Measuring oral and nasal airflow in production of Chinese plosive

Conference Paper · September 2015

5 authors, including:


Yujie Chi
Tianjin University


Jianguo Wei
Tianjin University

Hui Feng
Tianjin University


Jianwu Dang
Japan Advanced Institute of Science and Tec…

Measuring oral and nasal airflow in production of Chinese plosive

Yujie Chi[1], *Kiyoshi Honda*[1], *Jianguo Wei*[1,*], *Hui Feng*[1], *Jianwu Dang*[1,2]

[1] Tianjin Key Laboratory of Cognitive Computation and Application, Tianjin University, Tianjin, China

[2]School of Information Science, JAIST, Ishikawa, Japan

yujiechi@qq.com, khonda@sannet.ne.jp

Abstract

This study reports a new integrated device for measuring oral and nasal airflow with lossless speech recording with preliminary results on Chinese plosives. An acoustically transparent fiber-made mask acts both as a material for airflow resistance and a support for partitioning oral and nasal chambers. Two pressure sensors placed directly on the mask measure air pressure signals that correspond to the airflow rate. With this device, oral-nasal airflow rate and speech sound were recorded from two Chinese speakers. The corpus was made of two-syllable Chinese words (CVCV, C is a plosive) in a carrier sentence. Slow and rapid variations of nasal airflow signals were obtained with clear speech and oral airflow data, which were analyzed to speculate the elevation and vibration of the velum. The results show that (i) AC component of the nasal airflow is greater in /i/ than in /a/, which supports transvelar nasal coupling during high vowels and obstruents; (ii) word-medial plosives tend to be partially or fully voiced, as inferred from augmented AC amplitude of the nasal airflow during the closure; (iii) short rise or drop of the nasal airflow appears to reflect a quick elevation or decline of the velum.

Index Terms: fiber mask, oral-nasal airflow, plosive, nasal resonance, velum

1. Introduction

Simultaneous measurement of oral and nasal airflow data during speech production is not only meaningful for examining velopharyngeal functions, but also effective to understand articulation and resonance during vowels and consonants. This is because oral and nasal airflow reflects various events during speech production, such as closure and release timing for consonants, subtle rise and fall of the velum during oral sounds, etc. [1]. Further, those events could be more detailed when high-quality speech signals were recorded simultaneously [2] [3]. This study is aimed at realizing such combined measurements by resolving technical difficulties for data acquisition. To do so, a nasal-oral pneumotachography system was built using an acoustically transparent mask, and it was tested by analyzing Chinese vowels and plosive consonants in speech utterances.

The technique for simultaneously measuring oral-nasal airflow was developed as early as 1964 by Lubker & Moll using a two-chamber mask covering on the mouth and nose [4]. Combined with cinefluorographic images of velopharyngeal movements, the work demonstrated nasal airflow caused by rapid elevation of the velum at velopharyngeal closure. However, simultaneous recording of clear audio was not attained because the rigid mask to measure airflow distorts output speech sound. A possible solution to this problem is to develop a dual-chamber airflow mask that reduces acoustic distortion while keeping enough sensitivity.

Rothenberg invented a “dual chamber circumferentially vented mask” for measuring oral-nasal airflow [5]. The vented mask has relatively low acoustic impedance and was used for quantifying nasal resonance problems [6]. While the Rothenberg mask supports low-distortion audio recording, a few problems have been noted for the mask: spectral attenuation near 2 kHz caused by the single-chamber vented mask [7], and leakage of sound resulted from the vibration of the internal chamber separator [5]. More recently, researchers in France developed a new pneumotachography with a mask of synthetic fibers [3]. This technique solved the problem of spectral distortion due to the mask on the face. They attempted to build a dual-chamber type of the mask, but the hand crafting was difficult and time-consuming. Our work is based on their mask-type pneumotachography with an aim to further advance the device for phonetics studies.

We describe below the details of our dual-chamber pneumotachography system that has sufficient stability and sensitivity for airflow data acquisition, which is followed by preliminary results to document nasal resonance and velar elevation during production of Chinese plosives.

