THE EFFECT OF THE FLOW MASK ON PHONATION

R. Orr and B. Cranen

Department of Language and Speech, Radboud University Nijmegen, The Netherlands

The aim of this study was to investigate whether the presence of the flow mask affects voice quality. Microphone and flow recordings were inverse filtered and were compared to examine the possible effects of the flow mask. Closing quotient (ClQ), open quotient (OQ) and the amplitude difference between the first and second harmonics (H1-H2) were the parameters used to characterise the inverse filtered signals. The presence of the flow mask used for the recording of oral flow had an effect on these parameters, which is interpreted as being indicative of a more tense or more efficient voice quality in the presence of the mask.

I. INTRODUCTION

One way of obtaining objective measures to characterise phonation is to inverse filter either of the oral flow or the sound pressure wave, e.g. [1,2,3]. The former is registered with a flow mask [1] and the latter with a pressure sensitive microphone. Similar parameters are commonly extracted from both, with the exception of DC flow, which cannot easily be measured from microphone recordings.

Each type of voice recording has its own advantages and drawbacks. Microphone recordings are non-intrusive, may produce more natural voicing, and are more practical for field-work, but do not easily provide a measure of DC flow. Flow recordings do provide DC flow, but the experimental setup is more complicated, and the flow mask may affect voicing behaviour. Obvious practical advantages make the microphone the instrument of choice for speech recordings in voice research. Furthermore, the information gained from the DC flow measure is not fully understood.

DC flow has been investigated by numerous researchers, e.g [4.5.6], and is used in some measures of breathiness and vocal efficiency [4,6]. However, the precise relationship of DC flow to voice quality is unclear. Large values of DC flow indicate insufficient glottal closure extending into the membranous part of the vocal folds during maximum glottal closure, while small values [6] may indicate glottal opening in the cartilaginous part of the vocal folds during maximum closure. In [7], it is suggested that there may be two types of incomplete glottal closure, each of which has different implications for the shape of the glottal waveform.

A glottal chink in both the membraneous and cartilaginous parts of the vocal folds (diag. a in Fig. 1), may indicate more gradual glottal opening/closing, leading to a more sinusoidal waveform. When the glottal chink is found in only the cartilaginous parts of the vocal folds, the glottal waveform may show abrupt changes in flow, despite presence of DC flow (diag. b in Fig. 1). Thus, DC flow may
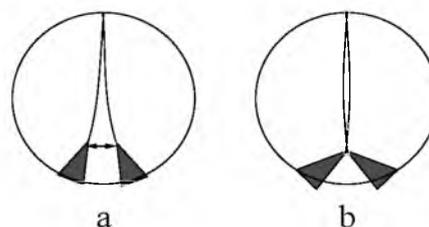

Fig. 1. Two ways to model glottal leakage: a) a linked leak created by abduction and b) a parallel chink in the cartilaginous portion of the glottis. Taken from [7].

not be an accurate means of determining voice efficiency. It is even suggested [6] that small amounts of DC may be due to vertical phasing as a result of a mucosal wave, in voices where there is complete closure.

As long as the relationship between DC and the glottal waveform remains unclear and even contradictory, there is no reason to prefer flow recordings over microphone recordings for voice analysis, and it may be preferable to focus on the parameters which do indicate a consistent relationship. This line of reasoning clearly depends on the validity of the assumptions made about the relationship between the flow and microphone recordings.

Theoretically, parameters extracted from either type of recording should represent the same information [8]. In earlier work [9], flow derivative and sound pressure were compared for a group of 70 subjects. The recordings were made in a loosely controlled situation, where subjects were asked to phonate as naturally as possible. The recordings were inverse filtered, and source parameters were extracted from each voicing condition for each subject. Since the subjects produced the two recordings within a relatively short period of time of about five minutes, large between-session variation was not expected.

The results of a microphone/flow comparison were not similar. When the data for each subject were analysed, the flow and microphone parameter values did not correlate. Moreover, analyses of variance indicated that there was a difference between flow and microphone results. A number of possible reasons were suggested for this. Normal within-subject variation may be large enough [4] that direct comparisons of separate utterances for the same speaker cannot be made in a loosely controlled experimental setup. Subjects were sometimes perceived to be uncomfortable with the mask, and this may have introduced physical tension in the voicing apparatus in general. The acoustic distortion produced by the mask may also have an effect on the data. Auditory feedback for the subject wearing the mask

is muffled, and this could also lead to some change in voicing strategy in phonation production in the presence of the mask.

While it is very difficult to make direct comparisons on different utterances, such large differences between flow and microphone data were not expected. It was therefore considered worthwhile investigating these differences in more detail. A more thorough understanding of the comparability of flow and microphone signals is important, as research on the voice source in the speech community comes from both flow and microphone data, and both are used to characterise voice quality for many purposes. A single subject experiment was designed to give maximal control over possible confounding factors suggested from the results of the previous work and we present here the data that was collected from such a setup.

