Improved glottal inverse filtering by Markov chain Monte Carlo methods

Samuli Siltanen
Professor of Industrial Mathematics
Department of Mathematics and Statistics
University of Helsinki, Finland
5th International Congress of World Voice Consortium, Luxor, Egypt
October 30, 2012
This work is part of the CSI Speech project, funded by Academy of Finland

Prof. Laukkanen (vocology)
Prof. Kaasalainen (inversion)
MSc Liu (inversion)

Doc. Kolehmainen
PhD Nissinen
(diffuse tomography)

Prof. Alku (acoustics)
Prof. Ilmoniemi (medical imaging)
Doc. Kurimo (speech processing)
Doc. May (cognitive neuroscience)
MSc Raitio (acoustics)
Prof. Sams (computational science)
Doc. Tiitinen (cognitive neuroscience)
1. Glottal inverse filtering (GIF)

2. Markov chain Monte Carlo (MCMC) method

3. Results

4. Conclusion
The goal of Glottal Inverse Filtering is to recover the glottal excitation and the vocal tract filter.

**Direct problem:** If we know the glottal excitation signal and the shape of the vocal tract, what does the microphone record?

**Inverse problem:** Given a speech signal recorded by a microphone, find the glottal excitation and the vocal tract.
An improved GIF algorithm has important applications

1. **Computational speech synthesis:**
   Clearer information announcements, more efficient automatic telephone-based services, and devices that help handicapped people express emotions.

2. **Noise-robust automatic speech recognition:**
   Efficient and reliable man-machine interfaces.
1. Glottal inverse filtering (GIF)

2. Markov chain Monte Carlo (MCMC) method

3. Results

4. Conclusion
We use the Klatt model for the glottal excitation.

The Klatt model has only one parameter called $K$, with values between zero and one.
We use the Klatt model for the glottal excitation. The Klatt model has only one parameter called K, with values between zero and one.
We use the Klatt model for the glottal excitation.

The Klatt model has only one parameter called $K$, with values between zero and one.
We model the effect of the vocal tract by a linear filter, much like a frequency equalizer.
Technically, the frequency response is described by an all-pole filter.
Looking at two first formants leads to four parameters for the vocal tract.

- Angle and length of the pole corresponding to the first formant.
- Angle and length of the pole corresponding to the first formant.
In this model, glottal inverse filtering means finding these five numbers:
But where does Monte Carlo come in the picture?
The Markov chain Monte Carlo method produces a long sequence of excitations and vocal tracts. The choosing of the members in the above chain is done in a controllably random way. As a result, the chain explores all combinations of glottal excitation and vocal tract filter that:

1. Produce closely the measured signal, and
2. Satisfy our a priori information (for example, if we know the vowel, we have a rough idea where the main formants are located).
M=40000 and N=10000, so we average over 30000 members in the chain.
1. Glottal inverse filtering (GIF)

2. Markov chain Monte Carlo (MCMC) method

3. Results

4. Conclusion
We perform glottal inverse filtering with two methods and compare the results

First, initial estimates of the vocal tract and glottal flow are evaluated by iterative adaptive inverse filtering (IAIF), see [Paavo Alku 1992].

Starting with the IAIF result as an initial point, MCMC refines the GIF model parameters in order to get optimal inverse filtering.

We apply the method to three sustained vowel signals collected from three Finnish female students.
Case 1 (Julia): $f_0=237$, frequency response by **IAIF** and **MCMC-GIF**
Case 1 (Julia): $f_0=237$, Klatt pulse recovered by MCMC-GIF

Inverse filtered results by **IAIF** and **MCMC-GIF**
Case 1 (Julia): $f_0=237$, Klatt pulse recovered by MCMC-GIF

Inverse filtered results by IAIF and MCMC-GIF

![Graph of Case 1 (Julia)](image)

![Graph of Inverse filtered results by IAIF and MCMC-GIF](image)
Case 2 (Tiina): $f_0=194$, frequency response by IAIF and MCMC-GIF
Case 2 (Tiina): $f_0=194$, Klatt pulse recovered by MCMC-GIF

Inverse filtered results by **IAIF** and **MCMC-GIF**
Case 2 (Tiina): \( f_0 = 194 \), Klatt pulse recovered by MCMC-GIF

Inverse filtered results by IAIF and MCMC-GIF
Case 3 (Taina): $f_0=197$, frequency response by IAIF and MCMC-GIF
Case 3 (Taina): $f_0=197$, Klatt pulse recovered by MCMC-GIF

Inverse filtered results by **IAIF** and **MCMC-GIF**
Case 3 (Taina): $f_0=197$, Klatt pulse recovered by MCMC-GIF

Inverse filtered results by **IAIF** and **MCMC-GIF**
For more details, and for MCMC-GIF results on synthetic voice, see

**Auvinen, Raitio, S and Alku:**
*Utilizing Markov Chain Monte Carlo (MCMC) Method for Improved Glottal Inverse Filtering,*
to appear in the proceedings of Interspeech 2012.
1. Glottal inverse filtering (GIF)

2. Markov chain Monte Carlo (MCMC) method

3. Results

4. Conclusion
Conclusion

MCMC-GIF can improve glottal inverse filtering over existing methods, especially for high-pitch voice.

We used a quite simple model for the glottal excitation, and modelled the vocal tract linearly using two free formants. Allowing more complicated models will improve the GIF.

The computational cost of MCMC is very high, so it is first applicable only in situations that are not time-critical.

The MCMC approach to GIF is very general and can possibly be extended to various pathological voices as well.