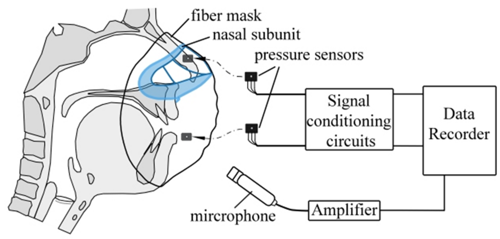


Figure 1: *Nasal-oral pneumotachography system.*

2. System description

As shown in Figure 1, the nasal-oral pneumotachography system is composed of two new components: a fiber mask with a nasal subunit inside to separate two chambers; two pressure sensors output to signal conditioning circuits and a data recorder. The speech signal is recorded by an electret condenser microphone on another channel of the data recorder. The pressure sensors and the nasal subunit are simply detachable from the mask without altering the mask structure.

2.1. Oral and nasal mask

The fiber mask was chosen to obtain lossless speech sound when it is worn on the face. The acoustic effect of the mask

*Corresponding author

was examined by analyzing the spectral envelopes of the speech signal (vowel /a/ produced by a male speaker) with and without the mask. Figure 2 supports that spectral distortion caused by wearing the fiber mask is minimal. Other advantages of the mask are lightweight, easily worn, and disposable as a dust mask.

To separate the oral and nasal air passages, the nasal subunit was crafted by modifying an anti-pollution nasal mask (MAIXINGREN, WH-KN A-1). The mesh of the mask was removed, and a rubber plate was attached to the plastic support of the mask as a chamber separator (see Figure 3). This nasal subunit is attached inside the mask with double-side tape, while the elastic fringe covers the nose hermetically. This type of fiber mask was chosen among a few types that have a sufficient cavity size to keep the nasal subunit inside. The chamber separator rests above the upper lip, similar to the masks in previours work, but our mask guarantees the stability of the separator and minimum sound leakage between the chambers.

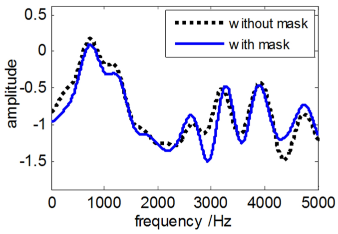


Figure 2: *Comparison of spectral envelopes in vowel /a/ recorded with and without mask.*

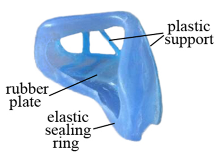


Figure 3: *Nasal subunit modified from a commercial nasal mask.*

2.2. Pressure sensor

A high sensitivity pressure sensor of a miniature type is used in this study to realize pressure-difference anemometry. An integrated gage sensor (HSCSAAN001NDAA5, Honeywell) was chosen because of its ultra-low pressure range (2.5 cmH_2O full scale, 2% FSS error band) and the small size (10 mm×10 mm). Two identical sensors are used for both oral and nasal channels, and they are attached directly on the fiber mask. In addition, the sensor owns a compensated temperature ranging from 0 to 50℃ and the stability at positioning in any orientations. The output frequency bandwidth is practically limited to 0 to 500 Hz, presumably due to the unique electronics built into the sensor unit [8].

2.3. Signal conditioning

The output voltage of the pressure sensor ranges from 0.5 to 4.5 V (zero pressure at 2.5 V) when supplied by 5V power. An external signal conditioner unit includes a level shifter circuit to convert the output voltage to fluctuate around 0 V. Then, amplifier and filter circuits convert the signals into the range of -8 ~ +8V. For the oral channel, the amplifier gain was set at about 30, and for the nasal channel it was about 90.

The outputs from the signal conditioner unit for the nasal and oral airflow channels and another output from a microphone amplifier were connected to a four-channel data recorder (DAS 40, Sefram) to sample and store the data.

