

Source Modeling of Severely Pathological Voices

Bruce R. Gerratt^{*}, Jody Kreiman^{*}, Norma Antonanzas-Barroso^{*}, Brian Gabelman^{*},
and Abeer Alwan^{**}

^{*}*Division of Head and Neck Surgery and* ^{**}*Dept. of Electrical Engineering,*
UCLA, Los Angeles, CA 90095 USA

Abstract: Our previous attempts at synthesizing severely pathological voices were hampered by perceptually-significant errors in vowel quality. To determine whether these were caused by methodological limitations or by source-tract interactions, normal and pathological voices were recorded using both a condenser microphone and a flow mask system. Synthetic versions of each signal were produced and compared perceptually to the original signals. Limitations of all-zero inverse filtering techniques as applied to pathological voices will be discussed.

INTRODUCTION

Attempts at modeling pathological phonation using analysis by synthesis depend on accurate modeling of the pathological voice source and vocal tract resonances. The source function is typically estimated using inverse filtering, either of the glottal flow signal (e.g., 1) or of the acoustic signal as transduced with a condenser microphone (2).

Because of the restricted frequency response of the flow mask system, our preliminary experiments synthesizing pathological phonation used estimates of vocal tract characteristics derived from a microphone signal, combined with estimates of vocal tract resonances derived from a flow mask signal (recorded from a different segment of the same utterance). These studies suggested that the frequency response characteristics of the flow mask system may limit its applicability to severely pathological phonation, particularly when precise source modeling is required. In addition, perceptually-salient changes in vowel quality occurred when we resynthesized voices using formants estimated from the microphone signal and sources estimated from the flow signal. These difficulties may be due to the fact that microphone and flow signals were measured at different points in an utterance, or to the fact that significant source-filter interactions occur in pathological voices (or both). To assess the extent to which flow and acoustic signals differ in their ability to capture source and vocal tract characteristics of pathological phonation, the following experiment was undertaken.

RECORDING METHOD

Eight speakers participated in this experiment. Four (two males, two females) were randomly-selected patients complaining of voice disorders; four (two males, two females) were volunteers free from vocal pathology. For half the subjects, a flow mask (Glottal Enterprises) was placed over the subject's face and a 1" condenser microphone (Bruel & Kjaer) was held a constant distance from the subject's lips outside the flow mask. The subject sustained the vowel /a/ for at least 2 seconds. Approximately half way through the utterance, the flow mask was quickly removed, allowing recording of the acoustic signal without any mask effects. The order of recording was reversed for the remaining subjects, with acoustic signals recorded first and flow mask applied half-way through the utterance. Thus, both flow and acoustic signals were recorded from a single utterance for each subject. All signals were low-pass filtered at 8 kHz and sampled at 20 kHz.

ANALYSIS AND SYNTHESIS TECHNIQUES

Signals were down-sampled to 10 kHz prior to analysis. Two sets of analyses were undertaken. The first (flow mask-microphone condition) used the flow signal to estimate the source function, and the microphone signal to estimate the vocal tract response. To derive the glottal pulse shape from the flow signals, formants and bandwidths for inverse filtering were estimated as follows. The beginning of the closed phase was located through LPC error analysis, and a covariance LPC analysis was performed (30 to 40 samples, order 12 or 14). If this procedure failed, or if there was no closed phase, an autocorrelation LPC analysis was performed instead (256 points, order 12 or 14). The flow signal was then inverse filtered using the all-zero filter method described in (3).

In the second condition (microphone only condition), inverse filtering was performed directly on the acoustic signal recorded with the condenser microphone. Formants and bandwidths for inverse filtering were estimated as described above, but from the acoustic recording. In both conditions, formants and bandwidths were manipulated interactively during inverse filtering to produce the "best" result possible. Our primary criterion for success was a smoothly decreasing source spectrum slope.

Formant frequencies and bandwidths for synthesis were estimated in both conditions from the microphone signals using autocorrelation LPC (256 points, order 12 or 14), supplemented with spectrographic analysis. Note that for the flow mask-microphone condition, the source and formant analyses were performed on different segments of a single utterance. For the microphone-only condition, both analyses were performed on the same segment of speech.

Three different versions of each voice were synthesized for each condition: a single pulse extracted from the original voice signal and concatenated to form a 1 second stimulus to avoid the perceptual effects of vocal perturbation and other sources of noise, a "reconstructed" version comprising the source function produced from inverse filtering (of the flow signal in the flow mask-microphone condition, or of the microphone signal in the microphone only condition) and the vocal tract resonances estimated from the microphone signal; and an "LF-fitted" synthesized version created by combining the estimated vocal tract resonances with a source that had been least-square fitted with a simplified LF source model (4).

PERCEPTUAL EVALUATION

Ten expert listeners assessed the similarity of each reconstructed and synthetic stimulus to the original (natural) voice on a visual analog scale. Perceptual analyses will compare the success of modeling normal and pathologic phonation and male and female voices, as well as the relative success of modeling efforts using the flow mask-microphone and condenser microphone only techniques.

ACKNOWLEDGMENT

This research was supported by NIDCD grant DC01797.

REFERENCES

1. Rothenberg, M., *J. Acoust. Soc. Am.* **53**, 1632-1645 (1973).
2. Ananthapadmanabha, T.V., *STL-QPSR* **2/3**, 1-24 (1984).
3. Javkin, H., Antonanzas-Barroso, N., and Maddieson, I., *J. Speech Hear. Res.* **30**, 122-129 (1987).
4. Qi, Y., and Bi, N., *J. Acoust. Soc. Am.* **96**, 1182-1185 (1994).