Measurement and Simulation of Aerodynamic Characteristics of Aspirated and Unaspirated Mandarin Chinese Stop Consonants

NIU Haijun[1], *BADIN Pierre*[2], *PU Fang*[1], *LI Deyu*[1], *FAN Yubo*[1]

[1] Department of Bioengineering’ The Beihang University, Beijing, China. 100083
[2] GIPSA-lab, DPC (ICP), UMR 5612, CNRS, Grenoble INP, Univ. J. Fourrier, Univ. Stendhal, Grenoble, France
yubofan@buaa.edu.cn, hjniu@buaa.edu.cn, pierre.badin@gipsa-lab.inpg.fr

Abstract

The production of aspirated and unaspirated Mandarin Chinese stops is complex and involves the careful coordination of laryngeal and supralaryngeal articulators. In this study, the aerodynamic characteristics of such consonants were investigated by means of aerodynamic measurements obtained from a Chinese speaker and of computer simulations based on Stevens’ (1993) low-frequency aerodynamic model. The simulation results are fairly well consistent with the experimental data, which means that (1) the low-frequency model is good enough to produce realistic results and that (2) we have been able to find suitable coordinated opening and closing gestures for both glottal and oral constrictions. This study draws the way to better knowledge of Mandarin speech production, and to developments in speech rehabilitation and articulatory speech synthesis.

1. Introduction

Speech production involves the careful coordination of the gestures of speech articulators such as tongue or lips, including vocal folds [1]. For stop consonants, contrasts such as voiced *vs.* unvoiced or aspirated *vs.* unaspirated – related to the nature and the dynamics of acoustic sources in the vocal tract – are determined by the precise time course and coordination between laryngeal and supralaryngeal articulations [2, 3]. Aspiration generation in stop consonants is generally attributed to the concomitance of a vocal tract state where vocal folds are adducted enough to produce turbulence, and where supralaryngeal occlusion is released enough to allow a significant escape of airflow [3]. Due the difficulty of measuring directly laryngeal and supralaryngeal constriction areas, parameters such as intra-oral air pressure or oral airflow have been extensively used to study and characterise these aerodynamic and acoustic phenomena [2, 3].

Most aerodynamic research on consonants has been focused on European languages: therefore, studying the case of Mandarin Chinese presents a high interest. Mandarin Chinese uses the aspirated *vs.* unaspirated contrast for stop consonants. Both Mandarin aspirated and unaspirated stops are voiceless. The aspirated *vs.* unaspirated contrast is found only in syllable-initial position in Chinese [6-7].

This paper presents a twofold approach to the characterisation of the aspirated vs. unaspirated contrast in Mandarin. It relies on (1) intra-oral pressure, oral airflow and acoustic measurements made on one speaker uttering a corpus of Mandarin aspirated and unaspirated stops, and (2) computer simulations with a low-frequency vocal tract aerodynamic model attempting to reproduce the observed data.

2. Measurements

A 33 year old male speaker of standard Mandarin Chinese was instructed to utter Mandarin speech material to obtain intra-oral pressure, oral airflow, and acoustic sound pressure measurements. Nonsense words /pVCV/ were repeated ten times during a single breath. For each item, vowel V was chosen from /a i u/, while consonant C was either the aspirated stop / p^h/ or the unaspirated stop /p/.

The measurements were carried out in a sound-shielded room at Institut de la Communication Parlée (which is now the Speech and Cognition Department of GIPSA-lab). Figure 1 is a schematic view of the experimental setup. During the recordings, the subject seated on a comfortable chair and pressed his face firmly against the mask to prevent air leakage. Three signals were recorded simultaneously: acoustic sound pressure, oral airflow, and intra-oral pressure. The oral airflow was recorded with a mouth rubber mask linked with the EVA2^{TM} aerodynamic measurement system ([10], see also http://aune.lpl.univ-aix.fr:16080/~sqlab/).

As the mask covers only the mouth, the airflow signal represents only the oral airflow. The airflow signal was digitized at 6.25 kHz with 16 bit resolution.

