Published in final edited form as:
Folia Phoniatr Logop. 2003 ; 55(1): 17–27.

Lip Kinematics for /p/ and /b/ Production during Whispered and Voiced Speech

Masahiko Higashikawa[a], **Jordan R. Green**[b], **Christopher A. Moore**[c], and **Fred D. Minifie**[c]

[a]Department of Otolaryngology, Osaka Medical College, Takatsuki, Japan

[b]Department of Communicative Disorders, University of Wisconsin-Madison, Madison, Wisc., USA

[c]Department of Speech and Hearing Sciences, University of Washington, Seattle, Wash., USA

Abstract

In the absence of voicing, the discrimination of 'voiced' and 'voiceless' stop consonants in whispered speech relies on such acoustic cues as burst duration and amplitude, and formant transition characteristics. The articulatory processes that generate these features of whispered speech remain speculative. This preliminary investigation examines the articulatory kinematics differences between whispered /p/ and /b/, which may underlie the acoustic differences previously reported for these sounds. Computerized video-tracking methods were used to evaluate kinematic differences between voiced and voiceless stops. Seven subjects produced the target utterances 'my papa puppy' and 'my baba puppy' in voiced and whispered speech modes. The results revealed that mean peak opening and closing velocities for /b/ were significantly greater than those for /p/ during whispered speech. No differences in peak velocity for either oral closing or opening were observed during voiced speech. The maximum distance between the lips for oral opening for /b/ was significantly greater than for /p/ during whisper, whereas no difference was observed during voiced speech. These data supported the suggestion that whispered speech and voiced speech rely on distinct motor control processes.

Keywords

Voiced speech; Whispered speech; Lip kinematics

Introduction

Cues for discriminating voiced and voiceless stop consonants in normally produced speech include voice onset time (VOT) [1], fundamental frequency [2], F_1 cutback [3], second formant transition [4], and amplitude of aspiration noise [5]. During whispered speech, however, contrasts resulting from changes in VOT and fundamental frequency are unavailable for distinguishing voicing cognates. The mechanism by which listeners

---

Masahiko Higashikawa, MD, Department of Otolaryngology, Osaka Medical College, Daigaku-cho 2–7, Takatsuki, Osaka 569-8686 (Japan), Tel. +81 726 83 1221, Fax +81 726 84 6539, masahiko@poh.osaka-med.co.jp.

discriminate whispered ‘voiced/voiceless’ pairs is unknown. Prior investigations of whisper have relied on measurements derived from the acoustic signal; none has evaluated speech movement. The present investigation was designed to evaluate differences in lip kinematics associated with the voicing contrast during whispered speech.

In the absence of an actual contrast in voicing, acoustic signals for voicing have been shown to include such features as burst duration and amplitude, and formant transition characteristics. Dannenbring [6] demonstrated increased noise burst duration for ‘voiceless’ compared to ‘voiced’ stops and the presence of a brief period of aspiration following the initial burst of ‘voiceless’ stops. Tartter [7] observed that the perceptual distinction of ‘voiceless’ consonants in consonant-vowel (CV) syllables relies on greater burst intensity and F_1 cutback. The F_1 transition is also of critical importance for the perception of ‘voicing’ in whispered stops [8]. Similarly, lower levels of high frequency noise in the following vowel are associated with ‘voiced’ stops [8]. Not surprisingly, overall speech intelligibility is relatively low during perception of whisper; the range for correct identification of ‘voiced’ consonants is from 50 to 80% for ‘voiced’ consonants [6-8].

The physiologic mechanisms that generate these acoustically distinct ‘voiced’ and ‘voiceless’ whispered stop consonants remain speculative. Whisper is characterized by lower tracheal pressure, higher translaryngeal flows, and lower laryngeal airway resistance [9]. Respiratory support does not appear to underlie these differences, as chest wall configuration during whisper is not significantly different from that observed for voiced speech production [9]. These observations imply a distinct coordinative organization of respiratory and laryngeal systems for whisper.

Tsunoda et al. [10] identified physiologic differences in laryngeal motor control for whispered ‘voiced’ and ‘voiceless’ stops and fricatives. Higher levels of posterior cricoarytenoid muscle activity were associated with ‘voiceless’ consonants. However, glottal aperture size, obtained nasoendoscopically, was not shown by these investigators to be significantly different for whispered minimal pairs. These findings suggest that there may be a reorganization of laryngeal control for distinguishing whispered ‘voiced’ and ‘voiceless’ consonants. Aerodynamic data do not help to explicate these apparent contradictions. Neither intraoral pressure nor the predicted glottal resistance was shown to be significantly different for whispered /p/ and /b/, although higher peak airflow rates were observed for /p/ than for /b/ [11]. Thus, aerodynamic and laryngeal EMG findings are equivocal, and we are led to evaluate other domains in search of the mechanisms that underlie these perceptually distinct phonemes. Recognizing this possibility, investigators have evaluated articulation, rather than laryngeal behavior, during whispered ‘voiced’ and ‘voiceless’ consonants [12, 13].

Schwartz [12] demonstrated significantly longer bilabial closure, measured using intra-oral pressure, during whisper. He also found longer closure durations for the whispered production of /p/ than for /b/. Schwartz [12] hypothesized that speakers conserve air by prolonging air-arresting articulatory gestures during whisper. Parnell et al. [13] observed greater durations for /s/ and /t/ than for /z/ and /d/ during whispered speech using spectrographic analysis. They also showed greater duration for the vowels /i/ and /a/ for whisper. Parnell et al. [13] rejected Schwartz's hypothesis, alternatively proposing that, during whisper, speakers enhance intelligibility by decreasing speech rate.

