RECORDING SPEECH DURING MAGNETIC RESONANCE IMAGING

T. Lukkari[1], J. Malinen[1], P. Palo[1]
[1]Institute of Mathematics, Helsinki University of Technology, Helsinki, Finland

***Abstract*: We discuss recording arrangements for speech during an MRI scan of the speakers vocal tract. The image and sound data thus obtained will be used for construction and validation of a numerical model for the vocal tract.**
***Keywords*: Speech recording, MRI**

I. Introduction

This article reports progress in development of a FEM-based numerical simulator for Finnish vowels. To obtain the anatomical geometry and to validate the model, we need formants from a speech signal that is recorded simultaneously with an MRI scan.

Magnetic resonance imaging (MRI) has been used for imaging the vocal tract for a long time [1]. Nowadays, the scanning can be carried out in well under 30 s [3]. The anatomical data produced by MRI is suitable for generating the computational mesh for the finite element method (FEM). FEM solvers for the wave equation have been used for simulating normal speech production acoustics [4; 5; 9], the effects of anatomical abnormalities, and oral and maxillofacial surgery on speech [2; 6; 8].

We shall carry out the imaging using a Siemens Magnetom Avanto 1.5 T machine. The environment in an MRI room is challenging from the viewpoint of sound recording. There is a static 1.5 T magnetic field within the MRI coil, and even the ambient field may be considerable. An imaging sequence produces an electromagnetic field at ≈ 64 MHz since the Larmor frequency of protons is 42.58 MHz/T. The peak power may reach several kilowatts. To further complicate things, there is acoustic noise of about 90 dB (SPL) over a range of frequencies that inconveniently overlap the expected formants.

The noise prevents the subject from hearing her/his own voice during the scan. Thus, the denoised, undelayed signal should be fed back into the subject's ear phones to improve speech naturality. As the experiments involve a human subject, safety and comfort must be taken into account.

Roughly speaking, the task is to separate a plane wave (i.e., the speech) from a cylindrically symmetric noise source (i.e., the environment), while paying attention to the complications described above.

II. Specifications and design

Because of the magnetic field, only negligible amounts of ferromagnetic material may be used in the experimental apparatus inside the MRI room. None at all is allowed in the sound collector within the MRI coil. All electronics inside the MRI room have to be shielded against overvoltage and radio frequencies. Of course, closed loops in all conducting material must be strictly avoided.

A. Sound collector and acoustic wave guides

A two-channel sound collector will be used, one channel for the speech and the other for the noise. The dimensions of the collector must be small compared to the formant wavelengths, and the collector must fit inside the MRI equipment.

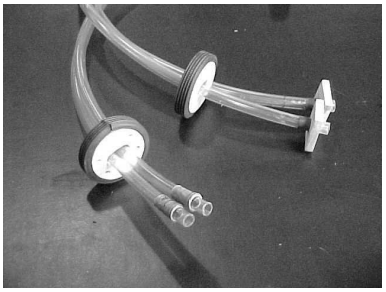

Figure 1: Acoustic wave guides and their suspension arrangement

The sound signals are transmitted to a microphone assembly by acoustic wave guides (see Fig. 1). They are constructed from soft PVC tube of inner diameter 9 mm. The length of each wave guide is 3.0 m, and they are suspended pairwise so as to cancel out external disturbances.

The medium in the collector and the wave guides is air. Sound transmission in the wave guide walls appears to be negligible. The frequency response of the acoustic wave guide between 0.42–3.3 kHz is given

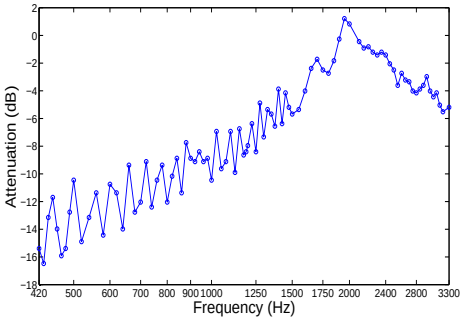

Figure 2: Frequency response of the wave guide

in Fig. 2. At lower frequencies in Fig. 2, longitudinal resonances of the wave guide appear. Below 1.5 kHz, there is ≈ 4dB attenuation per octave that can be easily compensated by, e.g., an RC filter.

B. Shielded microphone assembly and cabling

The microphone assembly is enclosed in a Faraday cage. The cage is made of 6 mm aluminium plate, which is thick enough not to buckle or resonate. Damping material can be used inside the cage, if necessary. The acoustic wave guides are brought into the cage through electromagnetic wave guides, designed to be opaque at frequences between 10–100 MHz.

The microphone assembly (see Fig. 3) consists of four Panasonic WM-62 condenser microphones (with sensitivity -45 ± 4 dB re 1V/Pa at 1 kHz, ∅ 9 mm) and a power source for them. The nominal frequency response of the microphones, as given by the manufacturer's data sheet, is essentially flat in the frequency range of interest. By a superficial measurement, sensitivities and frequency responses of such microphone units do not seem to differ from each other significantly, and hence we omitted more detailed calibration measurements.