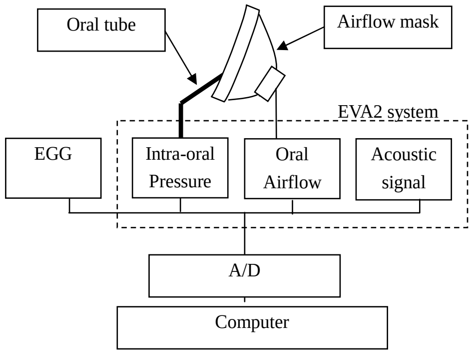
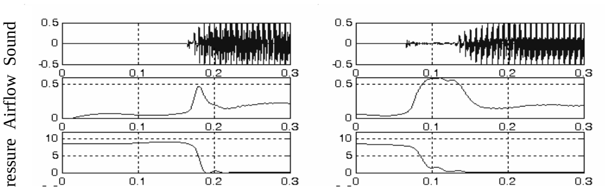


Figure 1: *Schematic diagram of the airflow-pressure experimental setup.*

The intra-oral air pressure was recorded using an oral polyethylene tube with a total length of 60 cm. The tube was bent to fit round the upper molars with the open end pointing downwards across the air stream near the midline in the oral-pharyngeal cavity, in which position it is least likely to give spurious pressure reading [2]. The outer end of the tube was led out between the face mask and the cheek and connected to the EVA2 pressure transducer. This setup allows recording labial and alveolar sounds only, excluding velar sounds for which the extremity of the tube may be obstructed by the tongue. The pressure signal was digitized at 6.25 kHz with 16 bit resolution.

The acoustic signal from a microphone located behind the oral mask was recorded and digitized at 25 kHz, with 16 bit resolution.

Figure 2 shows typical time trajectories of aerodynamic and acoustic signals for Mandarin aspirated and unaspirated stop consonants. It exemplifies the fact that the production of aspirated and unaspirated stop consonants is associated with changes of intra-oral pressure and oral airflow. One observe that for the unaspirated stop, the intra-oral pressure drops very rapidly after the beginning of the occlusion release; the oral airflow peak is sharp and short, corresponding to the frication part of the stop [4]. Oppositely, the intra-oral pressure drop is slower, while the airflow peak is more gradual and much longer, corresponding to the aspiration part of the stop.



Time (sec) Time (sec)

/pa/ /p^{h}a/

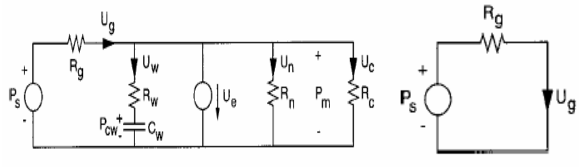
Figure 2: Time trajectories of acoustic s*ound pressure, oral airflow, and intra-oral pressure (from top to bottom) during the production of Mandarin unaspirated and aspirated stops.*

3. Computer simulations

3.1. Principles

Understanding the mechanisms of production of stop consonants requires knowledge on the laryngeal and supralaryngeal articulatory gestures and their coordination. As measurements of laryngeal and supralaryngeal constriction areas are not directly accessible through experimental setups, it is interesting to resort to simulations to attempt inferring these parameters from their measured aerodynamic consequences, and narrowing down the possibilities for realistic interpretations of the data.

One way to model speech production is to build a physical version of vocal tract, with movable articulators and air pressure source. However, this kind of model is difficult to develop and not very flexible. Therefore, numerical models based on equations or functions that represent crucial properties of the vocal tract constitute a very interesting alternative [8]. Electrical vocal tract analogues, which represent the vocal tract by means of electrical circuits, using electrical / aerodynamic analogies, constitute an important class of such models. Keating [9] gave a detailed description of a speech aerodynamic model. Stevens [4] simplified Keating’s model and developed a low-frequency model to examine vocal tract movements, airflows, and pressures occurring during stop consonant production. Hanson and Stevens [5] give a more exact description of this model. This aerodynamic model can serve as a testing ground for hypotheses concerning the less well understood components of articulation. In the present study, we have thus implemented an this low-frequency model to simulate intra-oral pressure and oral flows, as depicted in Figure 3. The comparison of simulated signals with measured ones will help testing the capabilities of the model and better understanding the principles underlying the production of aspirated and unaspirated stops in Mandarin.



(a) (b)

Figure 3: *Low-frequency aerodynamic equivalent-circuit model (a); simplified version that holds for non-nasalized unconstricted vocal-tract configurations (b) (from Stevens, 1993 [4]).*

3.2. Low-frequency aerodynamic model

This section describes our implementation of the low-frequency aerodynamic vocal tract model proposed by Stevens [4, 5].