Peak intraoral pressures are comparable for production of whispered /p/ and /b/ [11, 14, 15], whereas during voiced production peak intraoral pressures are observed to be higher for /p/ than for /b/ [14-17]. These intraoral pressure differences for /p/ and /b/ during voiced speech result from putative differences in airway resistance [16, 17], which may arise from differences in lip articulation or position (i.e., greater lip force, higher lip closing velocity, to maintain lip contact during closure interval). Lubker and Parris [18] failed to observe such differences in bilabial pressure, however, for normally produced /p/ and /b/ using a miniature pressure transducer, even though they did find significantly higher intraoral air pressures for /p/ than for /b/. This evidence suggests that greater interlabial compression is not necessarily required for increased intraoral pressure. Thus, the few physiologic studies of whispered consonants have not revealed the mechanisms for differentiating production of 'voiced' and 'voiceless' consonants during whispered speech. Direct measurements of lip and jaw kinematics during whispered speech, which might resolve the inconsistencies of earlier results, have yet to be completed.

Lip and jaw kinematics during voiced production of bilabial consonants have shown repeatedly that lip and jaw closing velocities for stops tend to be faster for the voiceless cognate [19-21]. Gracco [21] suggested that these velocity differences are a natural consequence of the timing requirements for the different consonants. However, in one of the most detailed investigations of these kinematic features, Löfqvist and Gracco [22] reported no consistent differences in lip and jaw closing velocities or in lip aperture changes for normally produced voiced and voiceless bilabial stops. These inconsistencies in lip and jaw kinematics during bilabial closure remain to be resolved for both whispered and voiced speech.

Reliable identification of voicing contrasts in whisper [6-8] entails nonlaryngeal, though distinct, speech movement. At least two types of mechanisms can be posited for differentiation of 'voicing' for whispered stops. It may be, for example, that whispered speech reveals latent properties of voiced speech motor control, which become more apparent in the absence of other highly salient cues used in voiced speech. That is, without the highly salient voicing cues afforded by the glottal source, whisper may rely more heavily on secondary cues, thereby exposing them to experimental observation. This type of mechanism is consistent with the observation that acoustic cues for whispered speech are absent in voiced speech. Alternatively, motor control for whispered speech production may be distinct from voiced speech and responsive to vastly different communicative conditions and physiologic limitations. The present investigation was designed to evaluate these differences in whispered and voiced speech. Differences in speech motor control were inferred from observed differences in lip kinematics during opening and closing gestures in the production of whispered 'voiced' and 'voiceless' consonants.

Method

Seven male subjects (mean age: 27.0 years; range: 21–37 years) participated. All subjects were native speakers of American English, and had normal speech and hearing, and negative histories of neurological disorder. Subjects produced two blocks of 40 target utterances each embedded in the carrier phrase ‘my ——— puppy’ (i.e., ‘my *papa* puppy’ and ‘my *baba* puppy’). The carrier phrase included the vowels /ai/ and /a/ adjacent to the target bilabial gestures (i.e., for /p/ and /b/) to reveal more clearly opening and closing movement.

A block consisted of 20 randomized repetitions of each of the two target phrases using either voiced or whispered speech during the entire block. The order of speech mode (i.e., whispered or voiced speech) was randomized across blocks. Subjects were instructed to produce the utterances at a comfortable rate of speech and at a normal loudness. The investigator demonstrated syllable stress to be on the first syllable in each target word. The interval between utterances was approximately 2 s and was paced by the experimenter’s cue.

A sample of sound pressure levels across entire trials for 4 subjects (subjects B, C, D, and G) confirmed the perceptual judgment by the investigator that subjects were successful in maintaining a relatively constant loudness level within each block. Sound pressure level, monitored with a sound level meter (Quest 1400, Quest Technologies) 20 cm from the subject’s lips, revealed speech levels of 68–72 dB SPL for voiced speech and 55–60 dB SPL for whispered speech. No systematic difference in sound pressure level was observed between /p/ and /b/ utterances.

The setup for data recording, digitizing, and signal conditioning was done based on the report by Green et al. [23]. Acoustic and video recordings were obtained in a large sound-treated booth located in the Laboratory of Speech and Hearing Sciences, University of Washington. A research quality wireless microphone (Telex, FMR-25) was attached to the shirt collar of each subject (approximately 20 cm from the subject’s mouth). The audio signal was recorded using a digital audio recorder (Panasonic, SV-3700). The video signal, obtained using an infrared-sensitive video camera (Burle, TC351A), was recorded using a videocassette recorder (Panasonic, AG-1980). The video image was subsequently digitized for computer-based tracking of lip and jaw movement. Distortion due to the shape of the camera lens was minimized by centering the video frame on the subject’s face and by using the maximum zoom setting on the camera.

Automatic tracking of videotaped movements relied on reflective markers, which were placed on the midline of the vermilion border of both the upper lip (UL) and the lower lip (LL), and just superior to the protuberance of the mandible (J). Two reference markers were placed on the nasion and the apex of the nose to isolate and provide correction for extraneous head movement. Length-calibration markers were placed 4 cm apart on the forehead. An infrared (i.e., invisible) light source was used to illuminate the markers (~2 mm in diameter).

