Glottal inversion with an approximate vocal tract filter

Lasse Lybeck, Robert Sirviö
University of Helsinki, Department of Mathematics and Statistics

UNIVERSITY OF HELSINKI

Introduction

- A synthetic human vowel sound consists of a periodic signal to simulate the glottal excitation signal at the glottis and a filter to simulate the vocal tract, which the glottal signal is filtered through.[6] With a given vocal tract filter the direct problem is *given a glottal excitation signal, create the vowel sound*. The inverse problem is *given a (recorded) vowel sound, find the glottal excitation signal*. We will be concentrating on the inverse problem using simulated vowel data.
- The inversion from a vowel sound to the glottal signal is an important part of creating synthetic human voices and speech generators. To create a synthetic vowel both the glottal signal and the vocal tract filter are needed. However, the glottal signal cannot be directly measured, but it can be approximated with inversion of a recorded vowel. This data can be used to create models for simulating the glottal excitation signal.

Materials and methods

- The *Rosenberg-Klatt model (RK-model)* [8] was used to simulate the *glottal excitation signal*. The parameters used by the model are the sound frequency f and the *klatt-parameter* Q. An example of the resulting airflow and pressure can be seen in figure 1.

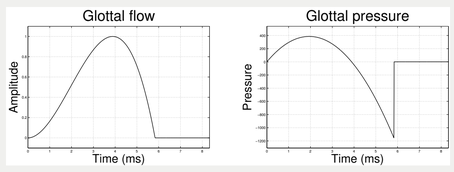

Figure 1: The airflow and pressure generated by the RK-model

- The measurement data for the inversion was created synthetically using the RK-model and a digital filter for a (female) vowel /a/. Some random errors were added to the data to simulate measurement noise.
- The vocal tract filter used in the inversion is a digital filter simulating the (male) vowel /a/. A matrix corresponding to the filter was constructed to arrive at a matrix model $m = Ag + \varepsilon$ for the vowel simulation.
- Even though the filter used in the inversion is not the same as the one used in simulating the data, they were assumed to simulate roughly the same sound when applied on the glottal impulse (the vowel /a/). Therefore the inversion could be done without the fear of inverse crime, while still expecting good results.
- The inverse problem was solved using the *Tikhonov regularization* with a customized penalty matrix. The regularization parameter was chosen using *Morozov's discrepancy principle*.

Results

- Figure 2 shows a naïve reconstruction attempt of the glottal excitation signal. In figure 3 the same inverse problem is solved with Tikhonov regularization using the regularization parameter $\alpha = 63.2$ attained with Morozov's discrepancy principle. The relative error for the naïve reconstruction was $\delta_{\text{rel}} \approx 219\%$, and for the regularized solution $\delta_{\text{rel}} \approx 63.1\%$. The measurement data was created with the parameter values $f = 90$ Hz and $Q = 0.7$.
- The relative error of the reconstruction with different values of the regularization parameter α is shown in figure 4. The least error $\delta_{\text{rel}} \approx 64.6\%$ was attained with the value $\alpha = 201$. Figure 5 shows the reconstruction of the same data done with Tikhonov regularization. The value of the regularization parameter $\alpha = 104.5$ was chosen with Morozov's discrepancy principle, and resulted in a relative error $\delta_{\text{rel}} \approx 65.0\%$. The measurement data was created with the parameter values $f = 120$ Hz and $Q = 0.5$.
- The relative error is calculated with the formula

$$\delta_{rel} = \frac{\|g - T_\alpha(m)\|_2}{\|g\|_2} \cdot 100\%, \qquad (1)$$

where $g \in \mathbb{R}^k$ is the original glottal excitation signal and $T_\alpha(m)$ is the reconstruction calculated from the noisy data m with the regularization parameter α.
- In the figures 2, 3 and 5 the green line shows the original glottal signal and the blue line shows the reconstruction.

