Pitch planning in English and Taiwanese Mandarin: Evidence from startle-elicited responses

Chenhao Chiu^{a)} and Bryan Gick^{b)}

Department of Linguistics, University of British Columbia, 2613 West Mall, Vancouver,
British Columbia, V6T 1Z4, Canada
chenhao@alumni.ubc.ca, gick@mail.ubc.ca

Abstract: Startling auditory stimulus (SAS) can trigger rapid release of pre-planned movement sequences and can effect a physiological perturbation of pitch level. An SAS-based paradigm was used to investigate planning of pitch profiles in English and Taiwanese Mandarin. Results show that startle-elicited pitch level is elevated, whereas pitch contours are preserved. These results support the view that pitch contours are pre-specified in speech plans, while absolute pitch level is not.

© 2014 Acoustical Society of America

PACS numbers: 43.70.Kv [AL]

Date Received: May 8, 2014 Date Accepted: September 14, 2014

1. Introduction

Researchers have long been concerned with uncovering the contents of speech motor plans (e.g., Henke, 1966; Keele, 1968). Startling auditory stimuli (SAS, >120 dB) provide a mechanism by which it is possible to examine the contents of a speaker's forward plans. When elicited by a SAS, a previously prepared upper limb movement (e.g., flexion or extension) can be released at a short latency (muscle pre-motor reaction time <70 ms) while the kinematics and accuracy of the movement remain largely unaffected; this latency has been claimed to be too short to allow somatosensory feedback correction in arm movements (Carlsen *et al.*, 2012).

Stevenson *et al.* (2014) extend the SAS experimental paradigm to speech production, showing that prepared English CV syllables elicited by SAS are initiated at shortened latencies, with lower-lip movements for [ba] initiated earlier in SAS-elicited responses (Mean = 75 ms) than in control responses (Mean = 116 ms). In English syllables thus triggered, lip kinematics and vowel F1 and F2 are unaffected by the presentation of a SAS, indicating that these parameters are encoded in the speech plan and are not dependent on afferent feedback. However, results for parameters relating to pitch control, such as pitch level and contour, are not reported by Stevenson *et al.* (2014). This leaves an important gap, as previous research shows that vowel formants may not be bound by a fixed relationship with pitch during vowel production (Liénard and Di Benedetto, 1999; Traunmüller and Eriksson, 2000), suggesting that these parameters may be dissociated in speech plans. Whether pitch information is also part of forward speech plans and thus resistant to change in response to SAS has not been investigated. The present study addresses this gap by examining how pitch is manifested under startle and non-startle conditions.

The SAS paradigm is ideally suited for investigating pitch control, both because a SAS can elicit rapid response of prepared movement sequences, allowing less time for effects of sensory feedback, and because the associated startle reflex provides a natural, physiological pitch perturbation. Following presentation of a SAS, a brief physiological startle reflex response occurs (muscle pre-motor reaction time around

^{a)} Author to whom correspondence should be addressed.

^{b)} Also at Haskins Laboratories, New Haven, Connecticut.

40 ms), followed by accelerated release of the prepared movement sequence (see [Stevenson *et al.*, 2014](#), for details); by the time of the voice onset (Mean = 204 ms), this initial startle reflex has long since been resolved. [Baer \(1979\)](#) used a startle methodology to show that a SAS can function as a physiological perturbation to pitch level: When an unexpected startling auditory stimulus was presented during constant phonation, the startle reflex caused an increase in laryngeal tension resulting in a momentary elevation in pitch level, followed by a correction (i.e., drop off) back to baseline pitch level.

Extrapolating from the findings of [Baer \(1979\)](#) for constant phonation to the production of spoken syllables, let us consider the sequence of pitch-related events beginning with presentation of the SAS. First, a speaker who is planning to produce a spoken syllable hears a SAS; after about 40 ms, the first premotor reaction associated with the startle reflex can be observed, initiating an increase in laryngeal tension. As soon as 20 ms after the startle reflex has begun to perturb laryngeal tension, any initial somatosensory (e.g., proprioceptive) feedback-based correction may begin to take place (cf. [Larson *et al.*, 2008](#)). Over 100 ms later, the physiological reflex and the initial response having resolved, voicing onset occurs (~204 ms); vocal fold vibration offers a second opportunity for somatosensory (e.g., vibrotactile, aerotactile) feedback to take effect at a further latency of ~20 ms after the pitch onset (cf. [Larson *et al.*, 2008](#)). If auditory feedback plays a role in correction, compensation would occur at an additional latency from as early as 100–150 ms ([Hain *et al.*, 2000](#)) up to 210 ms ([Jones and Munhall, 2002](#)) following voicing onset. Thus, if initial pitch level is pre-specified in the speech plan (i.e., if there exists a specified baseline to which the system might be corrected), proprioceptive feedback streams should have had sufficient time to correct for SAS perturbation to laryngeal configuration prior to vowel onset. However, if initial pitch level is not part of the speech plan, an uncorrected increase in onset pitch level may be expected to persist in SAS-induced responses. Taking the above observations into account, we expect vowel-initial pitch levels to be elevated by SAS perturbation.