2.4. Calibration procedure

The relation between the voltage value (U) and the airflow rate (Q, the target value for our result) is simplified as

$$U = R_m \times Q \qquad (1)$$

where R_m is the coefficient that is determined by airflow resistance of the mask and gain of the amplifier. Usually, the equation is integrated on both sides to obtain

$$R_m = \int_{t_0}^{t_1} U \div V \qquad (2)$$

where V is the total volume of air during t_0 and t_1. After the calibration procedure, R_m is calculated from the output voltage of the circuits and the known volume of air.

The oral and nasal airflow rate must be calibrated separately to obtain each coefficient R_m. Figure 4 shows our calibration device composed of four parts: a rigid plastic board; a rubber ring to fix and seal the mask’s outer edge; a plastic nose model with a nostril-like hole to fit the nasal subunit; and two air inlets. Two syringes of 200 ml (value of V in (2)), linked to the air inlets, provide the constant volume of air. During the calibration, the oral-nasal mask was placed on the rigid board hermetically to avoid air leakage; each syringe was pushed uniformly from 200ml to 0ml, which was repeated ten times. Finally, waveform data on a PC was analyzed to calculate averages of the integrated voltage values obtained by the syringe operations for each channel. The coefficients (R_m) obtained were 1.9 V/liter/s for the oral airflow rate and 16.9 V/liter/s for the nasal airflow rate.

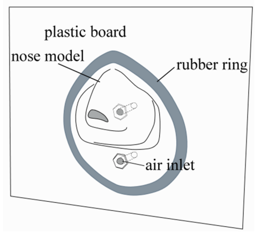


Figure 4: *Device for airflow calibration.*

3. Experiment

The experiment was planned to observe oral and nasal airflow patterns during production of Chinese plosives. This is because our preliminary airflow recordings suggested that nasal airflow exhibits nasal resonance and velar movement, while oral airflow marks closure and release for consonant articulation. Therefore, the aim is set at airflow-based characterization of aspirated and unaspirated plosives in Chinese.

3.1. Corpus and speakers

The corpus is made of Chinese disyllabic words with CVCV structure. The consonant C is one of plosives /p/, /t/, /k/, /p’/, /t’/ and /k’/; the first vowel V is /a/ or /i/ to have typical open and close vowels. The possible combinations of CVC were examined for preparation to compose possible disyllabic words. Considering the limited occurrences of /ki/, /k’i/, and /ka/ in spoken Chinese, 57 of the CVC sequences were chosen to build the corpus. Then, common and simple Chinese disyllabic words were selected to have syllables with balanced tones (marked by 1, 2, 3, 4 and 0 for the tones: high level, high rising, low dipping, high falling, and neutral). The second vowel was not controlled. Examples are /t’a^1 kê^1/ (他哥, “her brother”), /t’i^3 p’o^4 / (体魄, “body build”) and /ti^2 ku^0/ (嘀咕, “whisper”).

Two subjects, A (female, age 24) and B (male, age 23), are both from the northern China and speak standard Chinese naturally. The subjects read five words each time that were embedded in a carrier sentence as shown below. The test words were presented by the characters.

再读 ___ 一遍.

/tsai du ___ i pian/

"Say ___ again." in English

3.2. Experimental procedure

The experiment was conducted in a soundproof room with the procedures as follow:

(1) attaching the nasal subunit and two pressure sensors on the fiber-type mask

(2) linking the device to the data recorder (DAS 40), and setting a proper measuring range with 20-kHz sampling and 14-bit resolution

(3) adjusting the mask at a proper position. The subjects were asked to produce /m/ and /p/ to confirm no air leakage between the oral and nasal cavities

(4) instructing the subjects to pronounce test utterances naturally with a rest after finishing a row of five utterances.

