Hearing Research 324 (2015) 1–6

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research paper

Audio–vocal responses of vocal fundamental frequency and formant during sustained vowel vocalizations in different noises

Shao-Hsuan Lee [a], Tzu-Yu Hsiao [b], Guo-She Lee [a, c, *]

[a] Faculty of Medicine, School of Medicine, National Yang-Ming University, Taipei, Taiwan
[b] Department of Otolaryngology, National Taiwan University Hospital and College of Medicine, National Taiwan University, Taipei, Taiwan
[c] Taipei City Hospital, Ren-Ai Branch, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 25 August 2014
Received in revised form
24 February 2015
Accepted 26 February 2015
Available online 5 March 2015

ABSTRACT

Sustained vocalizations of vowels [a], [i], and syllable [mə] were collected in twenty normal-hearing individuals. On vocalizations, five conditions of different audio–vocal feedback were introduced separately to the speakers including no masking, wearing supra-aural headphones only, speech-noise masking, high-pass noise masking, and broad-band-noise masking. Power spectral analysis of vocal fundamental frequency (F0) was used to evaluate the modulations of F0 and linear-predictive-coding was used to acquire first two formants. The results showed that while the formant frequencies were not significantly shifted, low-frequency modulations (<3 Hz) of F0 significantly increased with reduced audio–vocal feedback across speech sounds and were significantly correlated with auditory awareness of speakers' own voices. For sustained speech production, the motor speech controls on F0 may depend on a feedback mechanism while articulation should rely more on a feedforward mechanism. Power spectral analysis of F0 might be applied to evaluate audio–vocal control for various hearing and neurological disorders in the future.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Speech communication relies on sophisticated sensory-motor integration of both central and peripheral nervous systems. The model of Directions Into Velocities of Articulators (DIVA) is one of the theoretical models that helps to explain the audio–vocal feedback system in terms of neural network and cortical interactions (Guenther, 2006). For keeping a stable speech, DIVA model suggests that the feed-forward control for speech output is performed on the basis of learned motor commands, while the auditory-feedback modification of phonation is mainly induced by the mismatches between the actual auditory feedback signals and the auditory sensory expectations (Tourville et al., 2008). A number of studies have confirmed that auditory feedback is one of the most important sensory information contributing to the learning and stability of phonation and articulation in human speech, and there are interactions between speech production and auditory reception which tend to induce active and reflexive control of vocal-fold vibrations and speech articulation in response to auditory interference. Speakers are likely to show significant changes in vocal fundamental frequency (F0), formant transitions, vocal intensity, speech rate, and/or nasal resonance when auditory feedback of self-generated voice is delayed, pitch-shifted, noise-masked, or greatly attenuated. These observations bolster audio–vocal feedback loop as a key to maintain speech stability (Chen et al., 2007; Ferrand, 2006; Hain et al., 2001; Larson et al., 2007, 2001; Lee et al., 2007).

Even in sustaining an as-steady-as possible vowel, F0s are not constant throughout the entire phonation (Titze, 1991; Titze et al., 1993). Rhythmic fluctuations of F0 do exist and were deduced to originate from the modulations of auditory feedback, aerodynamics of vocal production, or inherent irregularities in the nature of laryngeal muscle contractions (Titze, 1991). Each cycle of vocal fold vibrations is not exactly the same in time. The rhythmic fluctuations of vocal fold vibrations are different in frequencies and are

Abbreviations: ANOVA, analysis of variance; DIVA, directions into velocities of articulators; BBN, broadband-noise masking; F1, first formant; F1SD%, standard deviation of first formant frequency in percentage; F2, second formant; F2SD%, standard deviation of second formant frequency in percentage; NO, no-masking hearing status; EO, wearing headphone only; SN, speech-noise masking; HPN, high-pass noise masking; LFP, low-frequency power; MFP, middle-frequency power; HFP, high-frequency power; LPC, linear predictive coding.

* Corresponding author. Department of Otolaryngology, National Yang-Ming University, No. 155, Sec. 2, Li-Nong Street, Bei-Tou District, Taipei City 112, Taiwan. Tel.: +886 2 28267000x6155; fax: +886 2 28202190.

E-mail addresses: satomilee0701@gmail.com (S.-H. Lee), tyhsiao@ntu.edu.tw (T.-Y. Hsiao), guosheli@gmail.com (G.-S. Lee).

http://dx.doi.org/10.1016/j.heares.2015.02.005
0378-5955/© 2015 Elsevier B.V. All rights reserved.

generally classified as vocal wow (0–3 Hz), vocal vibrato (3–8 Hz), and vocal flutter (≥8 Hz). A vocal wow is a periodic variation of lower than 3 Hz underlying the vibrations of vocal folds. This essential instability cannot be totally suppressed even though the speaker has the experiences of voice or singing training. The low-frequency fluctuations imbedded in the signals of cycle-to-cycle vocal fold vibrations have been considered related with the audio–vocal interaction in our previous studies and tended to increase significantly while the speaker sustaining the vowel [a] under disturbed auditory input (Lee, 2012; Lee et al., 2004). It should be emphasized that it is the fluctuations of F0 below 3 Hz being analyzed rather than the vocal F0 itself. A faster pulsation of F0, usually between 3 Hz and 8 Hz, is known as vocal vibrato. A vibrato has been considered associated with active modulation of the laryngeal motor neuron pool (Hsiao et al., 1994) and the control of auditory system (Leydon et al., 2003). It can be deliberately produced, suppressed, or modified after training. The rhythms of faster than 8 Hz in F0 are another source of vocal fluctuations known as vocal flutters. The rapid oscillations in F0 might represent a natural oscillating of the glottal adductor–abductor control system during phonation (Aronson et al., 1992).

