Mask-less oral and nasal audio recording and air flow estimation for speech analysis

D. Derrick✉, J. Duerr and R.G. Kerr

Here is demonstrated *Rivener*, a mask-less oral and nasal audio recorder and air flow estimation system. This system records audio and low-frequency pseudo-sound from the nares and mouth. The system does not interfere with speech intelligibility, and minimally interferes with visual observation of the speaker. From these recordings, nasalance (a ratio of oral and nasal sound energy), oral air flow, and nasal air flow patterns may be estimated, all while allowing effective clinical observation. The first demonstration is a case-study comparison of the difference between hearing-impaired (HI) speech and non-impaired (NI) speech. Rivener records standard features of HI speech such as: (i) Atypically high or low speech amplitude; (ii) fundamental frequency (pitch) making individual words into intonational phrases; (iii) speech segment substitution; (iv) hypernasalance; (v) atypical air flow in HI speech, including low air flow during plosive release. The second demonstration is a comparison of Rivener and the Rothenberg NAS-1's ability to record nasalance among 26 New Zealand English speakers. The NAS-1 can differentiate low, medium, and high nasalance passages, whereas Rivener differentiates only medium and high nasalance passages consistently.

Introduction: To date there have been no instruments that allow unobstructed observation of a hearing-impaired (HI) speaker while instrumentally recording of all of the major qualities of HI speech. This is because HI speech traits include any or all of: (i) Inappropriate speech volume; (ii) inappropriate prosody; (iii) hypernasality; (iv) reduction of frication during speech; (v) and phoneme alterations [1]. These traits require sound-isolated capture of audio volume and air flow at the lips and nares, requiring the use of an oral and nasal mask.

Oral and nasal face masks muffle the audio component and hide the face such that listener identification of HI speech impediments becomes difficult. Oral and nasal masks are also uncomfortable. Because the masks and air pressure tubes contact mucous membranes, the cost of such equipment increases due to medical device regulatory requirements. Fortunately, a system for recording (HI) speech does not need to be so exact; only requiring a way to record oral and air-flow information at the mouth and nares well enough to capture the features of HI speech.

These traits are captured by recording oral and nasal pressure variation ranging from about 1 Hz through to the speech range of 20 kHz using small unfiltered audio microphones. Separation of the oral and nasal signal is accomplished by directing the relevant energy to individual microphones that are physically isolated from each other. Placing the microphones at the centre of Venturi tubes causes speech air pressure to drop such that the pressure-sensing microphones do not overload from speech with high pressure air flow (especially that of the sound /p/). This lower-pressure direct air flow impacts on the microphone diaphragms and is recorded along with higher-frequency longitudinal sound waves [2].

The low-frequency portion of the audio recording contains broadband pressure fluctuations impinging on the microphone baffle, also known as hydrodynamic noise or pseudo-sound. This low-frequency information allows air-flow estimation; the most useful frequencies for air-flow estimation are below the fundamental frequency of vocal fold vibrations in speech, which is generally not lower than 80 Hz in adult men. These lower frequencies are sufficient as pseudo-sound comes from the speech air flow stream, which typically varies dynamically across acoustic segmental, word, and phrase spans – much longer spans than a typical vocal fold vibratory pulse.

In contrast, frequencies above 80 Hz mostly contain longitudinal sound energy. Therefore microphone recordings unshielded against low-frequency information can contain both auditory and pseudo-sound information, separable using low pass and high pass (HP) filters.

A headset (hereafter referred to as Rivener [3]) was developed to meet the specifications above, as seen in Fig. 1. Rivener uses InvenSense ICS-40300 microphones, with a HP corner frequency of 6 Hz and 130 dB sound pressure level handling. The custom audio interface itself has a HP corner at 0.12 Hz, such that it does significantly impact the low-frequency response of the microphones. We demonstrate the use of Rivener in a case study comparing a HI and non-impaired (NI) speaker, and a comparison of Rivener's nasalance recordings to the Rothenberg NAS-1 system [4].

