ACOUSTIC, NON-INVASIVE MEASUREMENT OF VELOPHARYNGEAL APERTURE USING A HIGH FREQUENCY TONE

Kevin B. McGowan*, Michael T. Johnson**, Aleah Combs*, Mohammad Soleymanpour**

*Department of Linguistics,University of Kentucky, Lexington, KY USA 40506

**Electrical and Computer Engineering, University of Kentucky, Lexington, KY USA 40506
kbmcgowan@uky.edu, mike.johnson@uky.edu, aleah.d.combs@uky.edu, m.soleymanpour@uky.edu

ABSTRACT

Many linguists and other speech researchers need access to the real time articulatory state of the velopharyngeal port to investigate the timing and extent of nasal gestures alongside the acoustic consequences of those gestures. While numerous methods exist (e.g. airflow, velotrace [10], nasometer, etc.), they tend to be invasive or to collect nasal data at the expense of the acoustic speech signal. We present an inexpensive method and procedure to investigate nasal gestures using low frequency ultrasound. We provide plans for a 3d printed coupler or a vinyl tube connection that allows researchers to play a 20 kHz tracer tone into a nostril which can then be collected at the mouth using microphones typically used in speech research. Sample nasal traces of nasal consonants and nasalized vowels are provided along with calibration and analysis procedures.

Keywords: nasalization, speech production

1. INTRODUCTION

It is frequently necessary to understand the articulatory state of the velopharyngeal port, the muscular connection between the mouth and the nasal cavity. This mechanism is an important part of natural speech, controlling nasal airflow and speech nasality [5]. Measurements of this velopharyngeal mechanism are necessary to provide insight into both the timing and the ultimate extent of nasal gestures.

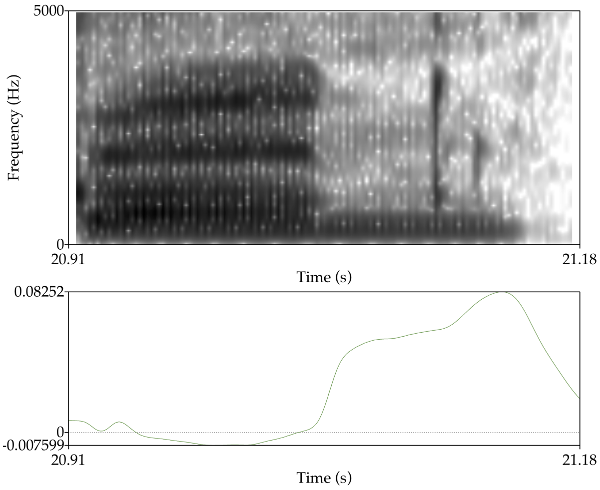
Methods of investigating articulatory gestures of the velopharyngeal mechanism can be direct or indirect [2]. The most invasive direct methods such as nasoendoscopy, in which a fiberoptic camera is inserted through a nostril; or velotrace, in which a series of rods connected to a lever are inserted through a nostril to rest on the soft palate; are rarely used in speech production research. A number of non-invasive direct imaging solutions exist but are also seldom used. Radioflouroscopy has fallen out of favor due to legitimate concerns regarding exposure to ionizing X-ray radiation. Ultrasound imaging of the velopharyngeal mechanism is possible, by placing an ultrasound probe on the soft tissue behind the ear or on the upper portion of the neck posterior to the mandible. However, the distance from the probe to the area of interest and the restricted viewing angle available make this use of ultrasound unappealing for speech research.

The method of indirect investigation of velum movement most naturally available to speech researchers is acoustic analysis [13, 14]. Important features include spectral tilt; formant frequency, bandwidth, and amplitude; the presence of anti-resonances (in nasal stops); and A1-P0 [4]. These features are typically used in conjunction with other direct or indirect methods of assessment.

The state of the art indirect method of investigating velum movement is a pressure sensor attached to a divided mask or pair of tubes which permits the collection of nasal and oral airflow data. Airflow itself is of direct, diagnostic use in, for example, speech pathology. In linguistics, though, airflow measurements serve merely as a proxy for the actual measurement of interest: acoustic coupling between the nasal and oral cavities (the coupling required for speech to sound 'nasal'). Unfortunately, the masks typically used for the collection of airflow data collection severely muffle the speech signal and, with most designs, even interfere with speech articulation [8] (although c.f. Derrick et al. [6] for a promising alternative).

As can be seen in Figure 1, the relationship between airflow and acoustic coupling of the resonating cavities is not simple. Here we see a natural production of the word *bend* paired with airflow measurements. While nearly the entire word is perceived by listeners as 'nasal' due to acoustic coupling throughout, the airflow signal is actually negative during the nasalized vowel and rises only during the nasal consonant. In tokens such as this one, acoustic coupling has occurred, and is perceptually available during the vowel [1] but this is difficult to determine from the airflow – particularly when processing large amounts of variable data across multiple speakers and prosodic environments.

