A Noninvasive, Low-cost Device to Study the Velopharyngeal Port During Speech and Some Preliminary Results

Xiaochuan Niu, Alexander B. Kain, Jan P. H. van Santen

Center for Spoken Language Understanding,
OGI School of Science & Engineering at OHSU
20000 NW Walker Road, Beaverton, Oregon 97006, USA
{xiaochua,kain,vansanten}@cslu.ogi.edu

10.21437/Interspeech.2006-308

Abstract

It is desirable to monitor the status of the velopharyngeal port during speech. This paper reports on the design and usage of a noninvasive device that measures the static nasal airflow from the nose during speech. The signal can be recorded directly from a generic sound card. Neither does the usage of this device interfere with the articulatory process of speech, nor does it introduce distortions to the simultaneously recorded acoustic signal. We successfully tested the device in an experiment to analyze the velopharyngeal status during normal speech.

Index Terms: speech production, nasal airflow, speech pathology, velopharyngeal function

1. Introduction

The appropriate manipulation of the velopharyngeal (VP) opening is important in normal speech production. The production of a nasal consonant involves the opening of the VP port and the closing of the oral tract at a certain point, in order for air puffs from the vocal fold to propagate only through the nasal tract. During vowel productions, when the VP port is open and the air propagates through both oral and nasal tracts, the vowel is nasalized. There are also nasal vowels in languages such as French and Portuguese, in which the precise control of nasal-oral coupling is needed. Some consonants may also be nasalized in the context of nasalized vowels [1]. In addition, the closure of the VP port is essential for most plosives and fricatives, because sufficient air pressure is needed to produce these phonemes. Inappropriate opening or closing of the VP port during speech may cause nasalization problems, which is the characteristic of some groups of disordered speech [2]. It is of interest to monitor the status of the VP port during speech. This information, accompanied by the acoustic signal, can be used for a more refined modeling of speech production, and for the analysis and enhancement of disordered speech with nasalization problems. However, because of the special physiological position of the VP port, it is difficult to monitor the VP port directly without some level of interference with the articulatory process of speech, usually involving invasive, expensive, or bulky instruments. It is therefore desirable to develop techniques that can extract valid information of the VP port status that are less invasive, less interfering, and less expensive.

Speech researchers and clinicians have developed many techniques to study the VP function during speech. There is a thorough review of them in Baken and Orlikoff's book [2] mainly for clinical purposes. Among them, the most direct but invasive way is to record the VP port images with an endoscope inserted into the nasal cavity [3]. The images obtained in this way may be unstable and distorted, which makes it difficult to make quantitative measurements. A photodetection system [4] has been developed to measure the intensity of light transmitted through the VP port. It needs a light source and a light detector to be placed at different sides of the VP port. A relatively less invasive technique is the speech movement tracking system, using either X-ray microbeam [5, 6] or magnetometry [7]. Pellets or coils can be attached to the velum in order to track its movements. These systems are usually costly and only a few research sites have them. An indirect but less costly way is to measure the nasal airflow with a pneumotachograph [8] or a hot-wire anemometer [9] that is connected to a nasal mask. The mask seals the air from the nose, but it also introduces distortions to the acoustic signal and sometimes interferes with articulation.

In this paper, we present the design and usage of a noninvasive, low-interference device that simply adopts a differential pressure sensor to measure the airflow from the nose during speech. Some simple circuits are designed to supply power to the sensor and to convert the signal in order to be recorded by a generic sound card as found in most consumer-grade computers. The device is then used to analyze the VP status during the speech of a normal speaker.

2. Device design

2.1. Measurement principles

The goal of our design is to monitor the VP status while simultaneously recording the acoustic speech signal. In order to avoid interference with articulation and distortion of the acoustic signal, we choose an indirect way to measure the air velocity outside the nostril during speech.

The velocity of the moving air can be measured according to the same principle as those of a Pitot tube. Assuming that air is incompressible, when the opening of a tube faces an oncoming stream of air, it senses the summation of the static pressure P_s and the dynamic pressure P_d; When air flows in the opposite direction, the tube senses the difference between P_s and P_d. The air velocity inside the tube is zero, and the pressure in the tube is

$$P_t = P_s \pm P_d = P_s \pm \frac{1}{2}\rho V^2, \quad (1)$$

where ρ is the density of air, and V is its velocity.