Methods – case study: Participant 1 was a 51-year-old woman from New Zealand and born with a hearing impairment. She has no hearing in the right ear, but retains 85% of hearing in the left ear, with profound hearing impairment in the lower frequencies.

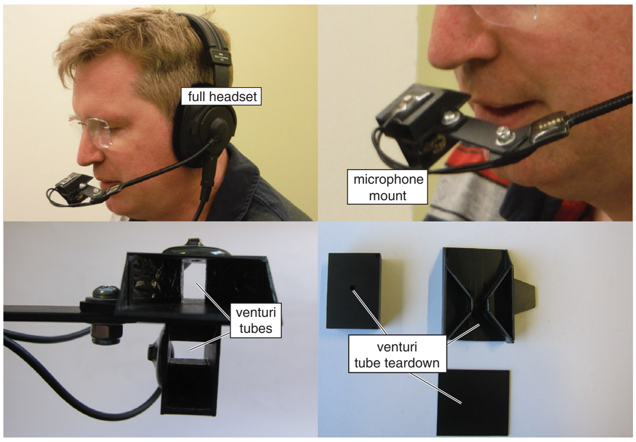

Fig. 1 *Photos of Rivener, showing headset, microphones, microphone placement within Venturi tubes, and a teardown of those Venturi tubes*

Participant 2 was a 22-year-old male. He was born and raised in New Zealand. Although both of his parents are HI, he learned English as his first language and then learned New Zealand sign language. He is fluent in both and has no hearing or speech impairment.

Participants read from the caterpillar passage [5]. The caterpillar passage is written to encourage speakers to read using differing intonation, prosody and speed, throughout the passage, highlighting the key areas of interest in HI speech.

The procedure was explained to our participants, informed consent obtained, and participants seated with the reading passage. The NI speaker translated the instructions into New Zealand Sign Language for the HI speaker. The participants were given as much time as they needed to read the passage to themselves and familiarise themselves with the vocabulary. Once they indicated that they are ready to read out loud, we then had the participant put on the Rivener headset and began recording.

For data analysis, intensity was obtained using PRAAT's intensity algorithm [6]. Nasalance was calculated through this formula:

$$\frac{A_n}{A_m + A_n} 100 \qquad (1)$$

where A, or amplitude, is the auditory signal amplitude from 30 Hz (to cut-off interference from pseudo-sound) to 12,000 Hz (to cut off high-frequency machine/room noise) filtered using a fourth-order Butterworth band-pass filter. An is the amplitude at the nares, and Am is the amplitude at the mouth.

Pitch was tracked using PRAAT/s pitch tracker [6] set to minimum 75 Hz, maximum 350 Hz, to analyse and compared the examples of pitch control between the two speakers.

Oral air flow estimation was obtained by low-pass filtering the audio signal recorded at the mouth using PRAAT's built-in Hann filter to include data between 1 and 70 Hz, with 1 Hz tails.

Results – case study: The results in Figs. 2 and 3 show examples of all of the features of HI speech. In this case study, it is easy to see that overall speech intensity is higher in 1b than 1a, and that the speech is segmented into words in 1a instead of phrases as in 1b. The fundamental frequency in 2a shows instability and is shaped around the word, whereas the flat example in 2b is indicative of speech during declarative reading. The example in 3a shows phoneme substitutions that do not exist for the NI speaker in that 'so' sounded like 'show'. The nasalance for the HI speaker (4a) is much higher (mostly higher than 40%) than that for the NI speaker (4b). (Note that Fig. 2 occludes nasalance during pauses.) Lastly, 5a shows a lack of air flow increase during /p/ release, which normally has the largest air flow of any segment in speech, as seen in 5b, indicating the nasal passage was open and releasing air pressure.