Compared with pitch level, pitch contour in Mandarin appears to be quite stable, possibly because of its more crucial role in maintaining lexical contrasts. [Liu *et al.* \(2007\)](#), for example, show that while Mandarin-speaking mothers use higher pitch to address infants than to address adults, the contours associated with phonemic contrasts are maintained and even exaggerated between adult-directed and infant-directed speech. Thus, lexical contrasts can be maintained by preserving the contour profiles that are used to differentiate lexical tones, despite changes in absolute pitch level. If absolute pitch level is part of the speech plan, a correction to baseline pitch level in SAS-induced responses is anticipated at some point during the vowel (i.e., auditory feedback-based correction), compromising pitch contour. However, if pitch contour is pre-specified but absolute level is not (as we might expect from previous literature we described above), the contour profile should be preserved when elicited by a SAS, even if at a higher pitch level. Thus, while we predict an elevated pitch level at vowel onset, we do not expect to see a rapid drop off in pitch signifying feedback-based correction following the physiological response. Rather, we expect to see pitch contours maintained, though at an elevated level. This may be especially true for a tone language such as Mandarin, in which pitch contour denotes phonemic contrast.

2. Method

Except as described here, apparatus and procedures for the present experiment were exactly as described for experiment 2 in [Stevenson *et al.* \(2014\)](#).

Participants sat in an upright chair facing a computer monitor (Viewsonic, VE175, 17", 75 Hz refresh rate) at a distance of approximately 0.5 m. They were instructed to look straight ahead at the monitor and respond to an acoustic stimulus, either 80 dB (control) or 124 dB (startle), by vocalizing the target syllable as quickly as possible. For the English experiment, acoustic data collected by a customized

LabView program from [Stevenson *et al.* \(2014\)](#), including pitch, F1, and F2, were reanalyzed.

For the Taiwanese Mandarin experiment, the testing was recorded and videotaped.

Acoustic data were collected by a microphone directly connected to the computer and recorded using the computer program PsychoPy (sampled at 44.1 kHz). Data collection began 160 ms before the presentation of the stimulus and was terminated 2000 ms later.

Video was recorded using Lumix digital camera (Panasonic Model DMC-LX3). A visual display of the syllable [ba] was presented on the monitor concurrently with the acoustic stimulus. In the Taiwanese Mandarin experiment, the syllable [ba] was matched with each of the four lexical tones and presented in blocks, so that all syllables in each block contained the same tone. Target syllables were visually displayed using Taiwanese phonetic orthography (Zhuyin bopomofo), with tonal information specified. All the testing was conducted in a sound-attenuated room at National Chiao Tung University, Hsinchu, Taiwan.

Fifteen Taiwanese Mandarin native speakers were recruited. Prior to testing, participants signed an informed consent release and were uninformed about the research question under investigation. The experiment was conducted following the ethical guidelines established by the University of British Columbia (UBC). Video recordings of all the participants were examined and coded for startle reflex behavior by a trained speech researcher (the first author). Seven participants (two male and five female) exhibited consistent startle reflex behavior, including shoulder shrugging and facial grimace, in response to SAS. These participants were defined as consistent startlers, and only data collected from these seven startlers were analyzed.

A total of 34 of the 700 Taiwanese Mandarin trials (4.8%) were excluded from analysis, due either to slow voice onset (defined as 2 SD from the mean) or to participants' unfamiliarity with the trial event at the beginning of the testing. For both English and Taiwanese Mandarin, acoustic signals were processed using Praat. Syllable boundaries were visually examined by the experimenter and manually marked. Before marking syllable boundaries, the acoustic waveforms were first filtered via de-emphasis at 50 Hz.

The formants were extracted using the LPC formant tracker, with a window from 0 to 2000 Hz, and a window length of 0.0125 s. Data were then analyzed across normalized durations. Voicing onset was identified and marked based on the appearance of substantial acoustic wave bursts; the end of the syllable was identified and marked as the end of the periodic waveforms. Acoustic burst release times and syllable durations were used to measure reaction time and submitted as dependent variables. Measurement analyses were performed using a two stimulus $\times$ four tone repeated measures analysis of variance (ANOVA). Pitch and formant transitions from both English and Taiwanese Mandarin were analyzed using a smoothing spline analysis of variance (SS ANOVA) to examine the fit of each variant ([Davidson, 2006](#); [Derrick and Schultz, 2013](#)).