This work was partially supported by NSF grant 0117911, "Making Dysarthric Speech Intelligible". We also thank Bob Nelson for helping us in selecting a sensor and building an initial prototype.

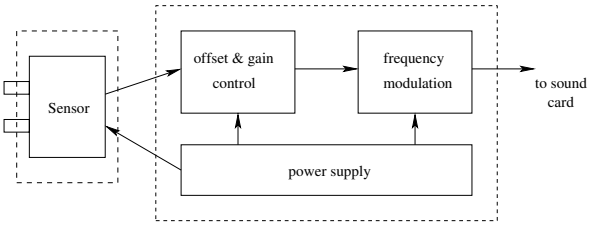

Figure 1: *The schematic structure of the measuring device. The sensor has two air intakes to measure differential pressures. Arrows indicate the flow of signals between modules.*

A differential pressure sensor with one intake connected to the probe tube and another open to the static air can be used to measure the pressure difference, $P_t - P_s$. According to Equation (1), the sign of this measurement represents the direction of the airflow, and the absolute value is proportional to the square of the velocity of the airflow.

2.2. Data acquisition

Thanks to the prevalent usage of personal computers (PCs), it is convenient and inexpensive to acquire acoustic signals of speech with a sound card in a PC. However, a generic sound card usually uses AC coupling for the input signal, thus filtering out direct current (DC) and low-frequency information of the signal. Because we want to analyze the DC and low-frequency components in the airflow signal, a sound card can not be used directly for the acquisition of airflow signals.

One solution to this problem is to use a multi-channel data acquisition card to collect the airflow signal and the acoustic signal of speech, but it is expensive and not equipped in most consumer-grade computers. Another solution is to pre-process the airflow signal with a voltage-controlled oscillator (VCO), so that the input signal is frequency modulated (FM) with a carrier frequency in the audio range (2–5 kHz). Then the FM signal is recorded through one input channel of the sound card, while the acoustic signal is recorded through another input channel simultaneously. The recorded FM signal, once captured, can be demodulated algorithmically to recover the original airflow signal.

2.3. Device implementation

Figure 1 shows the schematic structure of the device. A low pressure sensor (1 MBAR-D-4V, All Sensors, CA) is encapsulated in a small plastic box that can be attached to a headset. The operating pressure range of the sensor is $\pm$ 1 mbar. The sensor is connected to a processing box that contains the power supply, offset and gain control, and frequency modulation modules. The power supply module includes batteries and regulators. It supplies stable voltage for the sensor and other components. The module for offset and gain control simply includes op-amps and potentiometers. They are tuned to convert the output signal of the sensor into the operating range of the VCO in the frequency modulation module. A waveform generator chip (NTE 864, NTE Electronics, INC., NJ) is used as the VCO. Several resistors and one capacitor are chosen in order to set the highest frequency of the VCO to about 5 kHz. The sine wave output of the VCO is then amplified to standard audio device line levels. This signal is sent to the line input of the sound card of a PC. As a last step, the offset of the VCO input is tuned when zero differential pressure is applied to the sensor (e.g. when the sensor is inert), so that the frequency of the output signal is about 3 kHz.

3. Signal processing

3.1. Demodulation

A FM signal can be represented as

$$y(t) = K \cos\left[2\pi f_c t + \phi(t)\right], \tag{2}$$

where f_c is the carrier frequency, and K is a constant. The instantaneous frequency, $\omega(t)$, of the FM signal is proportional to the input signal $x(t)$, that is

$$\omega(t) = 2\pi f_c + \frac{d}{dt}\phi(t) = 2\pi[f_c + f_d x(t)], \tag{3}$$

where f_d is the frequency deviation. When $x(t)$ is a narrowband signal, the Hilbert transform of $y(t)$ is

$$\hat{y}(t) = K \sin\left[2\pi f_c t + \phi(t)\right]. \tag{4}$$

The analytical signal of $y(t)$ is

$$y_a(t) = y(t) + j\hat{y}(t) = Ke^{j[2\pi f_c t + \phi(t)]}. \tag{5}$$

Algorithms for frequency demodulation are usually based on calculating the derivative of the phase of $y_a(t)$ [10, 11].