3.3. Measurements

A script was written with Python to extract, display and measure data for the target CVC sequences. For the word-initial plosive, the following segmentation was made as illustrated in Figure 5:

(1) initiation of closure of the plosive (marker 1)

(2) end of prolonged nasal resonance (marker 2)

(3) release of the plosive (marker 3)

(4) vowel onset (marker 4)

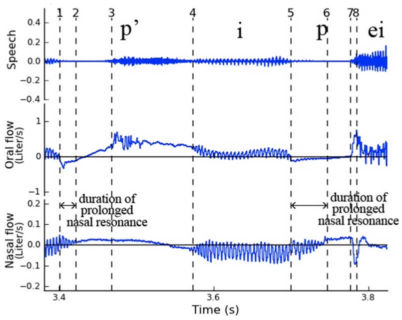


Figure 5: *Segmentation of a CVC sequence (/p'ip/ in /p'i^2 pei^4/ from subject A).*

Markers 1, 3 and 4 are determined mainly by examining the speech channel. Marker 2 indicates the end of "prolonged nasal resonance" for partially voiced plosive segments, where the oscillation amplitude of the nasal airflow rate decreases to zero. Special cases include (a) no oscillation in the nasal airflow during plosive segments (overlapped marker 1 and marker 2), and (b) continuous nasal airflow oscillation during fully voiced plosive segments (overlapped marker 2 and marker 3). The marker lines for the word-medial plosives were also made similarly (markers 5 to 8). The duration of prolonged nasal resonance is also illustrated in Figure 5.

4. Results

Examples of the CVC sequences obtained from subject A are shown in Figure 6. Six panels showing speech, oral and nasal airflow waveforms are presented by three rows (a ~ c) and two columns (1 ~ 2). Each row contains consonants with the same place of articulation, (a) alveolar, (b) bilabial, and (c) velar. Each column indicates the same phonation type but different places of articulation. In addition, the first two rows are words with vowel /i/, while the third row has the words with vowel /a/ (no /ki, k'i/ combinations in Chinese). The DC component of oral and nasal airflow was drawn with red lines.

4.1. Oral airflow

The oral airflow signals show rises of the DC component at the release burst and vowel production. The comparison between aspirated and unaspirated plosives reveals that the former has the longer duration of the release burst (about three times) than the latter (see each row in Figure 6, between markers 3 and 4).

4.2. Nasal airflow

The DC component of nasal airflow signals demonstrates a subtle rise near the closure onset (near marker 1) and a drop near the release (near marker 3). In each column of Figure 6, alveolar plosives (/t/, /t'/) have the largest rise followed by a slow drop of DC nasal airflow during the closure toward the release; bilabial plosives (/p/, /p'/) have the smaller rise followed by stationary or slow rise of airflow; velar plosives (/k/, /k'/) have the smallest rise followed by airflow attenuation. The fluctuation of DC airflow is supposed to reflect the rise and fall of the velum, caused by the velopharyngeal muscles or augmented intraoral pressure. In each row, the drop at the release is larger in unaspirated plosives than in aspirated plosives. The drop of nasal airflow at the release may be explained as a rapid fall of the velum due to the decrease of the intraoral pressure. In fact, the release of unaspirated plosives is shorter and faster.

Observing the AC component of nasal airflow, airflow oscillation sustains during the vowel segment (between marker 4 and 5) and the following plosive closure (between marker 5 and 6). It is obvious that the high vowel /i/ has the weaker sound intensity but the stronger nasal resonance in comparison to the low vowel /a/. Prolonged or enhanced nasal resonance (as seen by nasal airflow oscillation) following a vowel is found almost in every plosive. The duration of voicing during the closure is essential to judge a voiced plosive acoustically, and it is assumed to be longer and larger in word-medial plosives, especially for /t/ and /p/ (see the first two panels of the first column in Figure 6). A statistical analysis on the duration of the prolonged nasal resonance was conducted to confirm the observation above, which is shown in the next subsection.

4.3. Prolonged nasal resonance

Figure 7 shows two box plots summarizing the duration of the prolonged nasal resonance for the word-initial and word-medial plosives. It is observed that only the unaspirated /t/ and /p/ in the medial position have the mean duration longer than 40 ms, while no marked difference is found among other plosives. In fact, the maximum two durations appeared in the sequences /pap/ (102 ms) and /kat/ (82 ms) from the data of subject A (as shown in Figure 8), and these two word-medial plosives are fully voiced. Also, as shown in Figure 6 (panel in row (a) and column (1)) and Figure 8, the enhanced oscillation amplitude of the prolonged nasal resonance appears with the longer duration, suggesting a certain voicing effort for co-producing the vowel and the following plosive.