Table 1. Average acoustic burst release times and syllable durations (in ms) across 4 Taiwanese Mandarin tones. Standard deviations are shown in parentheses.

	Burst release time		Syllable duration	
	Control	Startle	Control	Startle
Tone 1	297.07 (70)	225.49 (49)	275.32 (103)	252.38 (79)
Tone 2	272.37 (67)	233.01 (48)	349.49 (127)	313.03 (109)
Tone 3	244.24 (74)	207.96 (55)	271.07 (124)	257.12 (122)
Tone 4	209.40 (68)	180.06 (43)	206.47 (60)	209.99 (63)

Fig. 1. Taiwanese Mandarin F1 and F2 range for control and startle responses. SS ANOVA results of F1 and F2 frequency range across 4 tones. Each shaded band represents a 95% confidence interval. The white space between the transition lines for each condition represents a statistically significant difference between the given measures.

3. Results

As with the English results reported in [Stevenson *et al.* \(2014\)](#), in Taiwanese Mandarin, significantly shorter latencies for acoustic burst release time were reported for SAS-induced responses than control responses (Table 1). Main effects were observed for stimulus [$F(1, 6) = 10.87, p = 0.02$], tone [$F(3, 18) = 8.94, p < 0.01$], and the stimulus $\times$ tone interaction [$F(3, 18) = 4.9, p = 0.01$]. *Post hoc* analyses revealed that the reaction time latency for tone 4 was significantly shorter than that for tone 1 and tone 2. The main effects for tone and stimulus $\times$ tone interaction may be interpreted as a practice effect, as reaction time latencies gradually decreased from tone 1 to tone 4. Syllable durations are summarized in Table 1. No effect for stimulus or tone was

Fig. 2. SS ANOVA comparison of pitch over English [ba] durations. Other legends follow Fig. 1.

reported. No statistically significant difference was found for either F1 or F2 (as indicated by the overlaps between the control and startle fits in Fig. 1), suggesting that formant profiles, acting in a phonemic role, are preserved in SAS-induced responses.

For English, SAS-induced pitch levels were significantly higher than those of control responses, as indicated by the distinct SS ANOVA curves shown in Fig. 2. The onset frequency in SAS-induced responses was elevated by 40–50 Hz, compared to control responses. It is also noted that pitch contour, although not phonemic in English, was unaffected by the SAS.

For Taiwanese Mandarin, as with English, SAS-elicited pitch levels were elevated and distinguished from control responses throughout most of their duration (Fig. 3). At vowel onset, substantial pitch elevation was observed for all four tones, with an increase of 39 Hz (≈ 222 cents) for tone 1; 27 Hz (≈ 189 cents) for tone 2; 23 Hz (≈ 166 cents) for tone 3; and 30 Hz (≈ 167 cents) for tone 4. Pitch contour for each tone, on the other hand, remained largely unchanged by the SAS. Also visible in Fig. 3, tone-specific variance is maintained across the different tones (i.e., tones 3 and 4 exhibit more variance than tones 1 and 2 in control responses, and this variance is preserved in startle conditions).

4. Discussion

The results for English and Taiwanese Mandarin indicate that the introduction of a SAS triggers a rapid release of prepared syllables. SS ANOVA results for F1 and F2 in both languages confirm that formants remain largely unaffected by a SAS, supporting the view that formant profiles are pre-specified in the speech plan. A prepared target syllable ought minimally to include phonemic information (i.e., that which is necessary for lexical contrast); as formants play a phonemic role in both English and Taiwanese Mandarin, their pre-specification is thus expected. When a SAS is presented, the prepared program is elicited at a shorter latency, but the phonemic goals are achieved as intended.

For English, the pitch level of SAS-induced responses was found to be significantly elevated throughout the duration of the syllable. As reported in Hain *et al.* (2000), pitch compensation can occur as early as 100–150 ms after the onset of auditory feedback. Thus, for SAS-induced responses, if any feedback correction in pitch

Fig. 3. SS ANOVA results of pitch profiles across 4 Taiwanese Mandarin tones. Other legends follow Fig. 1.

were to occur, it could start as early as 100 ms after the voice onset. No evidence of feedback-based correction of pitch level to a pre-specified baseline was observed at any stage in the production. Thus, it appears that absolute pitch level is not pre-specified in forward plans for English. The elevation can be explained as relating to a physiological startle reflex, with an increase in laryngeal tension presumably acting as a protective maneuver in response to the SAS (Baer, 1979).