When the FM signal $y(t)$ is sampled with a frequency of F_s ($F_s \gg f_c$), the resulting discrete signal is $y[n]$. Its corresponding analytical signal $y_a[n]$ can be obtained through the discrete Hilbert transform. Given the carrier frequency f_c, an auxiliary signal $y_x[n]$ is obtained by

$$y_x[n] = y_a[n] \cdot e^{-j2\pi f_c n/F_s} = Ke^{j\phi[n]}. \tag{6}$$

The phase signal $\phi[n]$ is the unwrapped angle of the signal $y_x[n]$. The derivative of the phase is approximated by differencing operations. Because we are only interested in the relative changes of the input signal $x[n]$, the constant factor f_d is ignored. We use this algorithm as implemented by the function *demod* in Matlab's signal processing toolbox [12]. Before demodulation, we apply a zero-phase low-pass filter to the FM signal. The cutoff frequency of the filter is set to 5 kHz in order to filter out higher-order harmonics.

3.2. Zero calibration

When no airflow exists at the sensor, the frequency of the output signal is equal to the carrier frequency f_c. Though this frequency has been tuned to about 3 kHz during the construction of the device, it can drift due to temperature changes and other environmental conditions. An accurate value for f_c can be calculated from a section of the FM signal, $y_0[n]$, that is recorded under the zero-airflow condition.

The demodulation procedure described above can be represented as a transform from the signal $y_0[n]$ into the signal $x_0[n]$, with the parameter f_c,

$$x_0[n] = \mathcal{D}\left\{y_0[n]; f_c\right\}. \tag{7}$$

The optimal value f_c^* minimizes the root mean square (RMS) of $x_0[n]$,

$$f_c^* = \arg\min_{f_c} \mathrm{RMS}\left\{\mathcal{D}\left\{y_0[n]; f_c\right\}\right\}. \tag{8}$$

A simple line-search algorithm is used to find f_c^*, starting from an initial guess of 3 kHz. The optimal carrier frequency is then used in the demodulation operations of other signals recorded during the same session.

4. Experiment

4.1. Speech materials

Currently, the purpose of our experiments is to examine whether the proposed device can provide useful information about the VP status during speech in addition to the acoustic signal. To observe the change of the VP status, we design four groups of words in the forms of CVN, NVC, NVN, and CVC. In these words, C is a consonant chosen from /t/, /d/, /s/, /z/, V is a vowel chosen from /i:/, /@/, /A/, /u/, and N is the nasal /n/ [1]. The groups are supposed to represent the opening process, the closing process, the complete opening, and the complete closing of the VP port within a monosyllable, respectively. Table 1 lists the set of 16 words used in the experiment. During the recording procedure, each word is inserted in the carrier sentence, "Say _ please".

CVN	NVC	NVN	CVC
/d @ n/	/n @ d/	/n @ n/	/d @ d/
/s A n/	/n A s/	/n A n/	/s A s/
/t i: n/	/n i: t/	/n i: n/	/t i: d/
/z u n/	/n u z/	/n u n/	/z u z/

Table 1: *The list of recorded words.*

The data were recorded by a male speaker. A head-mounted AKG HSC 200 condenser microphone was used to record the acoustic signal. The microphone was placed off-axis, 5 cm away from the edge of the speaker's mouth. The small box that holds the pressure sensor was attached to the frame of the headset. A 25-cm-long surgical tube was connected to one intake of the sensor with one end, and the other end of the tube was bent to point to the opening of one nostril of the speaker. The acoustic signal and the FM airflow signal were recorded simultaneously to a hard drive through two channels of a MAudio Delta 1010 system. Recordings were performed in an acoustically dampened booth. Waveforms were sampled at 44.1 kHz and stored in 16-bit PCM format. This sampling rate is far greater than the carrier frequency of the FM signal. At the end of the recording session, an extra 20 seconds were recorded with the headset taken off and left in the booth, for the purposes of calibration. After recording, the phoneme boundaries were manually marked by one of the authors according to the acoustic signals.

4.2. Analysis and results

Our first observation about the demodulated airflow signals is that they all contain strong harmonic components during the sections of voiced speech sounds. These components are caused by acoustic vibrations of the air that mainly propagates from the nostril. This information is redundant since the acoustic signal has been picked up by the microphone. To eliminate these harmonic components, we convolve each airflow signal with a 30 ms normalized Hamming window. This operation is equivalent to a weighted average of the signal within the window length, effectively low-pass filtering the signal with a cutoff frequency at about 20 Hz. The DC and low-frequency components remain, representing the average static airflow as it moves in and out of the nostril during speech.