The video segments containing target words were parsed from the continuous videorecording and were digitized (60 frames/s) for computer-based tracking of the reflective markers. Vertical position of the UL, LL, and J were obtained automatically using

Motus, a commercially available movement tracking system (Motus, version 2, Peak Performance Technologies, Inc.). Translation and rotation algorithms corrected for extraneous head movements prior to digital filtering ($f_{lowpass}$ = 30 Hz) and quantitative analysis. We have determined that the spatial resolution of this system under our recording conditions is better than 100 μm over a vertical range of 8 mm.

Analysis of lip kinematics was completed using custom routines written for Matlab (Mathworks, 1993), a commercially available signal processing package. The vertical separation of the lips at midline was obtained by subtracting the position of LL from that of UL. Segment boundaries [i.e., between the beginning point of closing (arresting) gesture for first ‘p’ and the ending point of opening (releasing) gesture for following vowel of first ‘p’] are identified in the shaded portion of figure 1. Operationally these points were identified by computing the first derivative of the lip separation waveform (i.e., change in vertical separation of the lips at midline over time, equation 1) and locating the negative-going zero crossings closest to the closing gesture associated with the first /p/ in the production of /papa/. First derivative functions for each of the articulator displacement functions (i.e., UL, LL, J) were also computed; a custom algorithm was applied to each of these functions to confirm the presence of zero-crossings coincident with that identified for the lip separation waveform, dy/dt. Utterances that did not exhibit all eight of these zero-crossings (i.e., at the onset and offset of the syllable production for each of the four derivative functions) were excluded from further analysis.

The following parameters were obtained from the lip separation waveform:

$$Rate\ of\ change\ in\ vertical\ separation\ of\ lips\ at\ midline = dy/dt \qquad \text{Eq. 1}$$

- **a.** Minimum of dy/dt (labeled ‘a’ in fig. 1), which was identified as the moment of most rapid bilabial closure in the production of the target syllable.
- **b.** Maximum of dy/dt (labeled ‘b’ in fig. 1), which was identified as the moment of most rapid bilabial opening in the production of the target syllable.
- **c.** Maximum vertical separation of the lips (y_{max}; labeled ‘b’ in fig. 1) during oral opening; identified as the peak value in the lip separation signal within 50 ms of the zero crossing in dy/dt associated with the opening gesture at the end of the first syllable in /papa/.

Paired t tests ($\alpha = 0.05$) were used to analyze differences among subjects’ means for peak oral closing velocity (PVC), peak oral opening velocity (PVO), and maximum oral opening (MLSO) for combined /p/ and /b/ productions for voiced and whispered speech. Separate paired t tests ($\alpha = 0.05$) were also used to evaluate differences (i.e., in PVC, PVO, and MLSO) in productions of voiced and whispered productions of /b/ versus /p/.

Results

All subjects accepted into the subject pool successfully completed the experimental conditions. The final data set of 433 utterances included 67, 69, 77, 42, 65, 67, and 46 samples for subject A through G, respectively. Distributions of whispered and voiced productions of /pa/ and /ba/ are shown for each subject in table 1.

Data Inclusionary Criteria

Speech tokens were included only if zero crossings were clearly identifiable at the onset and offset of the gesture in the derivative functions of each of the individual movement functions (i.e., UL, LL, and J) and the derivative function of the lip separation trace (i.e., dy/dt). This factor was especially important with respect to those subjects who exhibited minimal upper lip movement during speech.

Oral Closing

Figure 2a, b show each subject’s means for peak closing velocities during lip closure for /p/ and /b/ during voiced and whispered for /p/ and /b/ for PVC. Phonemic effects are shown in figure 2c in which the difference between the two speech tokens (i.e., ΔPVC, equation 2) is shown. For each subject and for each voicing condition (i.e., voiced and whispered):

$$\Delta peak\ closing\ velocity = peak\ closing\ velocity_{/\mathrm{b}/} - peak\ closing\ velocity_{/\mathrm{p}/} \quad \text{Eq. 2}$$

As seen in figure 2a–c, phoneme effects were quite irregular in magnitude and direction across subjects. Higher values of PVC were observed during voiced productions of /b/ in 4 subjects (B, E, F, and G; fig. 2a) and higher values for /p/ in 3 subjects (A, C, and D). These differences were not statistically significant. During whispered speech, 6 of the 7 subjects (subject E was the exception) exhibited higher PVC values for /b/ than for /p/ (fig. 2b). This difference was statistically significant [$t(6) = 4.171$, $p < 05$]. As observed in the voiced speech conditions, results of the comparison of ΔPVC for voiced and whispered speech revealed that 4 subjects (A, B, C, and D) exhibited smaller differences for voiced than for whispered speech; the remaining 3 subjects exhibited larger differences for voiced speech (fig. 2c). As one might anticipate, there was no statistical difference for ΔPVC in voiced and whispered speech conditions.

Oral Opening

Results of the analysis of peak velocities of lip opening are shown in figure 3. This figure includes mean peak velocities for each subject in each phoneme condition (i.e., /p/ versus /b/) and in each voicing condition (i.e., voiced versus whispered). Higher values of PVO during voiced speech production for /b/ were observed in 6 subjects (A, B, D, E, F, and G); a lower mean PVO for /b/ was seen only for subject C (fig. 3a). However, these differences were not sufficiently large to be statistically significant. In contrast, the results for the analysis of the phoneme effect during whispered speech did yield a statistically significant difference [$t\ (6) = 2.922$, $p < 0.05$]. Six of the 7 subjects (A, B, C, D, F, and G) exhibited higher PVO values for /b/ than for /p/ (fig. 3b). Similarly the same 6 subjects generated higher ΔPVO values for whispered speech than for voiced speech. This difference in ΔPVO between voicing conditions was statistically significant [$t\ (6) = 2.534$, $p < 0.05$; fig. 3c].

