

Magnitude and latency of fundamental frequency response within syllables under frequency shifted auditory feedback and public speaking

Thomas DONATH, Ulrich NATKE, and Karl Theodor KALVERAM
*Heinrich-Heine-University Düsseldorf, Institute of Experimental Psychology,
Universitätsstr.1, 40225 Düsseldorf, Germany
thomas.donath@uni-duesseldorf.de*

Abstract. Previous studies have shown that, during continuous vocalization, voice fundamental frequency (voice F_0) is modified by frequency shifted auditory feedback. In this study, voice F_0 contours were determined for the first two syllables of a non-sense word. This way effects of auditory frequency shift and a social stressor (public speaking) on voice F_0 were investigated. Results showed that voice F_0 is auditorily controlled with a high latency and on supra-segmental level in speech. Public speaking did not have an effect on voice F_0 control, which is attributed to its failure to induce stress when speaking a non-sense word repeatedly.

1. Introduction

As one dimension of prosody, voice F_0 encodes information in speech. Linguistic research has shown that prosodic information can have an effect on syntactic analysis and is used to comprehend discourse structure. In tone languages such as Thai or Cantonese, F_0 contours in single syllables distinguish words lexically (for a review see Cutler, Dahan & van Donselaar, 1997). Furthermore, correct production of prosody is of immediate biological importance because it conveys the emotional states of speakers.

Since correct production of voice F_0 is important in communication, it seems reasonable to assume a system which monitors and controls voice F_0 . Studies have repeatedly demonstrated the existence of an auditory-vocal system, which monitors and controls voice F_0 through a closed-loop negative feedback system. When artificially increasing or decreasing the pitch of auditory feedback during vocalization, subjects change their voice F_0 in the opposite direction to compensate for the difference between perceived and intended pitch (e.g., Larson, 1998).

Most previous studies employed a paradigm in which pitch of auditory feedback is modified unexpectedly while subjects produce a vowel continuously for several seconds. Although this paradigm has provided much insight into the audio-vocal system, it is hardly comparable to voice production during speech, which is characterized by rapid onsets and offsets of phonation. Natke and Kalveram (2001a) introduced a paradigm, in which subjects had to utter the non-sense word [tatatas] with different stress patterns. Responses in the opposite direction of frequency shift were found in long stressed, but not in unstressed and therefore short syllables. In second syllables, opposing responses were found regardless of the syllable's duration. The authors concluded that high response latencies prevent responses in unstressed syllables, but long stressed and subsequent syllables are affected by feedback. In the study of Natke and Kalveram, average voice F_0 was calculated for entire syllables and latencies of response were not determined.

Continuing the previous line of research, in the present study effects of frequency-shifted auditory feedback on voice F_0 contours in successive syllables were investigated. This way the continuous change of voice F_0 within syllables was addressed. In addition, the potential effect of socially induced stress on voice F_0 control was addressed in an explorative way. There is evidence that responses to frequency-shifted feedback are smaller or even absent in persons who stutter (Natke & Kalveram, 2001b). Nudelman, Herbrich, Hoyt and Rosenfield (1987) found that persons who stutter reproduce a frequency-modulated tone by humming less accurately than persons who do not stutter. These exemplary studies point towards a potentially deviant audio-vocal control in persons who stutter. Interestingly, stuttering is reduced when persons speak to young children, animals, or in absence of others (Bloodstein, 1949). Therefore, if stuttering is related to a deviant function of the audio-vocal system, social context seems to have an effect on the latter. Consequently, a "public speaking paradigm", which is known to induce stress (e.g., Rohrman, Hennig & Netter, 1999), was introduced in this study to investigate whether it has an effect on audio-vocal control in persons who do not stutter.

