Recording and post-processing speech signals from magnetic resonance imaging experiments

Theoretical and practical approach

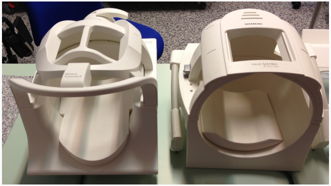
Juha Kuortti and Jarmo Malinen

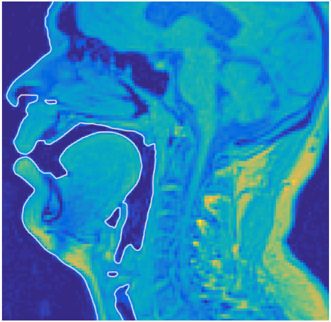
November 28, 2017

Aalto University
juha.kuortti@aalto.fi, jarmo.malinen@aalto.fi

The Measurement

Data acquisition: Magnetic Resonance Imaging

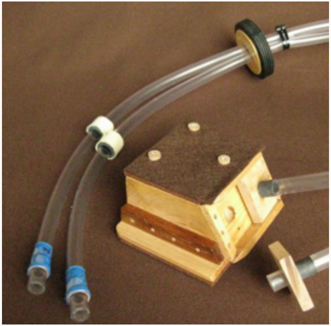


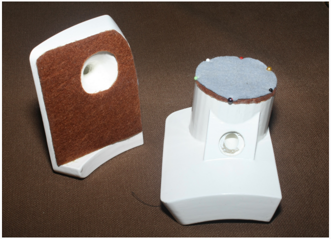


- Non-invasive, safe 3D imaging method.
- Strong electromagnetic fields make sound recordings during imaging difficult.

A. Ojalammi, J. Malinen Automated Segmentation of Upper Airways from MRI: Vocal Tract Geometry Extraction, BIOIMAGING 2017, 77-84

Data acquisition: Sound in MRI



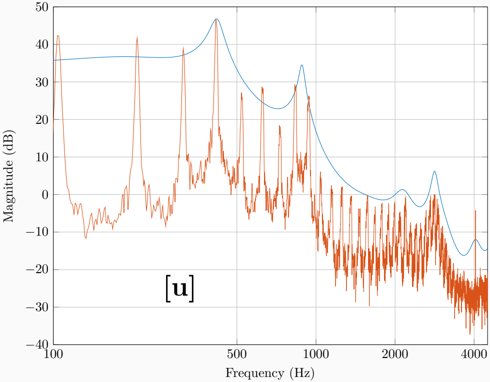


- MRI scanner itself produces about 90 dB(SPL) of noise that will be present in the speech sample.
- Record both speech and noise for post-processing with custom MRI-proof dipole sound collectors.
- Keep the actual microphones away from the scanner; acoustic waveguides transfer sound.

The Target

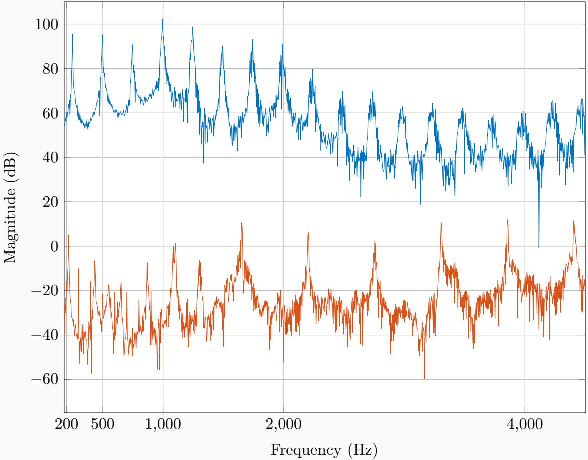
The goal

The goal of the algorithm is not (necessarily) to remove all noise. Rather, we seek to retain and *accurately* measure appropriate characteristics of the signal in its *spectral envelope*.



Why is this not trivial?

The pure MRI noise (red) and pure speech have intertwining spectral peaks, making direct spectral subtraction or AEC difficult.