Maximum lip separation during the opening gesture (MLSO) was greater for /b/ than for /p/ during voiced speech for 5 subjects (A, B, D, F, and G; fig. 4a), although this phoneme effect was not statistically significant. During whispered speech, however, all subjects exhibited this pattern of greater MLSO for /b/ than for /p/ (fig. 4b). This difference was statistically significant [$t(6) = 5.072$, $p < 0.05$]. The difference in MLSO between voiced and whispered speech (fig. 4c) was also statistically significant [$t(6) = 4.864$, $p < 0.05$].

Discussion

The present results demonstrated distinct patterns of lip movement for the production of bilabial stops in voiced and whispered speaking modes. These differences were most apparent in measures of peak velocity of lip opening (PVO) and in the maximum lip separation during the opening phase (MLSO) following release of the bilabial stop. These findings supported the suggestion that the motor control processes underlying whispered bilabial stops are different from those of voiced speech. Furthermore, we speculate that the differences in lip aperture are likely to have consistent and perceptible acoustic effects.

The data obtained using this video-based tracking method yielded no significant differences in the PVC or PVO between /p/ and /b/ during voiced speech production. This finding was in agreement with recent reports on lip opening and closing [for lip opening only: ref. 21; for lip closing and opening: ref. 22], but failed to support earlier investigations in which greater articulatory closing velocities for voiceless stops were observed to exceed those for voiced stops during voiced speech production [19-21]. These inconsistencies might be attributed to several factors, including: the small numbers of subjects observed in each investigation, methodological differences (i.e., phonetic context, targeted articulators, equipment used for measurement), and the anticipated ubiquitous variability of speech, including kinematic variability within and among speakers.

The present results represent the first quantitative description of both lip opening and closing velocities during whispered speech. PVC, PVO, and MLSO during whispered bilabial stops were each statistically greater for production of /b/ than for /p/. These findings were in sharp contrast with the bulk of earlier investigations of voiced speech articulation, most of which failed to find significant kinematic differences between voicing cognates. One interpretation of these results might rely on the tacit assumption that oral closing is associated with production of consonants (‘closants’) [24], whereas oral opening is associated with vowel production (‘vocants’) [21, 24]. It may be, for example, that higher closing velocities (PVC) enhance the acoustic contrast between bilabial stops. Similarly, higher opening velocities (PVO) and larger lip openings (MLSO) may enhance the acoustic contrast of bilabial stop consonants as well as the succeeding vowels.

Curiously, the consistent patterns of peak lip closing velocities during whispered speech observed in the present study were opposite to some of the reported articulatory velocities associated with voiced speech. Gracco [21] suggested that movement velocity changes during voiced speech are a natural consequence of the timing requirements for the different consonants. The present results have demonstrated that articulatory parameters of whispered bilabial consonants vary consistently with voicing contrasts. It may be that these alternations provide acoustic cues that enhance the discrimination of whispered /p/ and /b/; faster peak lip closing velocities may represent changes in the coordinative organization of stop consonant production. It may be that, in the impoverished acoustic conditions presented by whispered speech, faster opening velocities give rise to perceptual differences that distinguish /b/ from /p/. These cues may be irregular and of secondary importance during normally voiced speech, but assume a primary role in the absence of voice. The finding that ΔPVC was not significantly different for voiced and whispered speech further supports the notion that the changes observed in whispered speech are active alterations in speech motor control. Because the aerodynamic conditions of whispered and voiced speech are so different, passive changes would be expected to yield differences in PVC between voiced and whispered conditions. This difference was not observed.

Aerodynamic conditions, including airflow, transglottal pressure, and intraoral pressure, have frequently been shown to differ for stop consonant voicing cognates during normally voiced speech. Some of these differences have been shown to persist in whispered speech. Weismer and Longstreth [11] showed higher peak airflow for /p/ than for /b/ during whispered speech. Peak airflow is typically observed immediately following the burst release, and is proportionately greater for narrower vocal tract constrictions. Given the present observations of slower lip opening velocity for /p/ during whisper, one would anticipate momentarily higher airflow sufficient to enhance the turbulence at the lip opening relative to that associated with production of /b/. This proposition is supported by findings of increased noise burst duration for ‘voiceless’ stops during whisper [6] and increased burst intensity for ‘voiceless’ stops [7]. Complementarily, the rapid peak opening velocity for /b/ would be expected to yield comparatively brief turbulence at the lips and would likely alter the spectral tilt of the noise associated with the shortened noise burst. These factors could each contribute to the acoustic distinction of the whispered /b/.

Investigations of speech spectra further contribute to resolving the question of how voicing can be cued during whispered speech. Relatively low power levels of high frequency noise components have been observed during whisper for /ba/ than for /pa/. These observations were made of power spectra of the first 50 ms of the vowels in whispered CV syllables [8]. The motor organization underlying this contrast might, for example, mediate increased widening of the lip aperture following /b/ with respect to /p/. This increased lip aperture will reduce turbulence and reduce the power of high frequency noise. The present findings of greater lip separation at opening (MLSO) for /b/ than for /p/ support the operation of such a mechanism in generating the acoustic cues essential to discrimination of /p/ and /b/ during whispered speech.