Results: Figures

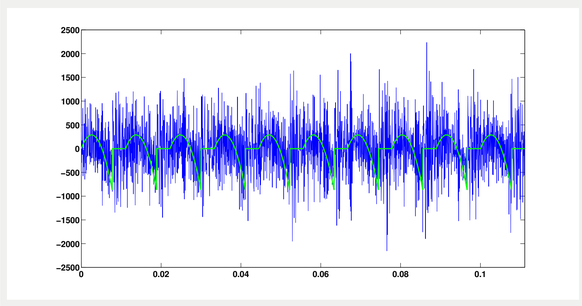

Figure 2: Naïve inversion

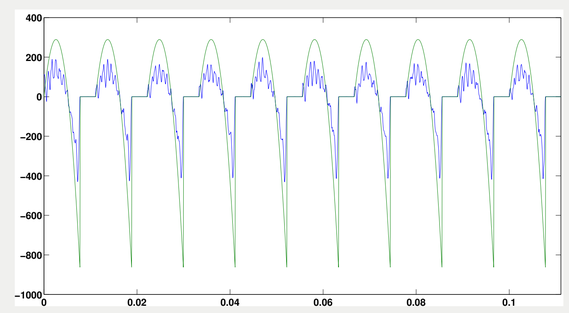

Figure 3: Tikhonov regularized solution

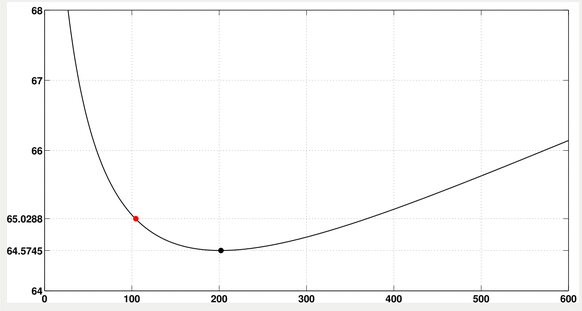

Figure 4: Relative error (%) by α-value

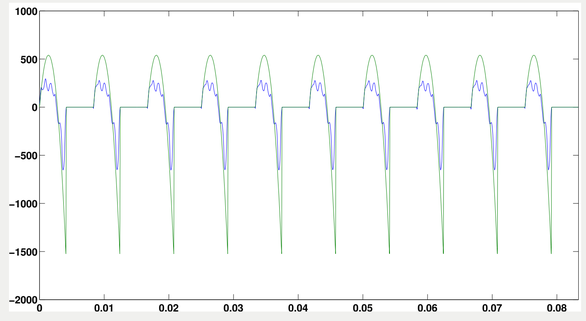

Figure 5: Tikhonov regularized solution

Discussion

- As can be seen from figure 2, the inverse problem cannot be solved with naïve inversion, as it fails to achieve any proper results. The problem is clearly *ill-posed*, and requires some regularization method to achieve reasonable results.
- The results acquired with the Tikhonov regularization strategy were good; the shape of the reconstruction is clearly the same as the original data, although the reconstruction is somewhat flattened. This can be seen from figures 3 and 5. Although the reconstructions were flattened, it does not affect much the generated *sound*, only its amplitude, as long as the shape of the pulse is correct. Therefore the relative error described in (1) might not be the correct number to show the quality of the reconstruction, but it could rather be displayed by a number comparing the *shapes* of the impulses. However, it turned out that this way of comparing the reconstruction with the original data yielded similar results to the relative error.
- The regularization parameters acquired with Morozov's discrepancy principle yielded good results. As can be seen from figure 4, the reconstruction with a regularization parameter acquired with Morozov's discrepancy principle gave almost as good a reconstruction as with the best possible value of the parameter, when using the relative error of the reconstructions as a reference point. Considering that the selection of the regularization parameter with Morozov's discrepancy principle works completely without the need of human interaction, the method selection can be regarded as a success for this particular problem.
- As this study is about generating sounds it is not reasonable to concentrate solely on graphs and error values, but also on how the the reconstruction sounds to a human ear. Because this cannot be measured in any way, it is up to each person to decide if the result is acceptable or not. Please ask the authors to play some sound samples from the reconstructions!

Bibliography

[1] Abramowitz, M & Stegun, I.A. (1965) *Handbook of mathematical functions with formulas, graphs and mathenatical tables* New York, NY: Dover, p.940

[2] Fant, G., Liljencrants, J., Lin, Q., (1985). *A four-parameter model of glottal flow*. STL-QPSR 26 (4), p. 1-13

[3] Fujisaki, H., Ljungqvist, M., 1986. Proposal and evaluation of models for the glottal source waveform. In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Vol. 11. p. 1605–1608.

[4] Mueller, Jennifer L. & Siltanen Samuli, (2012). *Linear and Nonlinear Inverse Problems with Practical Applications*. SIAM, 1st edition.

[5] Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. (2007) *Numerical Recipes: The Art of Scientific Computing*. New York: Cambridge University Press, 3rd edition, p.362

[6] Touda, K., (2007) *Study on numerical method for voice generation problem*. PhD thesis. The University of Electro-Communications.

[7] Rabiner, L. R., Schafer, R. W., (1987). *Digital processing of speech signals*. Englewood Cliffs: Prentice-Hall, p. 38-107.

[8] Rosenberg, A., (1971). *Effect of glottal pulse shape on the quality of natural vowels*. Journal of the Acoustical Society of America 49 (2B), p. 583–590.

[9] Evans, M., Hastings, N. & Peacock, B. (2000) *Statistical distributions*. New York: Wiley, p. 57

lasse.lybeck@helsinki.fi, robert.sirvio@helsinki.fi