According to Stevens [4], the acoustic resistances *Rg, Rn,* and *Rc* are determined by the cross-sectional areas of the constrictions at the glottis *Ag*, the velopharyngeal port *An*, and the oral supraglottal airway *Ac*, respectively. The volume velocities *Ug*, *Un*, and *Uc* through the constrictions are related to the pressure drops across the dynamic resistances and the corresponding airflows by the simplified generic equation:

$$R = \frac{\rho U}{2A^2}, \qquad \text{(Eq. 1)}$$

where ρ is the air density. The volume velocity *Uw* represents the rate of increase of the vocal-tract volume due to passive expansion or contraction of the vocal-tract walls. This expansion is a consequence of the increase or decrease of the intra-oral pressure *Pm*. *Cw* and *Rw* are the compliance and resistance of the walls respectively. The source *Ue* represents active expansion or contraction of the vocal-tract volume by the speaker. Positive values of *Uw* and *Ue* correspond to outward flow, or expansion of the oral cavity, while negative values correspond to the contraction of the oral cavity. In the present study no expansion was considered, and thus $Ue = 0$.

Given the subglottal pressure *Ps*, the resistances of the constrictions, and the compliance and resistance of the vocal-tract walls, flows and intra-oral pressure *Pm* can be calculated. For the production of most vowels and many sonorant consonants, there is no narrow constriction in the oral cavity and the velopharyngeal port is closed; in this case, $Rc = 0$ and $Rn = \infty$, and the equivalent circuit reduces to the simplified version shown in Figure 3(b), in which there is only one resistance, *Rg*, representing the glottal opening, and $Pm = 0$.

The circuit shown in Figure 3 has been implemented by means of the Matlab software. The control variables considered in our simulations were sublottal pressure *Ps*, glottal area *Ag*, and oral constriction area *Ac*. The resulting variables were intra-oral pressure *Pm* and airflow *Uc*. Resistances *Rc* and *Rg* are computed using Eq. 1. *Rc* is associated with the narrowest of the supraglottal constrictions (excluding the one at the nasal port). The resistance of the walls *Rw* is constant. The compliance of the walls *Cw* is derived from a modal value and the parameter *dc* (change in vocal-fold or wall compliance). When the values of constriction sizes and *Cw* have been determined, the circuit is solved for the pressures and flows.

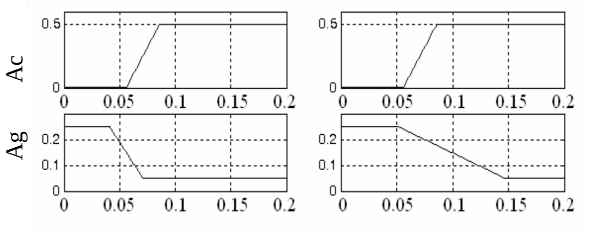
3.3. Simulations and results

The aim of the simulations was to fit as closely as possible the airflow and intra-oral pressure signals measured on the speaker with the simulated ones. As the aspirated *vs.* unaspirated contrast is found only in syllable-initial position in Mandarin Chinese, simulations have been performed only for consonants in such a position. We will exemplify this approach with a pair of aspirated /p^{h}a/ and unaspirated /pa/ bilabial stop + vowel syllables. Our approach was as follows.

All parameters were held constant at their default values (see Hanson & Stevens [5] for more details), except for the constriction areas. The measured intra-oral pressure during stop closure, which is a good estimation of subglottal pressure [3], was found very close on the average to 8 cm H_2O for both unaspirated and aspirated stops. Subglottal pressure was thus set to this constant value for all simulations.

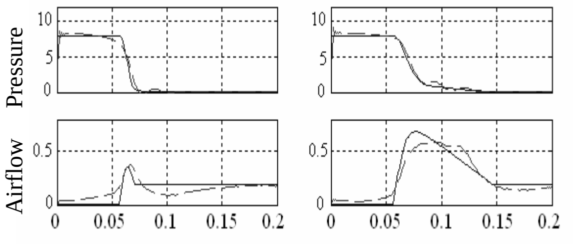
The time trajectories of cross-sectional areas of glottis constriction *Ag* and oral constriction have been modelled as linear transitions between steady state moments corresponding to the closure and vocalic parts of the syllable, as illustrated in Figure 4. Each stop was given a 30 ms long oral release gesture from fully closed to fully open. The glottal constriction area was estimated to 0.05 cm^2 for the vocalic part, using measured airflow and subglottal pressure; it was set to 0.25 cm^2 during the closure, value largely agreed upon in the literature. The oral constriction varies between complete closure and 0.5 cm^2 which can be considered as totally opened from the point of view of the aerodynamic model. The starting and ending time instants of the linear transitions have been empirically adjusted in order to get the best fit between measured and simulated signals.