We proposed two explanations for how speakers distinguish ‘voiced’ and ‘voiceless’ bilabial stops during whispered speech. It may be that latent mechanisms that underlie normally voiced speech production assume primary importance during whispered speech. Burst characteristics, which have been shown to vary systematically during whisper, may be subordinate to other, more potent features of voicing cognates; the most obvious feature, of course, is VOT. A second interpretation of the present results is that whisper entails a separate mode of speech motor control for bilabial stops. This proposition is far less parsimonious and, given the redundancy in speech motor control systems, far less likely.

Anecdotally, we can cite the observation that virtually all adult speakers are able to whisper, although proficiency is attained considerably later than normally voiced speech. This evidence may be taken to suggest that children must learn a new speech mode comprised of new physiologic phonetic sequences, or that learning to whisper entails suppression of extant motor patterns associated with voiced speech. Resolution of these competing models of the motor organization of whisper will depend on additional comparisons of other speech articulators (e.g., the tongue [24]; the lateral wall [25]) and a wider variety of phonemes during voiced and whispered modes of speaking.

Conclusion

Kinematic differences in lip movement for whispered and normally voiced bilabial stop consonants were observed in a group of 7 normal adult speakers of American English. Distinct productions of the bilabial voicing cognates /b/ and /p/ (i.e., ‘voiced’ and ‘voiceless’) were analyzed using a video-based computerized movement tracking system. Measures included high resolution tracking of the vertical separation of the lips at midline, as well as displacement and velocity functions for upper and lower lips. Analyses focused on the closing phase and the opening phase in the utterances ‘baba’ and ‘papa’.

The results revealed that the mean peak opening and closing velocities for /b/ were significantly greater than those for /p/ during whispered speech. No differences in peak velocity for either oral closing or opening were observed during voiced speech. Also, the maximum distance between the lips for oral opening for /b/ was significantly greater than for /p/ during whisper, whereas no difference was observed during voiced speech. These data supported the suggestion that whispered speech relies on a motor organization that is either altered or distinct from that used during normally voiced speech.

References

1. Lisker L, Abramson AS. A cross-language study of voicing in initial stops: Acoustical measurements. Word. 1964; 20:384–422.
2. Haggard M, Ambler S, Callow M. Pitch as a voicing cue. J Acoust Soc Am. 1970; 47:613–617. [PubMed: 5439661]
3. Liberman AM, Delattre PC, Cooper FS. Some cues for the distinction between voiced voiceless stops in initial position. Lang Speech. 1958; 1:153–167.
4. Stevens KN, Klatt DH. Role of formant transitions in the voicedvoiceless distinction for stops. J Acoust Soc Am. 1974; 55:653–659. [PubMed: 4819867]
5. Repp BH. Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. Lang Speech. 1979; 22:173–189. [PubMed: 513908]
6. Dannenbring GL. Perceptual discrimination of whispered phoneme pairs. Percept Mot Skills. 1980; 51:979–985. [PubMed: 7208245]
7. Tartter VC. What’s in a whisper? J Acoust Soc Am. 1989; 86:1678–1683. [PubMed: 2808917]
8. Higashikawa M. Perceptual, acoustical aerodynamic study of whispering. Nippon Jibiinkoka Gakkai Kaiho. 1994; 97:1268–1280. [PubMed: 8064511]
9. Stathopoulos ET, Hoit JD, Hixon TJ, Watson PJ, Solomon NP. Respiratory laryngeal function during whispering. J Speech Hear Res. 1991; 34:761–767. [PubMed: 1956183]
10. Tsunoda K, Niimi S, Hirose H. The roles of the posterior cricoarytenoid thyropharyngeus muscles in whispered speech. Folia Phoniatr Logop. 1994; 46:139–151. [PubMed: 8019596]

11. Weismer G, Longstreth D. Segmental gestures at the laryngeal level in whispered speech: Evidence from an aerodynamic study. J Speech Hear Res. 1980; 23:383–392. [PubMed: 7442198]
12. Schwartz MF. Bilabial closure durations for p, b, m in voiced whispered vowel environments. J Acoust Soc Am. 1972; 51:2025–2029. [PubMed: 5045258]
13. Parnell M, Amerman JD, Wells GB. Closure constriction duration for alveolar consonants during voiced whispered speaking conditions. J Acoust Soc Am. 1977; 61:612–613. [PubMed: 845324]
14. Murry T, Brown WS Jr. Peak intraoral air pressure in whispered stop consonants. J Phonetics. 1976; 4:183–187.
15. Klich RJ, Sabo WJ. Intraoral pressure differences in various degrees of breathiness. Folia Phoniatr. 1988; 40:265–269.
16. Isshiki N, Ringel R. Air flow during the production of selected consonants. J Speech Hear Res. 1964; 7:233–244. [PubMed: 14213499]
17. Arkebauer HJ, Hixon TJ, Hardly JC. Peak intraoral air pressure during speech. J Speech Hear Res. 1967; 10:196–208. [PubMed: 6082347]
18. Lubker JF, Parris PJ. Simultaneous measurements of intraoral pressure, force of labial contact, labial electromyographic activity during production of the stop consonant cognates /p/ /b/. J Acoust Soc Am. 1970; 47:625–633. [PubMed: 5439663]
19. Sussman HM, MacNeilage PF, Hanson RJ. Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. J Speech Hear Res. 1973; 16:397–420. [PubMed: 4744959]
20. Fujimura O, Miller JE. Mandible height and syllable-final tenseness. Phonetica. 1979; 36:263–272. [PubMed: 523517]
21. Gracco VL. Some organizational characteristics of speech movement control. J Speech Hear Res. 1994; 37:4–27. [PubMed: 8170129]
22. Löfqvist A, Gracco VL. Lip and jaw kinematics in bilabial stop consonant production. J Speech Lang Hear Res. 1997; 40:877–893. [PubMed: 9263951]
23. Green JR, Moore CA, Higashikawa M, Steeve RW. The physiologic development of speech motor control: Lip and jaw coordination. J Speech Lang Hear Res. 2000; 43:239–255. [PubMed: 10668666]
24. Raymond DK, Harold RB. Vocalizations of one-year-olds. J Child Lang. 1985; 12:491–526.
25. Svirsky MA, Stevens KN, Matthies ML, Manzella J, Perkell JS, Wilhelms-Tricarico R. Tongue surface displacement during bilabial stops. J Acoust Soc Am. 1997; 102:562–571. [PubMed: 9228817]
26. Minifie FD, Hixon TJ, Kelsey CA, Woodhouse RJ. Lateral pharyngeal wall movement during speech production. J Speech Hear Res. 1970; 13:584–594. [PubMed: 5507116]