Our previous findings showed that the low-frequency rhythms in F0 significantly increased in the normal-hearing speakers with noise masking (Lee et al., 2004, 2007) and in the post-lingual and the pre-lingual hearing-impaired speakers (Lee, 2012; Lee et al., 2013). The findings provided evidence that the involuntary modulations of vocal-fold oscillations was associated with the auditory feedback responding to the mismatch between anticipated and actual auditory information from self-generated speech. However, the speech material and the type of noise had been limited to vowel [a] and speech noise, so it remains unclear whether other speech sounds and/or a different type of noise masking will also alter the subsequent audio–vocal feedback modulations of F0 and even speech articulation in the same way. Therefore, we included three speech sounds with different formant frequencies to clarify if there is a dependence of F0 feedback on formant energy. We also used noise masking of different frequency bands to explore the responses of F0, as well as formant frequencies, to the information loss of formant energy. The audibility of vocalization was also evaluated to investigate the relationship between F0 feedback and auditory attention system. All speakers were requested to produce the vowels and syllable in tone 1.

2. Methods

2.1. Participants

Twenty participants (10 males and 10 females), aged between 20 and 40 years, having no medical history of neurological deficits, speech-language disorders, current upper respiratory infection, or the experience of voice singing training were enrolled. All participants passed the hearing screening test which was defined as a pure-tone hearing threshold level of better than or equal to 25 dB HL at the frequencies of 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz, and 8000 Hz. The participants were all native Mandarin speakers. The research procedures were approved by the Institutional Review Board of National Yang Ming University (IRB-960014), and the informed consent was acquired from each participant.

2.2. Sampling of voice

Voice recordings were conducted in a sound-treated room in which background noise was lower than 40 dBA monitored by a sound-level meter. On the assumption of different audio–vocal feedback for different speech sounds, all participants were instructed to sustain the open vowel [a], the close vowel [i], and the nasalized syllable [mə] as steady as possible for at least 6 s. The nasalized syllable [mə] was included because it elicits an coarticulation of adjacent vowel at the very beginning of the following schwa and serves as a reference for the speaker to purposefully continue the nasalized vowel quality. The vocal intensity was real-time displayed on a laptop computer to help the speakers maintaining their vocal intensity within the range of 70–80 dBA in all auditory conditions.

The microphone-to-mouth distance was maintained at a distance of 15 cm by a stand holder, and the frequency response of the microphone was flat from 31.5 Hz to 8000 Hz (IEC 651 TYPE II, TENMARS Electronics, Taipei, Taiwan). In order to investigate whether and how the different types of nose masking would interfere with the auditory feedback for the speech material, five auditory conditions were introduced to the speakers during vocalizations: no-masking hearing status (NO), wearing headphone only (EO), speech-noise masking (SN, plateau energy from 0.25 kHz to 1 kHz, attenuation by 12 dB per octave from 1 kHz to 11.025 kHz), high-pass noise masking (HPN, plateau energy from 1 kHz to 8 kHz, decay by 12 dB per octave below 1 kHz), and broadband-noise masking (BBN, plateau energy from 0.25 kHz to 11.025 kHz). Two as-steady-possible phonations were recorded for each speech material in each auditory condition, and the analytic results of the two phonations were averaged for later data statistics. The order of the speech sounds and the auditory conditions were both arranged in random for each participant. The introduced noises were generated by a lab-developed program and a built-in sound adapter (ASUS A43S/Realtek high definition audio) and were binaurally introduced to the speakers at the intensity of 85 dBA through the headphones (Telephonics, TDH-50). Calibrations of the noises were accomplished prior to the tests for each participant using a standard sound level meter and a 6-c.c. coupler at the intensity of 80 dBA (Larson Davis system 824, New York, US). To control vocal intensity within the range between 70 and 80 dBA, there was a real-time intensity meter displayed on the screen so as to help the participants control their own vocal intensity. In each listening condition, the phonations were repeated once to acquire averaged data for statistical analysis as our previous works (Lee, 2012; Lee et al., 2004, 2007). No participant reported difficulty of producing the speech materials during voice recordings. The voice signals were obtained with a sampling frequency of 44.1 kHz and stored in a 16-bit format. The software for noise production, hardware controls, signal sampling, and intensity displaying was lab-developed using LabVIEW for Windows (version 6.0i, National Instrument, Austin, Texas, US). For realizing whether or not there is an interaction between auditory awareness and audio–vocal feedback system, right after both vocalizations in each type of auditory conditions, all participants subjectively rated the auditory awareness of their own voices by marking a 12-cm visual analogue scale in which 0 cm denoted “no auditory perception of their own voice” and 12 cm stood for a clear perception of their own voice as in normal listening status.”

2.3. Contour of F_0 and conversion of cents

The procedure details for digital signal processing had been published in our previous study (Lee, 2012). In short, the 5-s voice signals starting at 0.5 s after the voice onset were extracted for signal processing. A 20-ms window including at least two glottal cycles was used to obtain the fundamental period by counting the time at which the autocorrelation function was maximal. That period is compatible with the interval of a glottal wave that repeats itself. Then, the analytic windows were shifted forward by the fundamental periods, and all fundamental periods were retrieved

Download English Version:

https://daneshyari.com/en/article/6287350

Download Persian Version:

https://daneshyari.com/article/6287350

Daneshyari.com