Resulting simulated intra-oral pressure and oral airflow signals are displayed in Figure 5, superposed to the signals measured on the speaker. The simulation results are rather well consistent with the experimental data. After release, for both unaspirated /p/ and aspirated /p^h/, the intra-oral pressure decreases and the oral airflow increases. The oral airflow after the release in the unaspirated stop is smaller than in the aspirated one. In unaspirated stop consonants, the oral airflow is usually only slightly larger than in the following vowel. The intra-oral pressure decay after release is slower for aspirated stop than for its unaspirated cognate. We confirm the result found for other languages [3]: the necessary condition to create aspiration is a delay in the glottal closing gesture following the release of the consonantal closure.



Time (sec) /pa/ Time (sec) /p^{h}a/

Figure 2: *Time trajectories of the cross-sectional areas of the glottis constriction* Ag *and of the oral constriction* Ac *used for the aerodynamic simulation of Mandarin stop consonants.*



Time (sec) /pa/ Time (sec) /p^{h}a/

Figure 3: *Simulated (solid lines) and measured (dashed lines) time trajectories of intra-oral pressure and airflow for Mandarin stop consonants (simulations performed with the control signals shown in Figure 2).*

4. Conclusions and perspectives

In order to understand the coordination between laryngeal and supralaryngeal articulators involved in the production of Mandarin Chinese aspirated and unaspirated stop consonants, we have obtained aerodynamic measurements from a speaker uttering a corpus of syllables containing such consonants. We have then implemented and used the low-frequency aerodynamic model proposed by Stevens and colleague [4, 5] to attempt to fit the measured signals with the simulated ones. The simulation results are fairly well consistent with the experimental data, which means that (1) the low-frequency model is good enough to produce realistic results and that (2) we have been able to find suitable coordinated opening and closing gestures for both glottal and oral constrictions. To summarise, the necessary condition to create aspiration is a delay in the glottal closing gesture following the release of the consonantal closure.

Differences between simulated and measured signals still exist, however. A more refined aerodynamic model, that would take into account the moving walls, for instance, could be envisaged. Besides, more consonants and more subjects could be considered in the future, as well as more exhaustive experiments, including vocal folds observation and aerodynamic measurements. Acoustic sources models controlled from the aerodynamic state of the vocal tract could also be developed, following Mawass et al. [11].

We can conclude that this study draws the way to better knowledge of Mandarin speech production, and to developments in speech rehabilitation and articulatory speech synthesis.

ACKNOWLEDGMENTS: The work has been supported by a grant from the K.C. Wong Education Foundation and CNRS and by the National Natural Science Foundation of China. (Grant No. 10772018, 10672014).

5. References

[1] Fant, G. 1973. Stops in CV-syllables, in *Speech Sounds and Features* (MIT Press, Cambridge, MA). 110-139.

[2] Hardy, J., 1965. Airflow and pressure studies. *ASHA Reports*, 1, 141-152.

[3] Scully, C., Mair, S. J., 1995. Relationships between different descriptive frameworks for plosive features of voicing and aspiration. in *Levels in Speech Communication: Relations and Interaction*, J. Schoentgen, J. M. Ramlot, C. Sorin, H. Meloni, and J. Mariani, Eds. Berlin, Germany: Elsevier. 51–62.

[4] Stevens, KN., 1993. Models for the production and acoustics of stop consonants. *Speech Communication* 13, 367-375.

[5] Hanson, HM. Stevens, KN. 2002. A quasi articulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn. *J. Acoust. Soc. Am.* 112, 1158-1182.

[6] Iwata R., Hirose, H., 1976. Fiberoptic acoustic studies of Mandarin stops and affricates, *Annual Bulletin of the RILP*, 10, 47-60.

[7] Wu, ZJ. 2004. *Wu zongji yu yan xue lun wen ji* [in Chinese]. Beijing: Shangwu Printing House.

[8] Ohala, J., 1976. A model of speech aerodynamics, *Rep. Phonology Lab.* (Berkeley) 1, 93-107.

[9] Keating P., 1984. Aerodynamic Modeling at UCLA. *UCLA Working Papers in Phonetics* 59, 18-28.

[10] Ghio, A., Teston, B., 2004. Evaluation of the acoustic and aerodynamic constraints of a pneumotachograph for speech and voice studies. *Proceedings of the International Conference on Voice Physiology and Biomechanics*, Marseille, France.

[11] Mawass, K., Badin, P. & Bailly, G., 2000. Synthesis of French fricatives by audio-video to articulatory inversion. *Acta Acustica*, 86(1), 136-146.