Methods – nasalance comparison: 24 participants were recorded (10 male, 14 female) using both the Rothenberg NAS-1 mask [4] and the Rivener nasalance system [3]. Participants were seated in a quiet room, and asked to read from each of five reading passages: The zoo passage [7], which contains words that are very low in nasal airflow, the rainbow [7], caterpillar [5], and grandfather [7] passages, which each contain a relatively normal amount of nasal airflow, and the nasal sentences [7] passage, which contains words with heavy nasal airflow. They read these passages while wearing the NAS-1 Rothenberg mask, and while using the headset-mounted Rivener recording system.

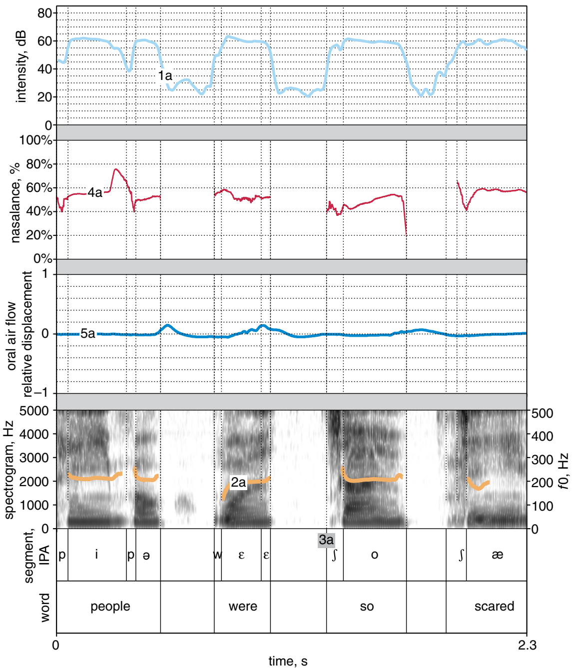

Fig. 2 *Participant 1: HI speech*

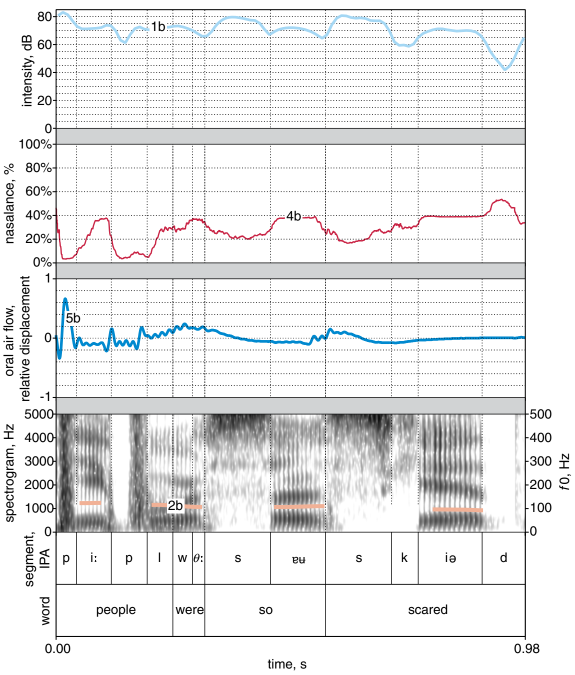

Fig. 3 *Participant 2: NI speech*

The auditory signals were recorded using customised recording software designed to capture from 0 to 22,100 Hz. Nasalance was then computed using the methods described with formula 1.

Results – nasalance: For each participant, the nasalance averages from the Rothenberg NAS-1 mask recordings were always lower for the low nasal passage compared to the three normal nasal passages, and highest for the high nasal passage, forming a clear categorical difference. With Rivener, the nasalance was always lower for the three normal nasal passages than for the nigh nasal passage. However, the low nasal passage only had lower nasalance than the middle nasalance passages for half (12) of the participants. Averages can be seen in Table 1.