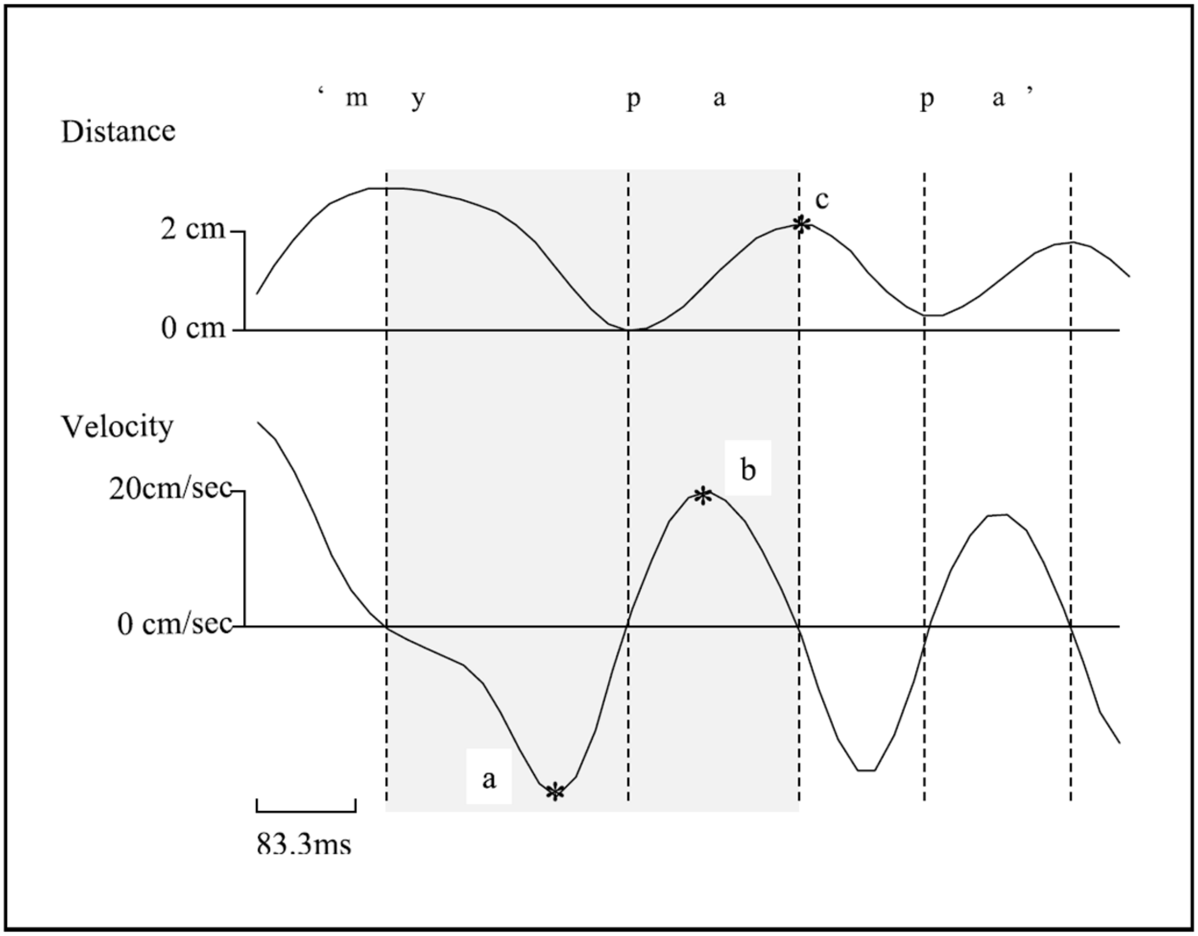

Fig. 1.

Kinematic signals during production of ‘my papa’ in whispered speech by subject C. The upper signal represents the distance of vertical separation of the lips at midline and the lower waveform represents the derivative of the upper signal. The vertical lines show the zero crossing of the derivative waveform. Minimum of dy/dt (labeled ‘a’) and maximum of dy/dt (labeled ‘b’) were obtained within the shaded portion. Maximum vertical separation of the lips (y_{max}; labeled ‘c’) during oral opening was identified as the peak value in the lip separation signal within 50 ms of the zero crossing in dy/dt associated with the opening gesture at the end of the first syllable in /papa/.