[1] All phonemes are denoted in Worldbet.

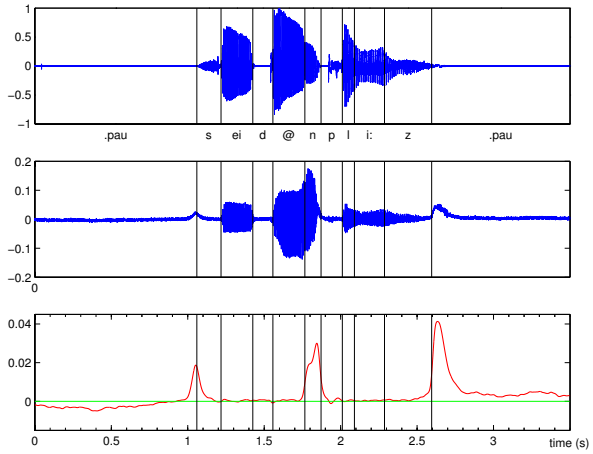

Figure 2: *Acoustic and nasal airflow signals of the sentence "Say /d @ n/ please". The normalized acoustic signal is plotted in the top panel; the demodulated airflow signal is plotted in the middle panel; and the static airflow is plotted in the bottom panel. The vertical bars represent phoneme boundaries.*

As an example, Figure 2 shows the acoustic signal, the demodulated airflow signal, and the static airflow signal of a sentence. The phoneme boundaries are manually marked in accordance with the acoustic signal. It can be seen from the middle panel that the device picks up the information of both the static airflow and the vibrations. The bottom panel shows three prominent peaks of the static nasal airflow. The first peak is located just before the beginning of the sentence, the second peak corresponds to the sound /n/, and the third peak is located right after the end of the sentence. A slight nasal inhalation can be observed as a negative signal section before the utterance, and a nasal exhalation is also observed after the utterance.

The negative nasal inhalation signal before the utterance in the above sample may or may not be observed in other sentences, because the speaker sometimes inhaled through the mouth instead of the nose. The signal of nasal exhalation after an utterance is always observed in each sample, but varies in both scale and slope of change among different sentences. This variation indicates the speaker could control the release of breath in different ways.

The peaks at the beginning and the end of an utterance are observed in all the static airflow signals of the recorded sentences. Since all the sentences in this study begin with the consonant /s/, whose target VP status is closure, the velum has to move from its rest position to close the VP port while the oral pressure increases at the beginning of an utterance. This action may push a certain amount of air out of the nose at the beginning of the utterance. At the end of each utterance, the velum always returns to its rest position. When the sentence ends with a pressure consonant, such as /z/ in our recordings, the air may be rapidly released from the nose, thus causing the flow peak at the end.

In order to compare the VP behaviors during the production of different words, we extracted the static nasal airflow signal corresponding to the phoneme sequence of each word and its adja-

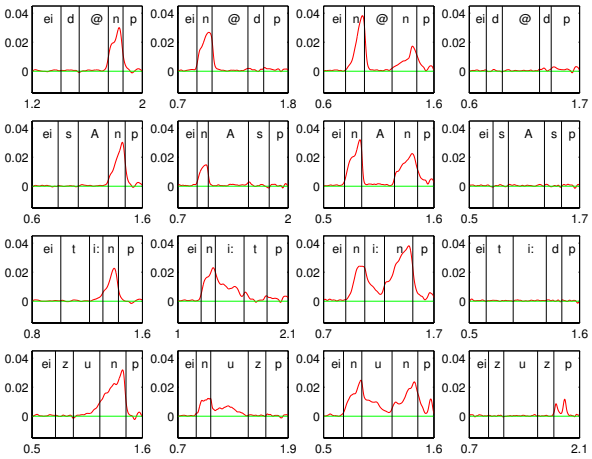

Figure 3: *Static nasal airflow signals of recorded words. Each section of the signal is cut from the carrier sentence. Only the phonemes that are adjacent before and after the word are kept.*

cent phoneme contexts. Figure 3 displays these signals in different panels. The signal in each panel corresponds to a word in Table 1. Each column of the panels belongs to a group of words. It can been seen that the patterns of the static nasal airflow signals within each group are similar. Airflow peaks are present during all nasal segments, while the signals are approximately zero during all CVC segments. In the CVN group, the positive rise of the airflow can start early in the vowel segment due to anticipatory coarticulation. In the NVC group, the positive airflow drops after the nasal but can extend into the following vowel, presenting the effect of carry-on coarticulation. In the NVN group, the signals indicate that the vowels are influenced by the nasals from both directions. The positive static nasal airflow signal during a vowel is an indicator of nasalization. It can also be seen that the positive static nasal airflow extends into the high vowels (/i:/ and /u/) further in time than into the low vowels (/@/ and /A/). This difference may be explained by the fact that the oral cavity of a high vowel has greater flow resistance than that of a low vowel.

