RECORDING AND REPRODUCING SPEECH AIRFLOW OUTSIDE THE MOUTH

Donald Derrick [*1], **Tom De Rybel** [†1], **and Romain Fiasson** [‡1]
[1]University of Canterbury, New Zealand Institute of Language, Brain, and Behaviour

1 Introduction

When we speak, air leaves the mouth and nose with time- and space-varying airflow, air pressure, directionality, and turbulence. Reproduction of which requires an ability to both estimate the airflow from speech and control a source of artificial airflow. Here we present a system that produces artificial airflow with a similar intensity envelope to that produced during speech, as measured 1 cm away from the lips.

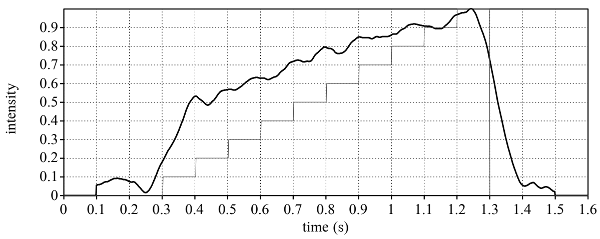

Figure 1: Relationship between direct voltage stimulation of air-pump, 10% increments every 100 ms. Input voltage (thin line). Five Hz low-pass filtered relative air-flow rate (bold line)

Researchers studying the effects of airflow on speech perception began with simple, solenoid-operated, compressor-based airflow device [1]. This system demonstrated that correctly time-aligned airflow enhanced the perception of aspirated stop consonants in syllables (/pa/, /ta/) produced in noisy environments. But our goal is to enhance speech perception during continuous speech, which requires a system able to produce both time and intensity-varying airflow. Our current system uses Murata's *microblower*, a 20x20x1.85mm piezoelectric air pump with up to 0.8 l/m flow, max 19.38 cm/H_2O pressure, and approximately 30ms 5-95% intensity rise time, allowing artificial approximation of continuously-varying airflow in speech [2]. This system is controlled using a circuit that converts a voltage output from a sound-card's right channel into a signal suitable to drive the air-pump, and routes the left channel to both headphone speakers, allowing direct presentation of both audio and airflow to the listener. The system allowed us to uncover speech perception enhancement in fricatives [3] and affricates [4], as well as stops.

However, the control system in use did not allow for a full range of airflow from zero to maximum, as the pump has a minimum airflow output of about 0.4 l/m when active, and exhibited other non-linearities. These effects are largely due to the electro-mechanical nature of the pump itself. Figure 1 shows the relationship between direct input and airflow output from our system. The solution presented here allows control of airflow from 0 to 0.8 l/m of airflow, as directly estimated during speech production.

2 Methods and Results

To begin, we made an airflow estimator system that does not interfere with speech production. The system consists of a ping pong ball mounted on a carbon fiber lever that flexes a thin, flat, carbon-fibre member to which two strain gauges are affixed on opposite sides. These strain gages are part of a sinusoidal-driven bridge, the output of which, after amplification, is an amplitude-modulated sine wave that is can be recorded using a standard XLR based pre-amplifier into a computer. This allows us to bypass the DC blocking circuitry of sound cards for the very low frequency information we intend to measure. The sine wave is then demodulated using an envelope detector and lowpass filtered (100 Hz) in custom software, and a 11.35 Hz resonance is notch-filtered out, along with the second and third harmonic of that resonance. These are due to the mechanical design of the sensor.

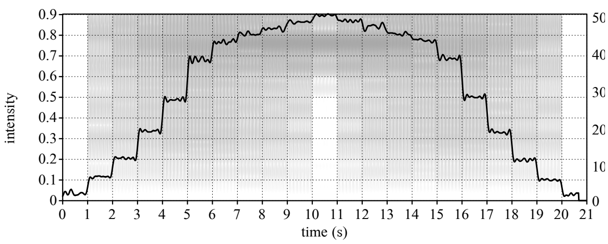

Figure 2: Relative airflow output (bold line) from linear 10 ms period-modulated 12kHz sine-waves sent to the air-pump, stepped at 10% intensity increments (blended background)

To overcome the non-linearities in the air pump output, we first had to overcome the DC blocking behaviour of the sound card. A 12 kHz sinusoidal carrier was chosen for this purpose. The air pump drive circuitry performs an implicit, envelope detection-based demodulation. Second, a form of period modulation was implemented to overcome the 0.4 l/m minimum flow rate. The modulation controls the number of 12 kHz periods during a fixed, 10 ms time window. By driving the pump with these bursts, and relying on the non-negligible quality factor of the piezo resonator in the pump to smooth the air flow, lower flow rates are obtained. Through experimentally-varying the width of these 12 kHz bursts, a 10 ms (100 Hz) burst width was found optimum, corresponding with the 30 ms step response of the *microblower*. We then generated a 21 second pyramid step signal with 1 second long 10% steps from 0 to 100% of signal strength and back. The period-modulated signal was then passed to the air pump. The audio output was recorded with a Sennheiser MKH-146 microphone, and the airflow output recorded using the air-

*. donald.derrick@canterbury.ac.nz
†. tom.derybel@canterbury.ac.nz
‡. romain.fiasson@canterbury.ac.nz

flow estimator. The output of one of the eight recordings can be seen in Figure 2. Although we could cover a much wider range of air flows, some non-linear behaviour remained.