In a tonal language such as Taiwanese Mandarin, pitch plays a phonemic role in speech production; on the assumption that forward plans encode phonemically contrastive information, one might predict that, for Taiwanese Mandarin, such a plan would include both absolute pitch level and pitch contour, while this may not be the case for a non-tonal language such as English. The results from the current study, however, show that both languages respond similarly to the SAS, with initial pitch levels being compromised (elevated) in both languages, while pitch contour is preserved throughout the syllable for all four tones. As summarized in Table 1, syllable durations across the four Taiwanese Mandarin tones ranged from 210 to 313 ms. The effects of afferent (sensory) feedback should thus have been expected to begin as early as 30%–50% of the way through the syllable's production. However, as shown in Fig. 3, our results show that for all SAS-induced tonal contours, no attempt to correct the pitch to an absolute or baseline level was observed.

Together with the findings in Stevenson *et al.* (2014), the observations we have made here reveal a good deal of similarity between speech plans and other SAS-induced (e.g., upper limb) movements. In SAS-induced speech, overall trajectories for a prepared response are generally maintained, whether relating to vowel formants, pitch, or lip movements. Likewise, while a SAS causes increased EMG muscle activity compared with control responses for upper limb movements (e.g., Carlsen *et al.*, 2012), similar effects are also observed in SAS-induced speech movements (cf. Stevenson *et al.*, 2014, experiment 2). In the present study, amplified activity in laryngeal muscles elicited by a SAS may also account for the increased laryngeal tension and concomitant elevation in pitch level.

The difference observed in the present study between pitch level and contour may relate to the observation that speakers are capable of producing reliable pitch contours despite a wide range of baseline vocal pitch levels in the speaking population. While the present study provides evidence supporting the view that pitch contour is included in speech plans and under feedforward control, absolute pitch level may be implemented at a later stage in production. This suggestion, that pitch contour and pitch height are introduced at different stages of production, calls for future research.

Acknowledgments

This work was supported by the National Institute of Health Grant to Haskins Laboratories. We acknowledge Dana Maslovat, Andrew Stevenson, Romeo Chua, Ian Franks, and Saulina Wong at UBC for their collaboration. We also give our thanks to Yu-an Lu at the National Chiao Tung University for participant recruitment and lab facilities.

References and links

- Baer, T. (1979). "Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes," *J. Acoust. Soc. Am.* **65**, 1271–1275.
- Carlsen, A. N., Maslovat, D., and Franks, I. M. (2012). "Preparation for voluntary movement in healthy and clinical populations: Evidence from startle," *Clin. Neurophysiol.* **123**, 21–33.
- Davidson, L. (2006). "Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance," *J. Acoust. Soc. Am.* **120**, 407–415.
- Derrick, D., and Schultz, B. (2013). "Acoustic correlates of flaps in North American English," in *Proceedings of Meetings on Acoustics*, Montréal, QC, Canada (Acoustical Society of America, Melville, NY), 5aSCb16.
- Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., and Kenney, M. K. (2000). "Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex," *Exp. Brain Res.* **130**, 133–141.

- Henke, W. (1966). "Dynamic articulatory model of speech production using computer simulation," Ph.D. thesis (Massachusetts Institute of Technology, Cambridge, MA).
- Jones, J. A., and Munhall, K. G. (2002). "The role of auditory feedback during phonation: Studies of Mandarin tone production," *J. Phonetics* **30**, 303–320.
- Keele, S. W. (1968). "Movement control in skilled motor performance," *Psychol. Bull.* **70**, 387–403.
- Larson, C. R., Altman, K. W., Liu, H., and Hain, T. C. (2008). "Interactions between auditory and somatosensory feedback for voice F_0 control," *Exp. Brain Res.* **187**, 613–621.
- Liénard, J.-S., and Di Benedetto, M.-G. (1999). "Effect of vocal effort on spectral properties of vowels," *J. Acoust. Soc. Am.* **106**, 411–422.
- Liu, H.-M., Tsao, F.-M., and Kuhl, P. K. (2007). "Acoustic analysis of lexical tone in Mandarin infant-directed speech," *Dev. Psychol.* **43**(4), 912–917.
- Stevenson, A. J., Chiu, C., Maslovat, D., Chua, R., Gick, B., Blouin, J.-S., and Franks, I. M. (2014). "Cortical involvement in the StartReact effect," *Neuroscience* **269**, 21–34.
- Traunmüller, H., and Eriksson, A. (2000). "Acoustic effects of variation in vocal effort by men, women, and children," *J. Acoust. Soc. Am.* **107**, 3438–3451.