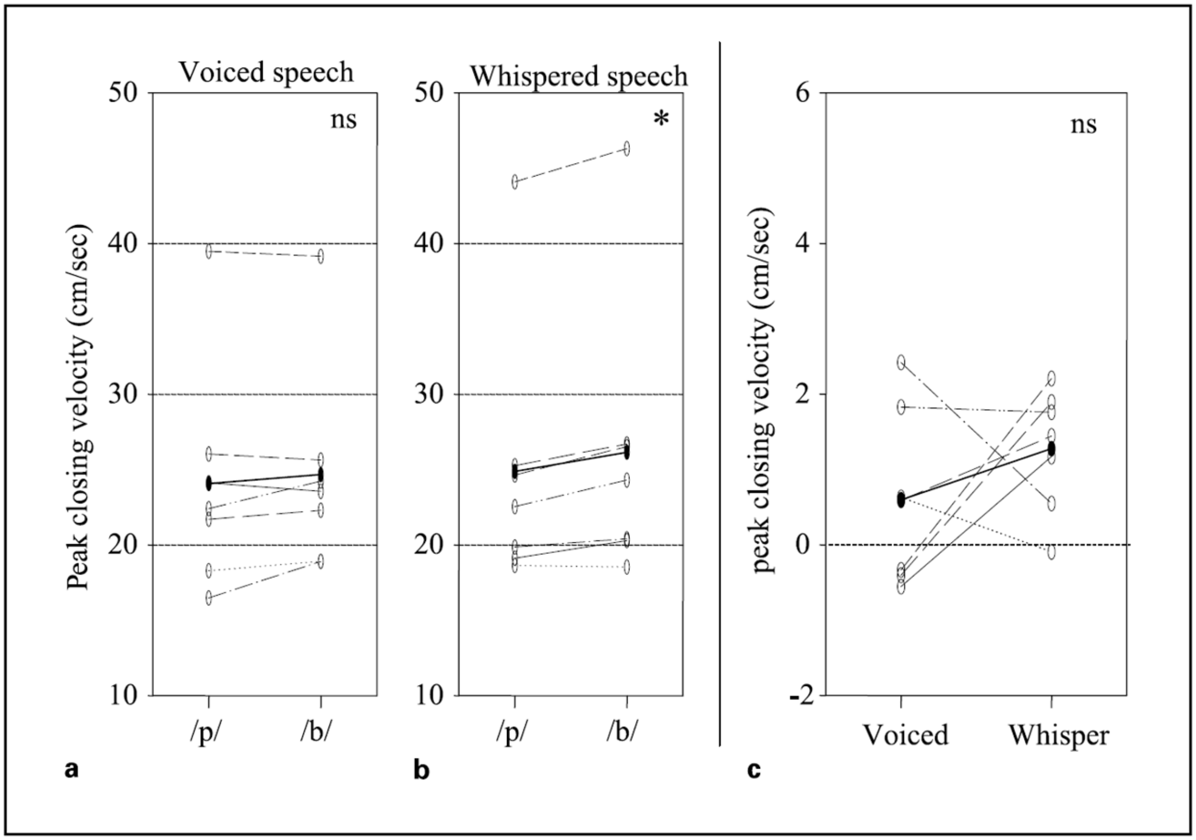

Fig. 2.

PVC (**a, b**) and ΔPVC (**c**). The open plots represent means of each subject and filled plots represent means of subjects' means. ΔPVC show the difference of PVC between /p/ and /b/ (i.e., PVC for /b/ minus PVC for /p/). * $p < 0.05$; ns = not statistically significant.