2. Method

Twenty-two adults (11 women, 11 men) between 19 and 29 years of age participated in this study ($M = 23$ years, $S.D. = 3.4$ years). Exclusionary criteria were a self-reported current speech disorder and/or a mother-tongue other than German. A hearing-screening assured that subjects had 20 dB HL or better pure-tone thresholds bilaterally (audiometric test: Hortmann DA 323, Neckartenzlingen, Germany).

The subject's voice was frequency-shifted using a commercial device (DFS 404, Casa Futura Technologies, Boulder, Colorado, USA), which works digitally (32 kHz, 14 bits). Auditory feedback was provided through sealed headphones (Blackhawk DSP 5DX, Flightcom, Portland, Oregon, USA), which have a built-in microphone to pick up subjects' voices. Feedback volume was adjusted in such way that a sine-tone of 440 Hz with 75 dB (A) at the microphone led to a headphones feedback volume of 70 dB (A). In order to mask bone conducted sound, low-pass-filtered white noise ($f_c = 900$ Hz) was presented at an intensity of 70 dB (A) during the two speech phases. Vibrations of vocal folds were recorded with an electroglottograph (Laryngograph, Kay Elemetrics, Pine Brook, New Jersey, USA) and stored digitally (11,025 Hz; 16 bits).

Subjects had to utter the non-sense word ['ta:tatas] at a speech rate and with a stress pattern as indicated by a tone sequence. Subjects were instructed to speak clearly and with normal volume. However, no instructions regarding fundamental frequency were given. Each speech phase consisted of 30 trials. The frequency of auditory feedback was shifted downwards by 100 cents in 20% of all trials. In these trials, digital frequency shift was automatically turned on before subjects uttered the word and turned off after they had finished the utterance.

The six voice F_0 contours of frequency-shifted trials and the six contours of trials immediately preceding frequency-shifted trials were averaged for each subject and then compared by calculating the deviation from each other in cents. This way the differences in voice F_0 contours due to frequency shift were revealed. By using trials within the experimental phase rather than a preceding baseline as a reference, error introduced by slow fluctuations of voice F_0 over the course of several trials should have been minimized. For a detailed description of the method, refer to Donath, Natke and Kalveram (submitted).

In order to test for an effect of frequency shift on the initial and last portion of voice F_0 contours as well as an effect of non-public speaking vs. public speaking, the contours for the first two syllables were arbitrarily divided into 25 ms intervals. Non-parametric Wilcoxon rank sign tests were calculated; one-tailed for the intervals in which we expected responses in the opposite direction of the frequency shift, two-tailed for other intervals. The response latency was determined with a non-parametric change-point test (Sigel & Castellan, 1988).

In one of the two speech phases (order randomized), the investigator and another scientist sat two meters in front of the subject and took notes in a standardized manner. Additionally, a video camera was placed behind the two observers and recorded the subject, as indicated by the red recording LED. To validate whether the public speaking condition caused stress in subjects, affect and heart rate were measured. The Positive Affect / Negative Affect Schedule Scale (PANAS-Scale, Watson, Clark & Tellegen, 1988) was used to measure subjective level of positive affect (activation, positively perceived involvement) as well as level of negative affect (fear, anger, nervousness). The questionnaire was given only once to each subject to prevent pseudo effects due to re-testing. Heart rate was measured with a device commonly used in sports, in form of a belt strapped around the subject's chest (Polar M21, Polar Electro, Oulo, Finland). To ensure that subjects had comparable levels of physical activation at the beginning and to reduce carry-over effects, two rest phases, each 15 minutes long, were added to the experimental design. Subjects sat quietly in a chair and could read before the first speech phase and in-between the two speech phases.

3. Results

Figure 1 shows the change in voice F_0 contours introduced by frequency-shifted auditory feedback for the first syllable. Under non-public speaking, there is no significant difference between frequency-shifted and regular trials in the interval 25-50 ms after vowel onset ($p_{2-tailed} = 0.714$, $Z = -0.382$, $n = 22$). Voice F_0 during frequency-shifted auditory feedback begins to increase at 157 ms ($p_{1-tailed} < 0.001$, $Z = -5.541$). The response velocity is 339 cents/s in the interval 150-250 ms after vowel onset, as measured by calculating the slope with a linear regression. In the interval 225-250 ms after vowel onset, mean voice F_0 is 28.3 cents higher than in regular trials ($p_{1-tailed} < 0.001$, $Z = -3.380$, $n = 20$).