II. METHODOLOGY

Subject: The subject was a female phonetician with experience in producing experimental speech, and who was familiar with the aims of the experiment. A single subject was chosen for this analysis for three reasons.

Firstly, laryngeal settings for normal voice production can differ considerably from subject to subject. Distinctive individual differences within a subject group may make the comparison and interpretation of mean scores spurious for this particular investigation.

Secondly, a speaker with some experience and understanding of the production of different voice qualities was required. In order to get an impression of whether the inverse filtering produced realistic parameter values for both flow and microphone utterances, the subject was required to produce three different voice qualities for which relative values are already established in other published research. If the relative values produced for the different voice qualities concur with those of other research, this would help to confirm that the signal processing procedure was robust.

Thirdly, it was necessary to control phonation in order to limit within-speaker variability as much as possible. The production of voice tokens which are as similar as possible requires insight and control which naïve subject groups may not have. We therefore wanted a speaker who was properly trained in the area of voice production.

Although the results of a single subject experiment cannot be generalised to the population as a whole, if a single trained speaker does not produce comparable voicing behaviour for mask and no-mask conditions, then we would reason that an untrained speaker is even less likely to do so.

Phonation Task: The subject produced the utterance / paepaepaepae/, at a rate of about 1.5 syllables per second, using *modal* voice, and also using assumed *breathy* and *creaky* voice. For each voice quality, 20 repetitions of the utterance were recorded first with a pressure sensitive microphone and then with a Rothenberg mask. In total, 40

utterances in each voice quality were recorded. A voice therapist was present during all recordings to ensure that the required voice qualities were actually produced. Fundamental frequency was kept constant at around 173 Hz, using a tuning fork for reference at the beginning of each sequence of utterances.

Measurements: Microphone recordings were made with a Bruel and Kjaer (B&K) microphone (4133) at approximately 10cm from the mouth and a B&K amplifier 2619.

Oral flow was measured with a circumferentially vented pneumotachograph mask (Glottal Enterprises) with a heated double screen wire mesh, in combination with a Glottal Enterprises amplifier (MS-100A2). Before and after the flow recordings, the flow sensors were calibrated in order to get absolute flow measures and to ensure the consistency of the measurements.

The signals were recorded simultaneously on an analogue 14-channel FM-recorder (TEAC XR510). The recordings were made at a tape speed of 19.05 cm/s allowing a flat frequency response up to 5kHz. The microphone signals were recorded on 3 different channels with low, medium and high input gains. In this way, at least one version of each signal would have an acceptable SNR. Flow signals were similarly recorded at two different levels on two different channels.

Signal Processing: All signals were synchronously digitised at a 10kHz sampling rate per signal and then demultiplexed.

The microphone signal, which prior to digitisation had already been filtered by means of an analogue high-pass filter (cut-off frequency 22.4Hz) in the B&K amplifier, was treated with a second digital high-pass filter to eliminate any remaining low frequency distortions, using a linear phase filter and with a cut-off frequency of approximately 20 Hz and a flat frequency response above 70Hz. It was then phase-corrected to compensate for phase distortion introduced by the analogue high-pass filter of the microphone amplifier. This signal was automatically inverse filtered by means of pitch synchronous inverse filtering using covariance LPC on the closed glottis interval (CGI). The start of the CGI was determined from the peak in the EGG derivative. The inverse filtered signal was then low-pass filtered at 1500Hz, again using a linear phase filter.

The calibrated flow signal was inverse filtered in the same way as the microphone signal and low-pass filtered with a linear phase filter with a cut-off frequency of 1500Hz.

Characterisation of the acoustic voice source: OQ and ClQ were extracted from each glottal cycle over the stable stationary parts of the vowel in order to compare results from the most reliably inverse filtered speech samples. A modal value was determined for each utterance. Fig. 2 shows the moments on the source wave which were used to calculate OQ and ClQ.
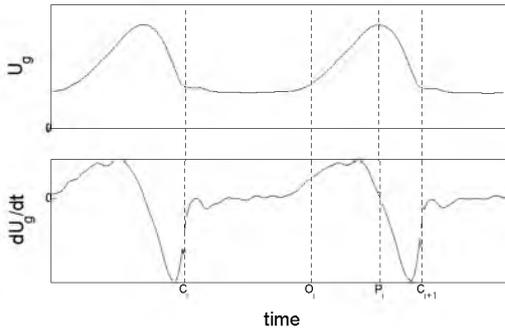
Fig.2 Moments on the glottal flow waveform (upper window_ and flow derivative (lower waveform) from which the time-related parameters OQ and ClQ were derived. OQ is derived as $(C_{i+1} - O_i)/(C_{i+1} - C_i)$ and ClQ is derived as $(C_{i+1} - P_i)/(C_{i+1} - C_i)$