Figure 1: Production of *bend* with nasal airflow tracing.



We present a device and method which uses low frequency ultrasound to directly measure acoustic coupling between the nasal and oral cavities. The intention is to provide an inexpensive, non-invasive, and effective tool which neither alters speech production behavior nor degrades the speech signal. Recording of the acoustic coupling data is done using the same high quality microphones typically used in acoustic analysis. The system uses a nasal olive (a small foam plug) to inject a 20 kHz tone from consumer grade earbud headphones into one nostril. 20 kHz has been chosen, first, because it is just beyond the range of most human hearing. Second, 20 kHz is within the sensitivity range of high-quality microphones sampling at, e.g., 44.1 kHz. In principal, higher tone frequencies could be used.

When the velum is closed (so there is no acoustic coupling between the nasal and oral cavities), this high frequency tracer tone is attenuated by the nasal cavity. However, when the velum opens in a nasal gesture, the tracer signal is detectable at the lips by the same microphone being used for voice recordings. The tracer signal can then easily be isolated from the voice recordings for analysis using software such as Praat [3] or Matlab [11].

2. TRACER TONE: INTRODUCTION, RECORDING, AND ANALYSIS

This section describes the construction and testing of nasal tone generation, introduction, recording, and analysis procedures and hardware. All recordings were made using an Electro-Voice RE20 microphone in a WhisperRoom sound-attenuated booth. The open source WaveSurfer [12] and Friture audio programs were used for recording and testing to allow real time spectrogram visualization. This real time visualization is extremely important as it allows researchers to ensure that amplitude of the inaudible 20kHz tone is neither too high nor too low.

2.1. Introduction of 20kHz tracer tone

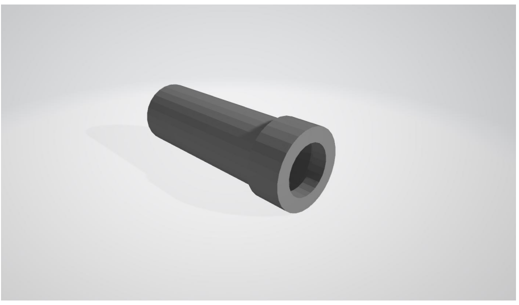
The 20kHz tracer tone was generated using an off-the-shelf high-resolution certified earbud headphone with hybrid dual drivers [7]. In order to comply with the high resolution audio industry standard and receive certification, a product must be capable of generating at least 40kHz. Any earbud with this industry certification will suffice for generation of the tracer tone. The second requirement to look for in a tone generation source is the physical interface between the driver and the ear. Suitable headphones will use replaceable silicone or foam tips; headphones where the speaker hangs directly in the intertragic notch of the pinna can not be coupled to the nasal olive using the methods described in this paper.

The silicone tip was removed from the earbud leaving a small cylindrical protrusion approximately 6mm long with a tubular lip that was also roughly 6mm in diameter. The earbud needed to be physically attached to the foam nasal olive to inject the tracer tone into the nostril. Two methods were tested to physically couple the earbud emitting the 20 kHz tracer tone through the nasal olive.

2.2. 3D printed coupler

Initially, a solid mechanical coupling tube was designed using freely available design software, and 3D printed on a Makergear M2 3D printer. The dimensions were adjusted to give a snap-fit to the earbud protrusion, and the inner structure was tapered to minimize acoustic loss. For testing, the coupling tube was press-fitted to the built-in plastic tube extending from the nasal olive, but there was found to be less acoustic loss if the built-in tube was removed and the foam of the nasal olive pulled directly over the coupling mechanism. As long as the body of the earbud itself is relatively sound-insulated, this method of physical coupling is ideal. It allows minimal signal loss from the generated tone to the wearer's nostril – increasing the amplitude of the 20 kHz tone available for subsequent analysis.

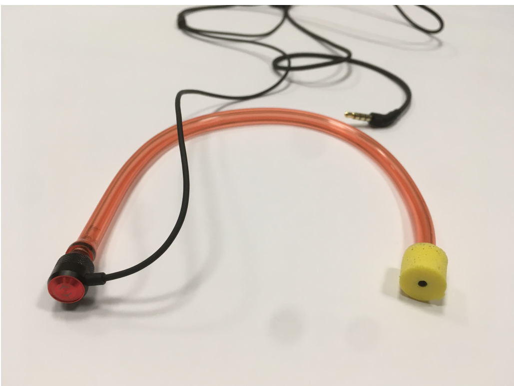
Figure 2: Rendering of 3D printed coupler