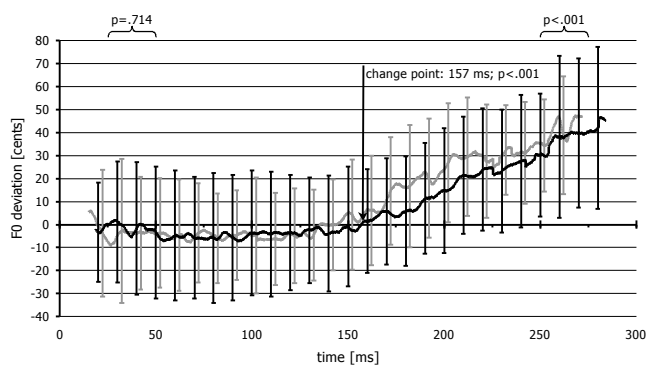


Fig. 1: Average deviation of voice F_0 contours of first stressed syllables (long) during frequency shift from the contours before frequency shift, with S.D. The contour ends at the point for which N becomes less than 8. — non-public speaking, - - - public speaking. The two p -values for intervals were calculated for non-public speaking to determine whether the interval is different from 0. Calculation of change point is based on non-public speaking trials.

Figure 2 shows the change in voice F_0 contours introduced by frequency-shifted auditory feedback for the second syllable. In the interval 25-50 ms after vowel onset, mean voice F_0 is 51.5 cents higher compared to regular trials ($p_{1-tailed} < 0.001$, $Z = -3.924$, $n = 21$). In the interval 75-100 ms after vowel onset, mean voice F_0 is 51.1 cents higher compared to regular trials ($p_{1-tailed} < 0.001$, $Z = -3.517$, $n = 17$).

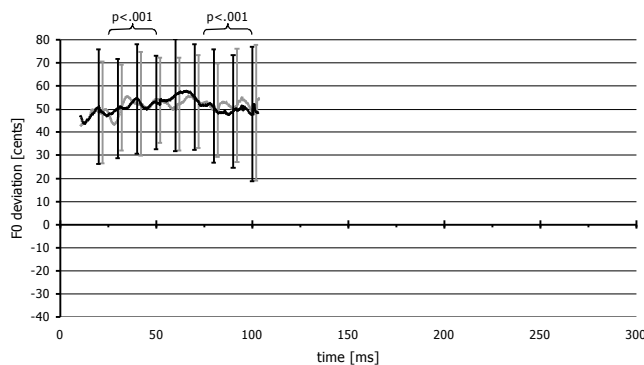


Fig. 2: Average deviation of the voice F_0 contours of second unstressed syllables (short) during frequency shift from the contours before frequency shift, with S.D. The contour ends at the point for which N becomes less than 8. — non-public speaking, - - - public speaking. The two p -values for intervals were calculated for non-public speaking to determine whether the interval is different from 0.

During public speaking, voice F_0 contours in frequency-shifted trials differ from contours in regular trials in the same way as under non-public speaking. (First syllable: $p_{2-tailed} = 0.500$ (25-50 ms), $p_{1-tailed} < 0.001$ (225-250 ms); second syllable: $p_{1-tailed} < 0.001$ (25-50 ms), $p_{1-tailed} < 0.001$ (75-100 ms).) There are no significant differences between non-public speaking and public speaking in any of the 25 ms intervals of either first or second syllable (all $p_{2-tailed} > 0.210$). There is no significant difference in heart rate between non-public speaking and public speaking ($p_{1-tailed} = 0.446$, $Z = -0.156$, $n = 21$). Also, there are no significant differences in positive affect or negative affect between non-public speaking and public speaking (Mann-Whitney-U-Test: positive affect: $p_{1-tailed} = 0.432$, $Z = -0.200$, $n = 11$; negative affect: $p_{1-tailed} = 0.379$, $Z = -0.330$, $n = 11$).