To resolve these nonlinearities, we recorded 8 instances of the output of this 21 second pyramid, measuring the intensities of each step, and fitting the data to polynomial regressions. We found that a 5^{th} order regression was best. The transformed modulation was then applied to a new 21 second pyramid, which was then sent to the air pump. An example of the much more linear output can be seen in Figure 3.

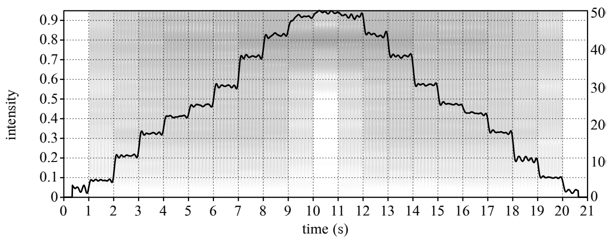

Figure 3: Relative airflow output (bold line) from 5th order polynomial fitted 10ms period-modulated 12kHz sine-waves sent to the air-pump, stepped at 10% intensity increments (blended background)

Our measurements also showed a 30 ms delay from the signal to air-flow from our pump. We used this information to realign, transform, and period-modulate our recorded airflow from speech to generate the signal to drive the air pump. We then played-back a long section of speech, recording the headphone and airflow output. The results show we can reasonably reproduce the envelope of airflow intensity from complex speech acts, such as the word "six" as seen in Figure 4.

3 Discussion

The results show that it is possible to reasonably reproduce the airflow envelope of speech using a low-cost and compact system. However, we do not know how linear the estimates from the airflow estimator are, and instead are relying on the fact that we are using the same measurement system for recording both speech airflow and artificial airflow.

The airflow estimator also has resonances and inertia that cause large impulses of air to appear to have a longer duration than they do. The issue can be seen in Figure 4 after the larger airflows from /s/ and /ks/ where there appears to be a second airflow burst that is produced by the under-damped resonances in the estimator instead of artificial airflow.

The results presented are also scaled, and the actual flow measurement intensity is 1/10th of that in speech. But despite this tiny airflow intensity, our research had already demonstrated that the airflow can alter speech perception, and, from experience, it is quite easy to feel and notice the airflow from this system. During experiments, the full-on flow from the air pump was even deemed uncomfortable by participants. Current research is using this new control system to test for speech perception enhancement of whole sentences.

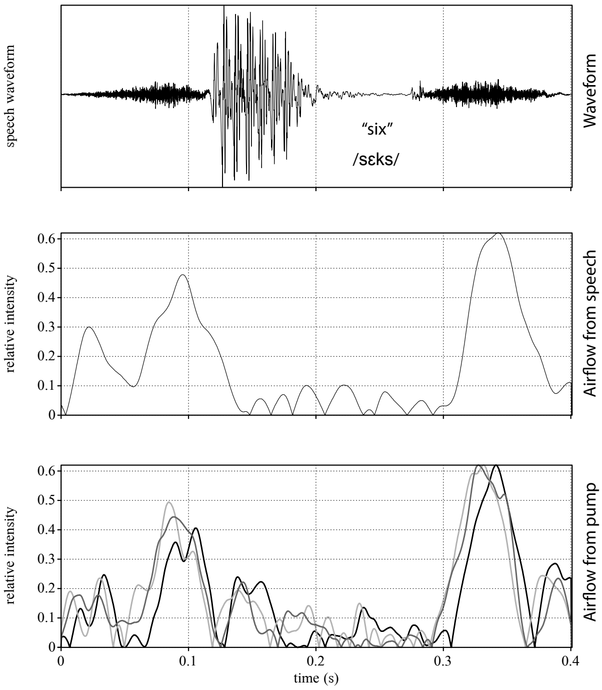

Figure 4: NZE "six" speech waveform (top). Scaled speech airflow from speaker (middle). Three overlapped tokens of scaled speech airflow from air pump generated from modulated input (bottom)

Acknowledgements

Research was funded by two New Zealand Ministry of Business, Innovation and Employment (MBIE) grants for Aero-tactile Enhancement of Speech Perception.

References

[1] Bryan Gick and Donald Derrick. Aero-tactile integration in speech perception. *Nature*, 462 :502–504, 26 November 2009. doi :10.1038/nature08572.

[2] Donald Derrick, Tom De Rybel, Greg. A. O'beirne, and Jennifer Hay. Listen with your skin : Aerotak speech perception enhancement system. In *Proceedings of the 15th Annual Conference of the International Speech Communication Association (INTERSPEECH 2014)*, pages 1484–1485, Singapore, 14-18 September 2014.

[3] Donald Derrick, Greg A. O'Beirne, Tom De Rybel, and Jennifer Hay. Aero-tactile integration in fricatives : Converting audio to air flow information for speech perception enhancement. In *Proceedings of the 15th Annual Conference of the International Speech Communication Association (INTERSPEECH 2014)*, pages 2580–2584, Singapore, September 14-18 2014.

[4] Donald Derrick, Greg A. O'Beirne, Tom De Rybel, Jennifer Hay, and Romain Fiasson. Aero-tactile integration in mandarin syllables. In Prep.

[5] Polynomial regression data fit - aracnoid.com, July 2015.