2.3. Tube coupler

In the process of testing signal reception, it was found that headphone signal was emanating from the plastic housing surrounding this particular earbud. The physical proximity of the earbud to the microphone caused a tracer signal to be present regardless of the status of the velopharyngeal port. To correct this, a second coupling method was developed in which a flexible 30 cm long plastic tube was used to connect the earbud to the nasal olive. This placed the earbud far enough from the recording microphone to eliminate tracer signal interference. The flexible tubing used in this project was a piece of aquarium airline tubing with an interior diameter roughly 2mm smaller than the earbud tip protrusion. The end of the tube was softened slightly with a soldering iron (although a heat gun would also suffice for this purpose) after which the tubing fit easily and snugly onto the earbud. The resulting device, show in Figure 3 was then clipped to the participants' shirt in the manner of a lavalier or lapel microphone.

Figure 3: Tube coupler with earbud and nasal olive



2.4. Signal Analysis & Calibration

A single microphone was used to record the subjects' speech and the tracer tone conducted through the nasal passage, velopharyngeal port, and mouth. The speech signal was digitized at a 48kHz sampling rate, and analyzed for tracer tone signal alternation. This sampling rate, typical in speech production research, has a Nyquist frequency of 24kHz which is well above the cut-off necessary to record the 20kHz tracer tone. However, care must be taken to consult the frequency response characteristics of the microphone selected for research. Many commercially-available microphones will not reliably record frequencies at 20kHz or higher and those that do will typically have some degree of roll-off in these frequencies which may decrease the tracer tone signal available for analysis.

The tracer tone is recorded alongside the speech signal and saved in the same audio file. An example spectrogram is shown on the bottom of Figure 4. This tracer tone was then separated from the speech signal using a narrowband bandpass filter, with a passband between 19,950 and 20,050Hz. The RMS envelope of the bandpass filter output was then calculated to create a tracer signal amplitude signal. A simplified excerpt of the Matlab analysis code is shown as follows:

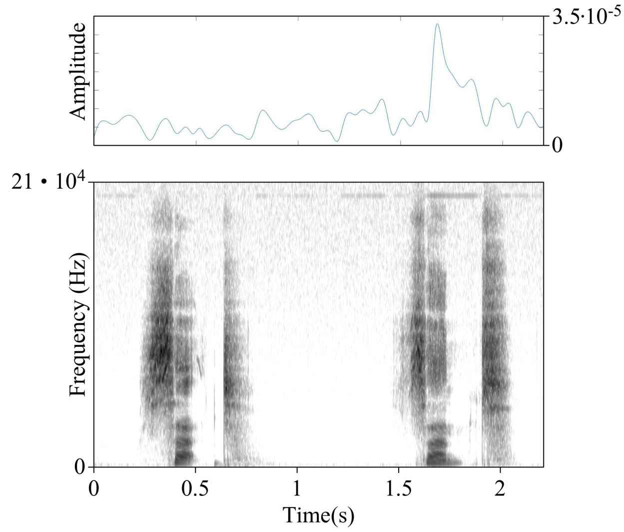
```
[b,a] = butter(4,[(19950/(fs/2))
    (20050/(fs/2))]);
tracersignal = filter(b,a,
    fullsignal)
traceramplitude = envelope(
    tracersignal, 1000, 'rms');
```

A threshold for identifying velopharyngeal port status was determined experimentally using the peak value of the signal during a nasalized vowel pronounced in context. As described below, further work will be required to determine how subject-specific and context-specific this threshold may need to be in order to consistently evaluation velopharyngeal port status.

3. PRELIMINARY RESULTS

An initial experiment with two subjects was implemented to evaluate system feasibility using the earbud and flexible tube coupling to the nasal olive, as described in the previous section. To gather comparative data within an identical prosodic context, the subjects read a wordlist consisting of 27 CVC, CVNC, and NVN target words. An example of the resulting data can be seen in Figure 4. The oral/-nasal word pair [ˈsɛt] *set* and [ˈsɛ̃nt] *sent* are shown here with the 20kHz tracer signal still present in the spectrogram (bottom). This single recording is then separated into speech and tracer signal recordings by lowpass and bandpass filtering, respectively.

Figure 4: [ˈsɛt] (left) and [ˈsɛ̃nt] (right) with 20 kHz tracer signal (top)



The differential presence of the tracer signal between the oral and nasal word conditions can be observed directly in the spectrogram. The resulting tracer signal amplitude envelope, shown above, confirms that the tracer signal level is significantly stronger during the nasalized portion of the vowel and nasal consonant than during any other portions of the recording. One potential problem is that, due to oral closure, the tracer signal may be attenuated during nasal consonants. Note that the tracer signal is also visible when subjects are breathing due to the velopharyngeal port being open when at rest [9].

