Opposing and following vocal responses to pitch-shifted auditory feedback: Evidence for different mechanisms of voice pitch control. *JASA*, *132*(4), 2468–2477.

[2] Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *JASA*, *103*(6), 3153–3161.

[3] Chen, S. H., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. *JASA*, *121*(2), 1157–1163.

[4] Franken, M. K., Acheson, D. J., McQueen, J. M., Hagoort, P., & Eisner, F. (2018). Opposing and following responses in sensorimotor speech control: Why responses go both ways. *Psychonomic Bulletin & Review*, *25*(4), 1458–1467.

[5] Franken, M. K., Acheson, D. J., McQueen, J. M., Hagoort, P., & Eisner, F. (2019). Consistency influences altered auditory feedback processing. *Quarterly Journal of Experimental Psychology*, 1747021819838939.

[6] Kalveram, K. T., & Jäncke, L. (1989). Vowel duration and voice onset time for stressed and nonstressed syllables in stutterers under delayed auditory feedback condition. *Folia Phoniatrica*, *41*(1), 30–42.

[7] Natke, U., & Kalveram, K. T. (2001). Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables. *Journal of Speech, Language, and Hearing Research*.

[8] Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347.