4. Discussion

No differences were found in voice F_0 contours under non-public speaking vs. public speaking. This must not necessarily indicate that the presence of an audience does not influence the audio-vocal system. It seems that the public speaking paradigm does not induce a significant amount of stress when subjects speak non-sense words. Speaking the same word repeatedly requires only very little cognitive effort and probably does not induce speech anxiety based on evaluation of speech. Future studies investigating effects of audiences on stuttering should take this result into account and employ free speech or a demanding reading task, which bring social evaluation with them, as it is common in public speaking paradigms.

Adjustment of voice F_0 can only occur after a duration, in which auditory feedback is processed, the efferents to the larynx are modified, and physical changes of vocal folds take place. The high latency of the response (157 ms) prevents auditorily controlled realization of intended voice F_0 in short, unstressed syllables and during the initial portion of long syllables. Therefore auditory feedback would not be very useful for adjusting voice F_0 in syllables as they are being produced, but more useful in the regulation of voice F_0 in later segments. This could serve to adjust for temporarily changed conditions of the larynx and to control supra-segmental prosody. By monitoring actually realized voice F_0 of syllables, adjustments could be made for later syllables, that means on a supra-segmental level, so that the relative differences in voice F_0 encoding the information are produced reliably.

The maximum voice F_0 response was approximately 50 cents in this study, even though a complete compensation would have required a response of 100 cents. This maximum response magnitude agrees well with earlier findings (e.g., Larson, 1998: approximately 30 cents; Natke & Kalveram, 2001a: approximately 30-60 cents), even though different paradigms such as continuous vocalization and non-sense words, and frequency shifts varying from 100 to 600 cents were employed. These magnitudes of less than a semi-tone may seem small, but the compensation was easily audible in the audio records and falls well within natural prosodic variations.

Although these findings support a closed-loop negative feedback model of voice F_0 regulation, no study has yet demonstrated a satisfying correlation between frequency shift magnitudes and resulting response magnitudes. This is not surprising because most studies have used frequency shifts of 100 cents and more, and only one frequency shifts as low as 25 cents (Burnett, Senner & Larson, 1998), while responses seem to be limited to 30-60 cents. Therefore the range in which a correlation between frequency shift magnitude and resulting response magnitude might be observed was not investigated systematically.

The question remains why the response does not exceed approximately 30 to 60 cents regardless of frequency shift magnitude. Besides auditory feedback, proprioceptive feedback plays a role in voice F_0 regulation (Tanabe, Kitajima & Gould, 1975). In an experimental frequency shift condition the two feedback channels carry

different information: proprioceptive feedback reports the laryngeal configuration for actual F_0 , whereas auditory feedback reports a deviation. The integration of both feedback channels might limit the maximum response non-linearly.

It might also be that voice F_0 simply is not controlled very tightly in speech. Sundberg (1987) showed that trained singers can match an external reference tone of 440 Hz with an accuracy of more than 1 Hz (approximately 4 cents). Furthermore, it was shown that an unpredictable frequency shift can be completely compensated for when singing scales without an external reference (Burnett, Senner & Larson, 1997). This indicates that the compensatory mechanism for voice F_0 can be very effective when a reference frequency is represented externally or internally. When speaking syllables, an intended value for voice F_0 in syllables may also be given. However, in speaking, relative changes in voice F_0 related to prosodic information seem to be more important than the production of absolute F_0 values. Secondly, in languages such as English or German, voice F_0 appears to be of little importance for precise comprehension of speech. Thus, it seems less important for the speaker to monitor and regulate voice F_0 within syllables. This of course does not apply to tone languages, in which syllable's voice F_0 contours have lexical function. Consequently, although a compensatory mechanism for voice F_0 at the syllabic level may exist, it seems that for languages such as English and German control of voice F_0 on supra-segmental level is more important (e.g., for word focus or conveyance of emotional states).