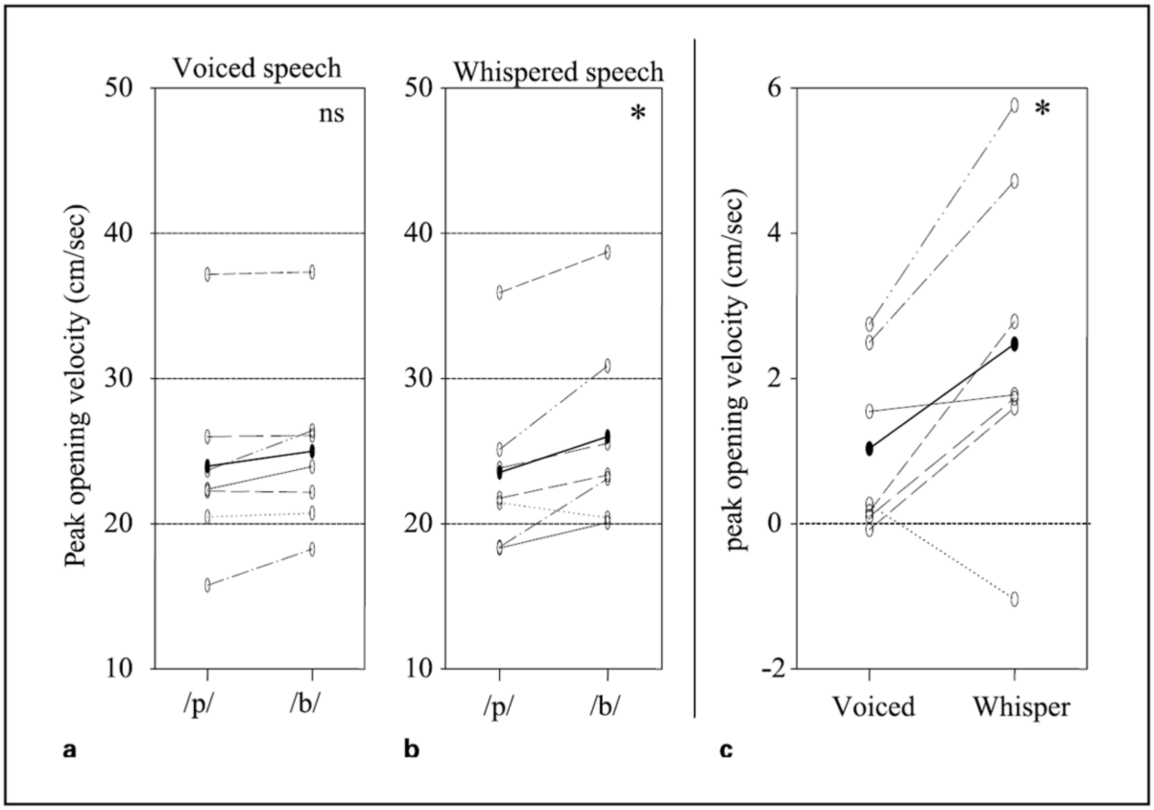

Fig. 3.

PVO (**a, b**) and ΔPVO (**c**). The open plots represent means of each subject and filled plots represent means of subjects' means. ΔPVO show the difference of PVO between /p/ and /b/ (i.e., PVO for /b/ minus PVO for /p/). * $p < 0.05$; ns = not statistically significant.

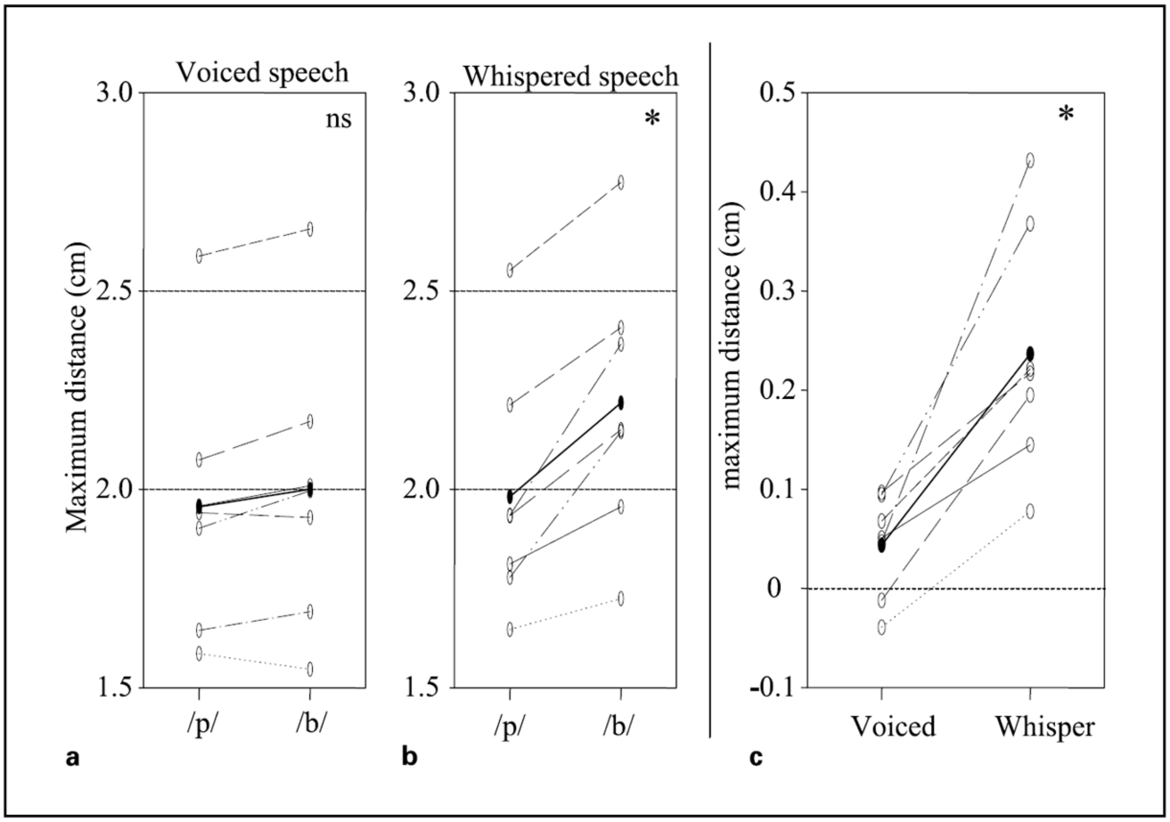

Fig. 4.

MLSO (**a, b**) and ΔMLSO (**c**). The open plots represent means of each subject and filled plots represent means of subjects' means. ΔMLSO show the difference of MLSO between /p/ and /b/ (i.e., MLSO for /b/ minus MLSO for /p/). * $p < 0.05$; ns = not statistically significant.

Table 1

Distributions of whispered and voiced productions of /pa/ and /ba/ for each subject

Subject	Total	Voiced speech		Whispered speech	
		/pa/	/ba/	/pa/	/ba/
A	67	15	16	20	16
B	69	15	16	19	19
C	77	19	20	20	18
D	42	11	13	8	10
E	65	16	14	18	17
F	67	18	15	16	18
G	46	10	13	12	11