Table 1: Rivener and Rothenberg's nasalance means (standard deviation inbrackets), by passage

Reading passage	Rivener	Rothenberg
Zoo	29.8% (4.4%)	16.7% (1.9%)
Caterpillar	31.9% (2.7%)	27.4% (3.8%)
Grandfather	31.9% (3.4%)	27.7% (3.4%)
Rainbow	32.6% (3.1%)	29.5% (3.4%)
Nasal	44.3% (4.5%)	46.9% (4.2%)

Discussion: The results of the case-study show that Rivener records the five main traits of HI speech, including components from audio, air flow, and nasalance. The results of the experiment show that Rivener provides a comparable measure of nasalance to that of the masked Rothenberg system for the normal (rainbow, grandfather, and caterpillar) and high nasality (nasal) passages, but not the low nasality (zoo) passage. Since Rivener cannot distinguish between low and normal nasalisation in NI speech as well as the Rothenberg system does, it is very important to use a high-nasality passage (i.e. nasal) to identify hyponasality, and a low nasalance passage (i.e. zoo) to identify hypernasality.

Nevertheless, the results are very positive. Rivener instrumentally captures the qualities of hearing-impaired speech while preserving speech quality. With simple real-time displays of audio, air flow and nasalance, Rivener can be used in biofeedback, which itself has been used effectively in speech therapy [8], meaning that the output of this system is potentially very helpful in clinical and self-teaching settings.

Acknowledgments: Research was funded by a New Zealand Ministry of Business, Innovation and Employment (MBIE) grant ONT-30003-HVMSSI-UOC for Aero-tactile Enhancement of Speech Perception. Thanks to Western Sidney University's MARCS institute for loan of the Rothenberg NAS-1 system, Tom De Rybel for co-invention of Rivener, and Matthias Heyne and Ruth Chaim for data collection.

Submitted: *29 May 2018*
doi: 10.1049/el.2018.5567
One or more of the Figures in this Letter are available in colour online.

D. Derrick (*New Zealand Institute of Language, Brain, and Behaviour, University of Canterbury, Christchurch, 8140, New Zealand*)

✉ E-mail: donald.derrick@canterbury.ac.nz

J. Duerr (*Speech and Hearing Sciences, Department of Medicine, University of North Carolina, Chapel Hill, NC, 27516, USA*)

R.G. Kerr (*Department of Linguistics, University of Canterbury, Christchurch, 8140, USA*)

D. Derrick: Also with Western Sydney University, MARCS Institute, NSW, Australia

References

1 Ling, D.: 'Speech and the hearing-impaired child: theory and practice' (The Alexander Graham Bell Association for the Deaf, Inc., Washington, DC, USA, 1976)

2 Ffowcs Williams, J.E.: 'Hydrodynamic noise', *Annu. Rev. Fluid Mech.*, 1969, **1**, pp. 197–222

3 Derrick, D., and De Rybel, T.: 'Maskless speech airflow estimation system'. PCT patent number WO 2018/021920A1, 2018

4 Rothenberg, M.: 'Pneumotachograph mask or mouthpiece coupling element for airflow measurement during speech or singing'. US Patent # US5454375A, 1995

5 Patel, R., Connaghan, K., Franco, D., *et al.*: 'A novel reading passage for assessment of motor speech disorders', *Am. J. Speech-Lang. Pathol.*, 2013, **22**, pp. 1–9

6 Boersma, P., and Weenink, D.: 'Praat: doing phonetics by computer', 2008, Version 5.0.27

7 Fletcher, S.G., Adams, L.E., McCutcheon, J.J., *et al.*: 'Communicative disorders related to cleft lip and palate, ch. cleft palate speech assessment through oronasal acoustic measures' (Little, Brown, Boston, 1989), pp. 246–257

8 Murray, E.S.H., Mendoza, J.O., and Gill, S.V., *et al.*: Effects of biofeedback on control and generalization of nasalization in typical speakers', *J. Speech Lang. Hear. Res.*, 2016, **59**, pp. 1025–1034