The patterns of the waveform are similar between subjects A and B, while some differences are also observed. Firstly, the intensity of speech and maximum airflow rate for B are larger than those for A (about 1.5 times), suggesting that B spent more volume of air at speaking. Secondly, the fluctuation of the DC nasal airflow tends to be smaller in the data from B. A possible explanation is that subject B has a thicker velum that moves smaller and slower than A.

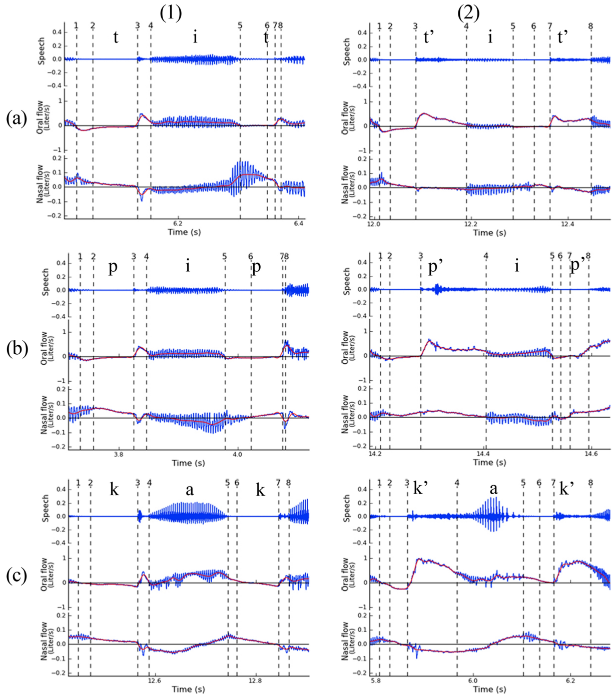
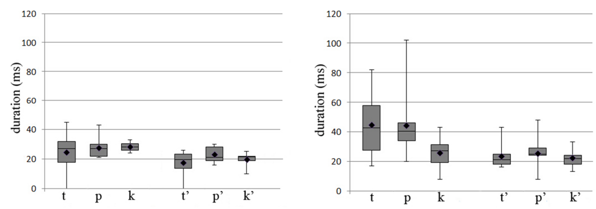


Figure 6: *Examples of CVC sequences containing six Chinese plosives (from subject A).*



(a) *word-initial plosives* (b) *word-medial plosives*

Figure 7: *Box plots for the duration of prolonged nasal resonance in each plosive.*

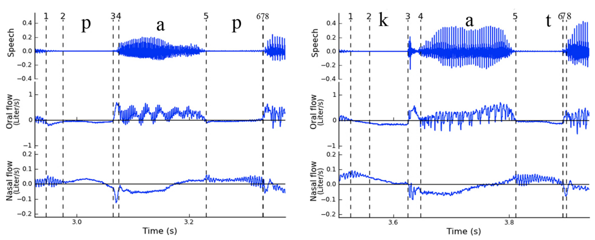


Figure 8: *Examples of voiced word-medial plosives.*

5. Discussions and conclusion

A pressure-difference pneumotachography system was devised to measure oral and nasal airflow signals using an acoustically transparent mask combined with a nasal subunit. The use of the high-sensitivity and high-stability pressure sensors guarantees the reliability of the low-frequency signals. The advantages of the system include the ease for building the system and the stability at sensing small airflow signals (critical to observe the voicing patterns). A problem remains however about the sensor characteristics [8]: the sensor's output signals are updated at the rate of 1 kHz, which interacts with the formant frequencies of vowel /a/, causing amplitude modulation of AC airflow signals (as seen in Figure 8). Therefore, the sensor used in this study is more adequate to record nasal airflow, while a different type of the sensor should be employed to record AC signals of oral airflow.