The spectral parameter H1-H2 was calculated from the stable sections at the beginning of the final vowel. Theseselections were divided into equal length sections of 1024 samples. The first harmonic peaks in the spectrum were detected, and their frequencies and amplitudes were recorded. H1-H2 repesents the difference in amplitude (dB) between f_0 and the component with double that frequency. The values used in the comparison were average values from the 1024 sample sections. As a mean value was used, we decided not to include the dying out part of the last vowel, where vocal effort would be reduced such that the laryngeal musculature would relax and produce a less efficient voice. This approach was tested on data from earlier work [9], and the spread of values per glottal cycle was smaller, giving more representative values for the utterance.

III. ANALYSIS

Fig. 3 shows scatterplots of the data separated for flow/microphone recordings (mask/no-mask conditions). The visual data already indicates that the mask has some effect on the source parameters. A power calculation could not be made for the estimation of an appropriate significance level, as there is insufficient normative data to be able to know what constitutes a perceptually relevant difference in our parameters. It was expected that, as the subject concentrated on maintaining a stable voice quality, much of the variance would be attribuatable to the experimental conditions, and the effect of the mask, if present, should be clear. Therefore, a real effect of the flow mask was considered present if an analysis of variance[1] showed a medium effect size, partial eta squared (η_p^2) where $0.15 > \eta_p^2 > 0.06$ [10]where $p \leq 0.01$.

---

[1] Type II ANOVA, calculated according to the principle of marginality, testing each term after all others, ignoring the term's higher order relatives [11, 12]

IV. RESULTS AND DISCUSSION

Relative parameter values mostly concur with other research. *Breathy* voice has larger OQ, smaller ClQ and larger H1-H2 values than *modal* voice. *Creaky* voice shows a greater range of values than *modal* voice and has larger ClQ and smaller H1-H2 values. *Modal* and *creaky* OQ were centred around similar values, but were more spread for *creaky* voice. *Modal* values are close to what has been observed for *pressed* voice. The voice therapist who was present at the recordings confirmed this perceptually . It is reasonable to expect that if the known differences between voice qualities are properly represented, then the unknown effect of the mask will also be properly represented.

Table 1. ANOVA results for effects of the mask factor.

	F ($df = 1$)	p	η_p^2
	Breathy		
H1-H2	0.00	0.96	0.00
ClO	39.69	0.00	**0.42**
OO	98.60	0.00	**0.71**
	Modal		
H1-H2	43.26	0.00	**0.54**
ClO	23.73	0.00	**0.36**
OO	1.53	0.22	0.03
	Creaky		
H1-H2	1.72	0.20	0.04
ClO	114.36	0.00	**0.68**
OO	39.61	0.00	**0.44**

In general, values for the three chosen parameters were lower for *mask* condition than for *no-mask* condition. M*ask* values were less spread than *no-mask* values. Of the nine combinations of voice quality and source parameter, only two showed mask values that do not conform to the overall result, namely H1-H2 for *creaky* and *breathy* voice. One other combination, while following the general trend of the values, did not represent a significant result according to the criteria that we set, namely, OQ for *modal* voice. Although an effect was not demonstrable for these three combinations of DV and *mask* factor, the effect is systematically present for the other six combinations. This is supported by the large effect size for these combinations.