4. DISCUSSION

Perhaps the most surprising discovery during the development of the nasal tracer system was how often normal parts of the speech signal have high frequency components extending to, and above, 20kHz. For some speakers, release bursts, aspiration, and fricatives (particularly the sibilant fricatives), had high frequency components of sufficient amplitude to be confusable with nasalization as detected by the nasal tracer approach. This discovery leads to two potential problems with the current approach. First, the proposed lowpass filtering to separate the speech signal from the nasal tracer signal is likely to result in loss of high frequency speech signal information for some speakers. It must be determined on a study-by-study basis whether this high frequency information is crucial to the research question under consideration. One possible solution to this problem is to raise the frequency of the tracer signal itself. This may be necessary, regardless of the research question, if the study includes children or teens or other talkers who may hear 20kHz. Raising the frequency signal only increases the need for a high quality microphone with a well-understood frequency response, and other possible consequences of this change have not yet been evaluated. Second, the presence of whisps of normal speech signal up in the 20kHz range suggests that the nasal tracer approach will not be suitable for the investigation of nasalized fricatives in a language like Scottish Gaelic [15] or nasal coarticulation of fricatives in other languages. Finally, it remains to be seen whether the nasal tracer system can be used to determine the *extent* of a nasal gesture in addition to the timing. Data like the RMS envelope in Figure 4 tantalizingly suggest that this is indeed the case, but much work remains to be done on this question.

5. CONCLUSION

A need still exists for an easy, inexpensive, non-invasive and reliable method of assessing velopharyngeal articulation during speech. Our method is unique in responding to this need by directly measuring acoustic coupling, the measure typically of interest in linguistics, rather than airflow or pressure. It remains to be seen whether the system presented here is reliable across participants and across research contexts (e.g. sociolinguistic fieldwork, paired with eye tracking for simultaneous collection of production and perception data, paired with fMRI using pneumatic headphones, or when paired with ultrasound imaging of the tongue or electromagnetic articulography). However, we believe the results presented here are promising. The nasal tracer system is affordable, relatively easy to construct and use, and non-invasive. Future work will include direct comparison with a state of the art airflow system and a large-scale, systematic test of the system to determine the reliability of the system across a varied population of talkers.

6. ACKNOWLEDGMENTS

We are grateful for feedback from our reviewers and would like to thank the University of Kentucky Office of the Vice President for Research for funding this project through an Igniting Research Collaborations grant to the first and second authors.

7. REFERENCES

[1] Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., Brasher, A. 2013. The time course of perception of coarticulation. *The Journal of the Acoustical Society of America* 133(4), 2350–2366.

[2] Bettens, K., Wuyts, F. L., Van Lierde, K. M. 2014. Instrumental assessment of velopharyngeal function and resonance: A review. *Journal of communication disorders* 52, 170–183.

[3] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer (version 6.0.43).

[4] Chen, M. Y. 1997. Acoustic correlates of english and french nasalized vowels. *The Journal of the Acoustical Society of America* 102(4), 2360–2370.

[5] Cohn, A. C. 1990. Phonetic and phonological rules of nasalization. In: *UCLA Working Papers in Linguistics* volume 76. University of California at Los Angeles.

[6] Derrick, D., Duerr, J., Kerr, R. October 2018. Mask-less oral and nasal audio recording and air flow estimation for speech analysis. *Electronics Letters*.

[7] Hi-res in-ear earbud headphones. https://www.amazon.com/Headphones-Earphones-Smartphone-Nintendo-Chalked/dp/B0799G4ZYX.

[8] Ghio, A., Teston, B. 2004. Evaluation of the acoustic and aerodynamic constraints of a pneumotachograph for speech and voice studies. *International Conference on Voice Physiology and Biomechanics* 55–58.

[9] Hoit, J., Weismer, G. 2018. *Foundations of speech and hearing*. Plural Publishing Inc.

[10] Horiguchi, S., Bell-Berti, F. 1987. The velotrace: A device for monitoring velar position. *The Cleft palate journal* 24(2), 104–111.

[11] MATLAB, 2010. *version 7.10.0 (R2010a)*. Natick, Massachusetts: The MathWorks Inc.

[12] Sjölander, K., Beskow, J. 2000. Wavesurfer - an open source speech tool.

[13] Styler, W. March 2015. *On the Acoustical and Perceptual Features of Vowel Nasality*. PhD thesis University of Colorado at Boulder.

[14] Styler, W. 2017. On the acoustical features of vowel nasality in english and french. *The Journal of the Acoustical Society of America* 142(4), 2469–2482.

[15] Warner, N., Brenner, D., Schertz, J., Carnie, A., Fisher, M., Hammond, M. 2015. The aerodynamic puzzle of nasalized fricatives: Aerodynamic and perceptual evidence from scottish gaelic. *Laboratory Phonology* 6(2), 197–241.