While a partial correction of a mismatch between intended and auditorily perceived voice F_0 occurs in first syllables with a latency, second syllables show a higher F_0 from the very beginning. Since syllables were separated by a voiceless consonant, phonation stopped in-between them. Therefore control of voice F_0 is continuous and not interrupted by the onset and offset of phonation itself. This is an indication that the audio-vocal system monitors voice F_0 independently of phonation boundaries and adjusts voice F_0 on supra-segmental level.

Acknowledgments

This research was supported by grant Ka 417/28-1 of the Deutsche Forschungsgemeinschaft (DFG).

References

- Bloodstein, O. (1949) Conditions under which stuttering is reduced or absent: A review of literature. *J. Speech Hear. Dis.* 14:295-302
- Burnett, T. A., Senner, J. E. & Larson, C. R. (1997) Voice F_0 Responses to pitch-shifted auditory feedback: A preliminary study. *J. Voice* 11:202-211
- Burnett, T. A., Senner, J. E. & Larson, C. R. (1998) Voice F_0 responses to manipulations in pitch feedback. *J. Acoust. Soc. Am.* 103:3153-61
- Cutler, A., Dahan, D. & van Donselaar, W. (1997) Prosody in the comprehension of spoken language: A literature review. *Language and Speech* 40:141-201
- Donath, T., Natke, U. & Kalveram, K. Th. (submitted) Effects of frequency-shifted auditory feedback on voice F_0 contours in syllables.
- Larson, C. R. (1998) Cross-modality influences in speech motor control: The use of pitch-shifting for the study of F_0 control. *J. Comm. Dis.* 31:489-503
- Larson, C. R., Burnett, T. A. & Kiran, S. (2000) Effects of pitch-shift velocity on voice F_0 responses. *Journal of the Acoustical society of America* 107:559-564
- Natke, U. & Kalveram, K. T. (2001a) Effects of frequency shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables. *J. Speech Lang. Hear. Res.*, 44
- Natke, U. & Kalveram, K. T. (2001b) Fundamental frequency and vowel duration under frequency shifted auditory feedback in stuttering and nonstuttering adults. In H.-G. Bosshardt, J.S. Yaruss & H. F. M. Peters (Eds.), *Fluency Disorders: Theory, Research, Treatment and Self-help. Proceedings of the Third World Congress of Fluency Disorders* in Nyborg, Denmark. Nijmegen: Nijmegen University Press, 66-71.
- Nudelman, H. B., Herbrich, K. E., Hoyt, B. D. & Rosenfield, D. B. (1987) Dynamic characteristics of vocal frequency tracking in stutterers and nonstutterers. In H. F. M. Peters & W. Hulstijn (Eds.). *Speech motor dynamics and stuttering*. New York: Springer
- Rohrman, S., Hennig, J. & Netter, P. (1999) Changing psychobiological stress reactions by manipulating cognitive processes. *Int. J. Psychophysiology* 33:149-161
- Siegel, S. & Castellan, N. J. (1988) *Nonparametric statistics for the behavioral sciences*. 2nd ed. Boston: McGraw-Hill.
- Sundberg, J. (1987) *The science of the singing voice*. Dekalb, IL: Northern Illinois University Press
- Tanabe M., Kitajima K. & Gould W. (1975) Laryngeal phonatory reflex: The effect of anesthetization of the internal branch of the superior laryngeal nerve – Acoustic aspects. *Ann. Otol. Rhinol. Laryngol.* 84:206-212
- Watson, D., Clark, L. A. & Tellegen, A. (1988) Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology* 54:1063-1070