The generally lower H1-H2, ClQ and OQ values could indicate a more efficient phonation. Auditory feedback of muffled speech produced with the mask is a likely cause of this effect. Muffled auditory feedback could also influence the speaker to put more effort into accurately producing the intended voice quality. The smaller spread of mask values may reflect extra focussing on the phonation task. It is interesting to note that the mask seems to have the effect of
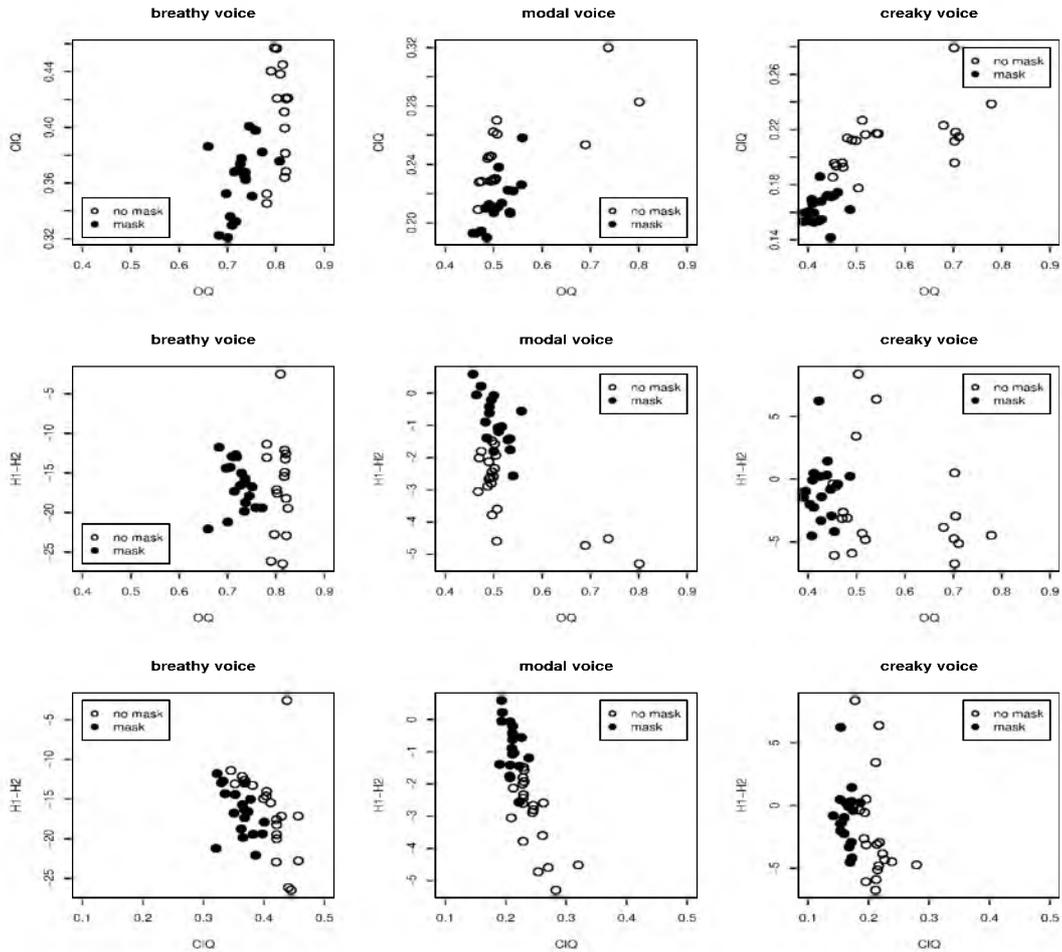
reducing parameter variability on a subconscious level, but the speaker was not able to consciously limit variability.

V. CONCLUSION

There was a systematic difference between source parameters extracted from flow and microphone speech. Flow mask recordings produced lower parameter values than microphone recordings. This may indicate a more tense voice, more efficient voiceing strategy caused by muffled auditory feedback or a combination of both.

REFERENCES

[1] M. Rothenberg, "A new inverse-filtering technique for deriving the glottal flow waveform during voicing", *J. Ac. Soc. Am.*, vol. **56**(6), pp. 1632-1645, 1993.

[2] P. Alku, "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering", *Speech Communication* **11**(2-3), 109-118, 1992.

[3] A. Ní Chasaide & C. Gobl, "Voice Source Variation" in *Handbook of Phonetic Sciences*, J. Laver & W. Hardcastle, Eds. Blackwell. 1995, pp. 427-462.

[4] E. Holmberg, *Aerodynamic Measurements of Normal Voice*, Stockholm University, Sweden, 1993.

[5] J. Iwarsson, *Breathing and Phonation – Effects of Lung Volume and Breathing Behaviour on Voice Function*, Stockholm, Sweden, 2001.

[6] S. Hertegård, *Vocal Fold Vibrations as Studied with Flow Inverse Filtering*, Stockholm, Sweden, 1994.

[7] B. Cranen & J. Schroeter, "Modelling a leaky glottis", *Journal of Phonetics*, 20, pp. 165-177, 1995.

[8] J. L. Flanagan, *Speech Analysis, Synthesis and Perception*, Springer, Berlin, 1972.

[9] R. Orr, B. Cranen & F. de Jong, "An investigation of the parameters derived from the inverse filtering of flow and microphone signals", in *Voice Quality: Functions, Analysis and Synthesis (VOQUAL '03)"*, C. d'Alessandro & K. R. Scherer, Eds. Geneva, Switzerland, 2003, pp. 35-40.

[10] J. Cohen, *Statistical Power Analysis for the Behavioural Sciences*, Lawerence Erlbaum Associates, Hillsdale, NJ, 1988.

[11] J. Fox, *Applied Regression, Linear Models, nad Related Methods*, Sage, 1997

[12] T. Rietveld & R. van Hout, *Statistical Techniques for the Study of Language and Language Behaviour*, Mouton de Gruyter, 1993.