In some panels of Figure 3, a slight amount of nasal emissions can also be observed between /n/ and /p/ segments of the signals, which can also be explained as coarticulation effects. Finally, it is interesting to notice that there are two adjacent airflow peaks during the /p/ segment after the last word /z u z/. One explanation for this is that the speaker might have added a "comma" before the word "please" during the recording of this sentence. The two peaks indicate the actual end and then the beginning of two adjacent utterances.

4.3. Summary

The analysis above shows that a quantitative measurement of the static nasal airflow can be effectively extracted from the signals obtained through our proposed device. The resulting signal of the static nasal airflow contains not only the non-speech information, such as inhalation and exhalation, but also the VP information about the detailed time-course of nasal, nasalized vowel, and nasal emission events during normal speech. The observations of the VP status during speech in our experiment are consistent with the findings of other researchers [5, 6].

5. Conclusions

In this paper, we present the design and the implementation of a noninvasive device that is used to measure the nasal airflow during speech. A small, low-cost differential pressure sensor is used to pick up the dynamic pressure of the airflow. The total raw cost of the device is less than $100. The airflow signal is frequency modulated so that it can be recorded by a generic sound card. The airflow signal is recovered by a demodulation operation. A filtering process extracts the static nasal airflow from the signal.

The sensor is light enough to be attached to a headset. The usage of this device does neither interfere with the articulatory process during speech, nor does it cause degradation of the simultaneously recorded acoustic signal, which is critical to the further analysis and enhancement of acoustic signals of disordered speech.

We carried out an experiment designed to verify the validity of the device. The obtained signals reflect various VP events such as nasals, nasalized vowels, and nasal emissions. We expect to make further use of this device in analyzing VP-related problems in dysarthric speech and enhancing the intelligibility of such speech.

6. References

[1] P. Ladefoged, *A Course in Phonetics*, Harcourt Brace, 3rd edition, 1993.

[2] R. J. Baken and R. F. Orlikoff, *Clinical Measurement of Speech and Voice*, Thomson Delmar Learning, 2nd edition, 2000.

[3] M. P. Karnell, E. J. Seaver, and R. M. Dalston, "A comparison of photodetector and endoscopic evaluations of velopharyngeal function," *Journal of Speech and Hearing Research*, vol. 31, pp. 503–510, 1988.

[4] J. J. Ohala, "Monitoring soft palate movements in speech," *Journal of the Acoustical Society of America*, vol. 50, no. 140(A), 1971.

[5] J. Dang and K. Honda, "Investigation of the acoustic characteristics of the velum for vowels," in *ICSLP*, 1994.

[6] K. L. Moll and R. G. Daniloff, "Investigation of timing of velar movement during speech," *Journal of the Acoustical Society of America*, vol. 50, pp. 678–684, 1971.

[7] W. Engelke, T. Bruns, M. Striebeck, and Hoch. G, "Midsagittal velar kinematics during production of VCV sequences," *Cleft Palate-Craniofacial Journal*, vol. 33, pp. 236–244, 1996.

[8] D. W. Warren, "Nassal emission of air and velopharyngal function," *Cleft Palate Journal*, vol. 16, pp. 279–285, 1967.

[9] B. Hutters and K. Brrndsted, "A simple nasal anemometer for clinical purposes," *European Journal of Disorders of Communication*, vol. 27, pp. 101–119, 1992.

[10] T. Oberg, *Modulation, Detection and Coding*, John Wiley & Sons, LTD, 2001.

[11] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal—Part 2: Algorithm and application," *Proceedings of the IEEE*, vol. 80, no. 4, pp. 540–568, 1992.

[12] MathWorks, *http://www.mathworks.com/*, The MathWorks, Inc.