Larynx movements and intonation in whispered speech

John Coleman, Esther Grabe and Bettina Braun

(Summary of research supported by British Academy grant SG-36269)

1. Context and Research Objectives. Speakers of all languages change the pitch of their voice to signal linguistic information. In intonation languages such as English, for instance, many questions are accompanied by high or rising pitch. In tone languages such as Chinese, pitch patterns distinguish otherwise identical words from one another. Articulatorily, pitch changes are caused by accelerated or slowed vocal fold vibration. Vocal fold vibration, in turn, is largely controlled by muscular activity in the larynx. In particular, the vocal cords are tenser for high-pitched sounds. Also, the larynx is higher in the throat for high-pitched sounds and lower for lower-pitched sounds. Pitch changes in speech have also been observed, however, when vocal fold vibration is absent. In whispered speech, the vocal cords are held open and they do not vibrate. In principle, listeners should not hear pitch movement. Nevertheless, speech perception research has shown that listeners can identify intonation patterns and tones in whispered speech. It is unclear how listeners can do this, so we examined two conflicting hypotheses:

Hypothesis 1: Whispered speech is produced in the same way as normal speech, but without vocal cord vibration. Consequently, the larynx movements are the same in whispered and normal speech, despite the fact that the larynx movements are apparently unnecessary. They could, however, affect the speech signal.

Hypothesis 2: Since there is no vocal fold vibration in whispered speech, the muscles that control pitch are not active and the larynx does not need to move. In this case, perceived pitch must have another cause.

In order to determine which of these is correct, we used dynamic Magnetic Resonance Imaging to investigate larynx position in normal (voiced) and whispered speech in six untrained volunteer speakers of Southern British English (3 male, 3 female). Recruitment of subjects and conduct of the experiment followed a protocol that was approved by the Oxfordshire Research Ethics Committee.

2. Method. Sequences of Magnetic Resonance Images were acquired at a rate of 4 frames/second (for four subjects) and 5 frames/s for another 2 subjects, as subjects were cued via an intercom to say either "R" or "E", in response to a prior sentence. Two intonation contours, a rising pitch and a falling pitch, were elicited by using prior sentences for which the letters are an affirmative reply (R! E!) or an echo-question (R? E?). These letter names are pronounced as long vowels, [ɑ] and [i]. The position of the base of the epiglottis (E) and of the front and rear of the first cervical vertebra (C1) were manually marked on every image (Figure 1). Data from one subject was marked independently by two annotators. A very high level of agreement was found between their measurements. The vertical distance between C1 and E indicates the height of the larynx. This distance was normalized by the z-transform, so that measurements from different subjects could be appropriately combined and compared.

3. Summary of results. In voiced speech, the larynx was found to be significantly lower immediately after the end of falling-pitch vowels, compared with rising-pitch vowels (Figure 2). The average difference is circa 3 mm, about 2% of the length of the vocal tract. There are significant differences between the vowels, with [ɑ] having a generally lower larynx position than [i]. This is probably due to the lower and more retracted tongue position of [ɑ] pushing the larynx down, through the tongue-root–hyoid bone connection.

In whispered speech, the larynx was also lower after falling-pitch vowels than rising-pitch vowels (Figure 3). Although the differences are not as large as in voiced speech, they are statistically significant for both vowels. Comparison of these final larynx positions with a baseline (the mean larynx position in non-speech) reveals that the larynx ends below the baseline in pitch falls and, except for whispered [ɑ], above the baseline in pitch rises.

Fig. 1. Measuring larynx position.

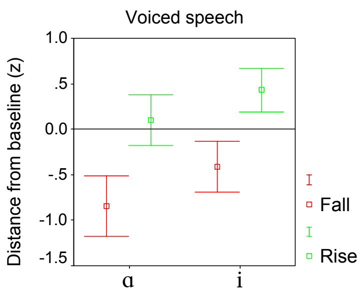

Fig. 2. Larynx position immediately after voiced vowels.

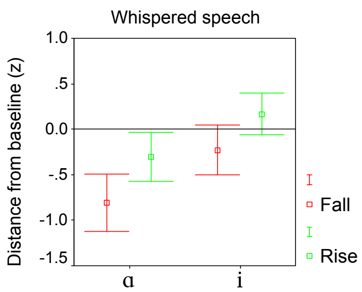

Fig. 3. Larynx position immediately after whispered vowels.

4. Discussion. Since the larynx movements responsible for pitch changes are comparable in voiced and whispered speech, we infer that the speech motor control plans are similar, at the expense of articulatory redundancy in whispered speech. This affords an explanation for perceived pitch in whispered speech: larynx movements change the tongue position and the shape of the oral cavity, thus altering the vocal tract acoustics so that the speech signal carries information that allows listeners to infer the articulatory movements normally associated with pitch changes.