Looking at the data: magnitude spectra

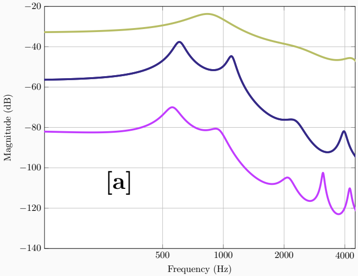


Figure 2: Spectral envelopes of Finnish vowels [ɑ, i] from a male subject.
Top curves: Without post-processing, recorded during MRI.
Middle curves: Post-processed by the proposed method.
Bottom curves: Optimal recordings in anechoic room.

Sound is heard, rather than seen

Looking at the data: spectrogram

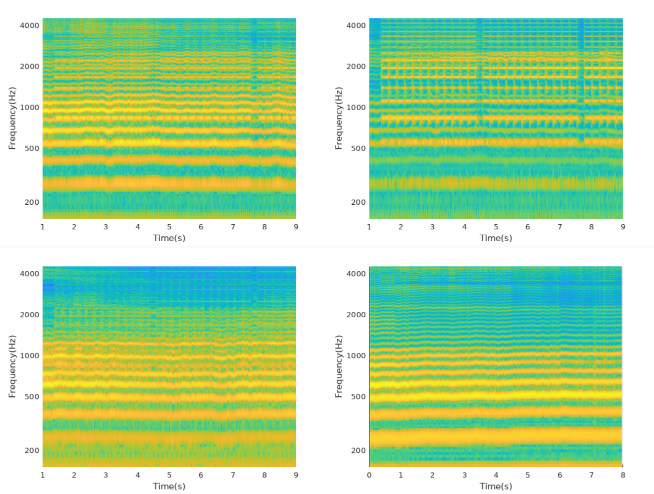


Figure 3: Top row: Spectrogram of noisy [a], spectrogram of the noise. Bottom row: spectrogram of filtered vowel [a] and spectrogram of an ideal recording.

Tools of the trade

Meet the problem

Problem statement: we wish to recover signal $x(t)$ from measurement $y(t)$ when

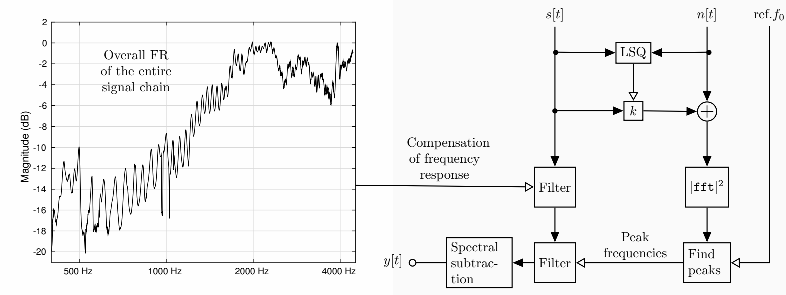
$$y = h \star (x + n).$$

We also have available the noise sample

$$\hat{n} = \hat{h} \star (n + \hat{x}).$$

- The responses h and $\hat{h}$ are not known, and they are impractical to measure due to circumstances in the MRI room.
- There is significant crosstalk between the two recorded signals y and $\hat{n}$.

Noise cancellation algorithm



1. **LSQ**: Speech channel crosstalk is optimally removed from noise signal minimisation.
2. **Frequency response compensation:** The magnitude response of the system is compensated. The peaks in the frequency response are due to the longitudinal resonances of the waveguides.
3. **Noise peak detection:** The noise power spectrum is computed by FFT, and the most prominent spectral peaks of noise are detected.
4. **Harmonic structure completion**: The set of noise peaks is completed by its expected harmonic structure to ensure that most of the noise peaks have been found.
5. **Notch filtering:** The noise peaks are removed by using notch filters.
6. **Spectral subtraction:** A sample of the acoustic background (helium pump etc.) of the room is extracted from the beginning of the recording. The averaged spectrum of this “silent sample” is subtracted from the speech signal using FFT and inverse FFT.

J.Kuortti et al Post-processing speech recordings during MRI, Biomedical Signal Processing and Control 39(C), (2018) 11-22

Other things we have tried

- Time domain subtraction: Try to estimate the responses h and $\hat{h}$. Then deconvolve and subtract:

$$\tilde{n} = n - \frac{\langle n, s \rangle}{||n|| \cdot ||s||} s \quad \text{and} \quad \tilde{s} = s - \frac{\langle s, \tilde{n} \rangle}{||s|| \cdot ||\tilde{n}||} \tilde{n}$$

  In practice, it is extremely difficult to find the kernels h and $\hat{h}$ without a proper reference.