The results from the present preliminary analysis of Chinese plosives indicate the following:

(1) The oral airflow shows rises of the DC component at release burst and aspiration.

(2) The nasal airflow demonstrates marked AC components during word-medial plosives and slow DC component during the CVC utterance.

(3) The above signals together with clear speech signals help us understand velar articulation and nasal resonance for the Chinese plosives.

The nasal resonance observed during the Chinese plosives reflects the "transvelar nasal coupling" [9], which refer to the propagation of acoustic pressure variation in the vocal tract behind the constriction into the nasal cavity via the velum. The transvelar coupling observed in the nasal airflow signals also suggests phonetic realization processes of plosive sounds. The Chinese plosives are phonologically voiceless, only contrasting the aspirated or unaspirated. Analysis of the prolonged and enhanced nasal resonance confirms intervocalic Chinese plosives to be partially or fully voiced, similar to other languages [10]. Among reports on Chinese, Wu et al. also reported the higher rate of voicing in medial plosives based on a spectrographic analysis [11]; and Deterding and Nolan noted a negative tendency of the word-initial sentence-medial plosives [12].

In conclusion, the nasal-oral pneumotachography system developed in this study was found to be capable of detecting velar movement and transvelar nasal coupling during production of Chinese plosive sounds. The comparison among the present data supports that voicing tends to be avoided in the word-initial position by the effort of initial strengthening, but it is free to occur in the word-medial position as assimilation. Voicing in the Chinese plosives may reflect speaker-to-speaker differences regarding anatomy of the velum, dialect of speakers, etc. It may also depend on types of tones in the syllables. Those topics will be examined in future.

6. Acknowledgements

This work is supported by the national 1000-plan project (WQ20111200010), and the National Natural Science Foundation of China (No.61175016, No.61471259) and Key Program (No. 61233009).

7. References

[1] N. Isshiki, "Air flow during the production of selected consonants," *Journal of Speech and Hearing Research*, vol. 7, pp. 233-244, 1964.

[2] M. Rothenberg, "Measurement of airflow in speech," *Journal of Speech and Hearing Research*, vol. 20, pp. 155-176, 1977.

[3] J. Vaissière, K. Honda, A. Amelot, et al. "Multisensor platform for speech physiology research in a phonetics laboratory," *The Journal of the Phonetic Society of Japan*, vol. 14, no. 2, pp. 65-78, 2010.

[4] J. Lubker, and K. Moll, "Simultaneous oral-nasal air flow measurements and cinefluorographic observations during speech production," *Cleft Palate Journal*, vol. 36, pp. 257-272, 1965.

[5] M. Rothenberg, "A new method for the measurement of nasalance," *http://www.rothenberg.org/Nasalance/Nasalance.htm*, 2006.

[6] M. Rothenberg, "Rethinking nasalance and nasal emission," *http://www.rothenberg.org/Rethinking-Nasalance-Nasal-Emission/Rethinking_Nasal_Emission_and_Nasalance.pdf*, 2014.

[7] S. Hertegard, and J. Gauffin, "Acoustic properties of the Rothenberg mask," *STL-QPSR*, vol. 33, no. 2-3, pp. 9-18, 1992.

[8] Honeywell, "TruStability board mount pressure sensors," *http://sensing.honeywell.com/index.php?ci_id=151133*, 2014

[9] J. Dang, K. and Honda, "An improved vocal tract model of vowel production implementing piriform resonance and transvelar nasal coupling," *4th International Conference on Spoken Language Processing*, 1996, pp. 965-968.

[10] J. Beckman, J. Michael, and R. Catherine, "Empirical evidence for laryngeal features: Aspirating vs. true voice languages," *Journal of Linguistics*, vol. 49, no. 2, pp. 259-284, 2013.

[11] Z. Wu, M. Lin, et al. *The Summary of Experimental Phonetics*, Beijing: Higher Education Press, 1989.

[12] D. Deterding, and F. Nolan, "Aspiration and voicing of Chinese and English plosives," *The 16th International Congress of Phonetic Sciences*, Saarbrücken, 2007, pp. 385-388.