- Noise component identification: The noise spectrum in the case of MRI is very band concentrated. To remove noise we can fit a bandstop filter on every identified noise component.

  Unfortunately, depending on SNR, it is difficult to identify energy concentrations (i.e., peaks in spectrum) that are related to MRI scanner due to channel crosstalk.

D. Aalto et al, Large scale data acquisition of simultaneous MRI and speech, Applied Acoustics 83 (1), (2014) 64–75

Also afloat

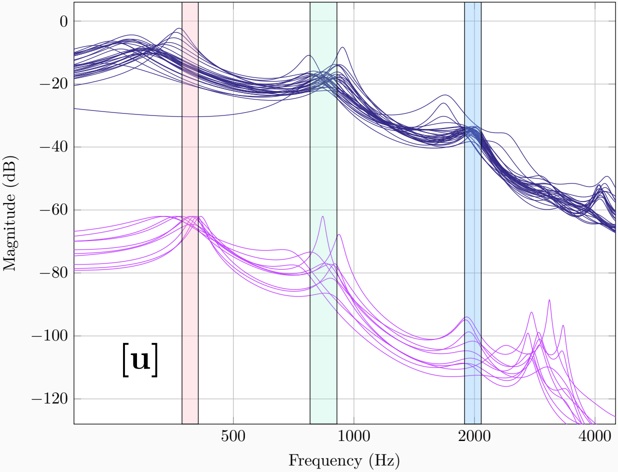
For other approaches see

- E. Bresch, K. Nielsen, K. Nayak, S. Narayanan, Synchronized and noise-robust audio recordings during realtime magnetic resonance imaging scans, JASA 120 (4) (2006) 1791–1794
- J. Přibil, J. Horáček, P. Horák, Two methods of mechanical noise reduction of recorded speech during phonation in an MRI device, Measurement Science Review 11 (3) (2011) 92–99
- J. Inouye, S. Blemker, D. Inouye, Towards undistorted and noise-free speech in an MRI scanner: correlation subtraction followed by spectral noise gating, JASA 135 (3) (2014) 1019–1022

The Validation

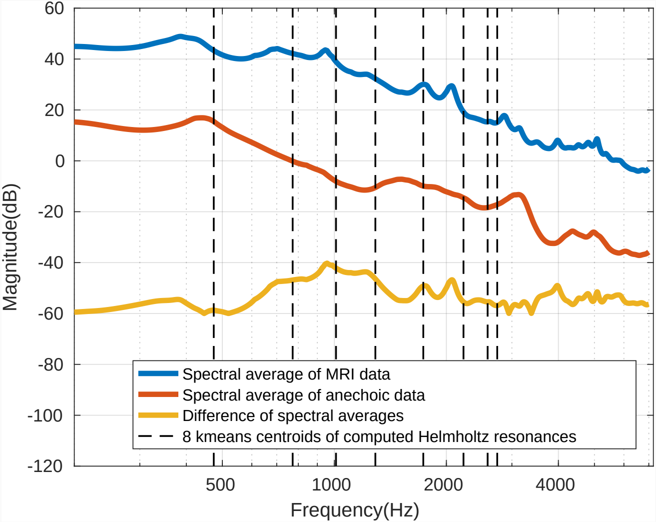
Resonant frequencies

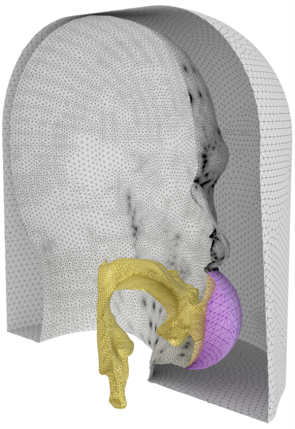
Removing the noise preserves the relevant spectral data, i.e., the resonant frequencies, i.e., the *vowel formants* quite well.



Retaining spectral characteristics

We even manage to reveal the resonance artifacts caused by the MRI head coil by comparisons with numerical Helmholtz modelling.





Thanks for your attention

Questions?