The microphones are embedded into a plate that is acoustically and electrically isolated from the walls of the Faraday cage. The sound waves enter the microphones through simple, adjustable acoustic impedance matchings (see the lower right corner in Fig. 3). These matchings are tuned experimentally by closing some of the holes (∅ 2 mm) in the walls of the tubes. Tuning is carried out in order to minimize the reflected wave from the microphone assembly, analogously to the termination of usual electric trasmission lines. This results in a partial supression of the longitudinal resonances of the wave guide (see Fig. 2).

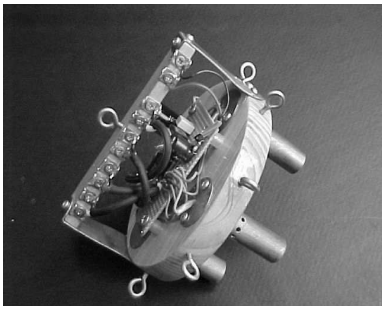

Figure 3: Microphone assembly

An energy dissipation of several dB's is seen in the frequency response of the system, depending on the number of impedance matching holes that have been closed. Since the matching consists of both open and closed partial terminations of the wave guide, the residual reflection takes place both with and without phase inversion. We remark that this corresponds exactly to the number of measured peaks in Fig. 2.

The signals are transmitted from the MRI room by two microphone cables (Tasker C116 4x0.14-26AWG); two channels in each. All cable endings are shielded against overvoltages by diodes. Since only two channels are used by the sound collector, the remaining microphones are a reserve.

C. De-noising amplifier and CMRR curves

The test subject needs to hear the de-noised signal in real time. Hence, we implement the de-noising system as an analog device. It is a summing amplifier (see Fig. 4) with one direct channel (for the signal) and three adjustable, inverted channels (for subtracting up to three noise signals). Before recording, the summing coefficients are adjusted manually by listening to the output. The device is constructed using six LM741's, and its input impedance is 3 kΩ.

The frequency response of the amplifier is flat between 0.2–5 kHz. Its optimal common mode rejection ratio (CMRR) between 0.42–3.3 kHz is given by the lowest, quite smooth curve in Fig. 5. This CMRR can be improved by reducing tolerances of the electrolyte capacitors in the amplifier.

The upmost, rather rough-looking curve in Fig. 5 is the measured CMRR of the whole system. This includes the wave guides and the acoustic impedance matchings at the ends of the wave guides. The difference between the two curves in Fig. 5 is mostly due to the physical properties of the wave guides and – unfortunately – the poor quality of the sound source used in the measurements.

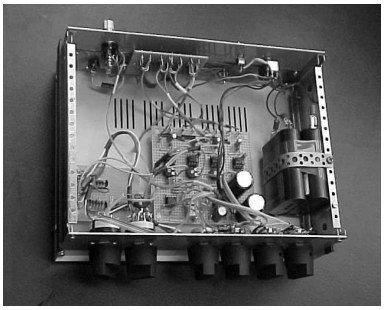

Figure 4: De-noising amplifier

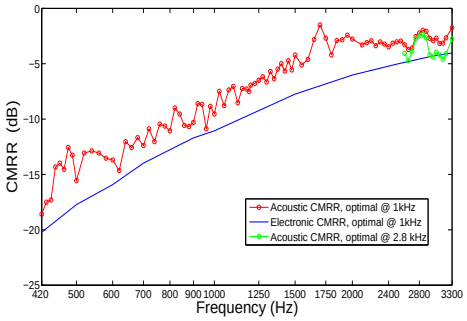

Figure 5: Optimal CMRR curves of the amplifier (lowest) and the acoustic wave guides (upmost)

D. Computer equipment and signal processing

The de-noised signal from the amplifier is digitized using a MacBookPro2,2 computer running MacOSX 10.4.9. The required signal processing and formant extraction will be done using Matlab 7.4, Signal Processing Toolbox, and custom made code. In particular, the longitudinal resonances visible in Fig. 2 can be compensated with Matlab. We remark that the frequency response must be remeasured in the final experimental setting since bending the wave guide will move the resonance frequencies [7].

III. Measurements

We proceed to explain in detail how the data in Figs. 2 and 5 was obtained.

A. Arrangement and equipment

A sine wave generator (Taylor 192A) was coupled to a two-channel sound source (see Fig. 6), and the sound pressure at the source was manually kept at a constant level 94 dB (SPL) for frequencies between 0.42–3.3 kHz. This was accomplished by measuring the reference microphones inside the sound source using an analog volt meter (Heathkit V-7 A) through a microphone preamplifier (Resound CVS908). An oscilloscope was used to detect possible distortion visually.

The produced sound signals were fed to the microphone assembly (see Fig. 3) through the wave guides (see Fig. 1). The wave guides were completely straightened out during the measurements, and the surrounding acoustical noise was controlled by various means.

From the microphone assembly, the two signals were brought to the direct and inverted channels of the de-noising amplifier. The amplification of the direct channel was set to 45dB. The amplification of the inverted channel was set so that the output of the amplifier was at its minimum when a 1 kHz sine signal was fed to both the direct and inverted channels.

All data for Figs. 2 and 5 was measured using a second analog volt meter (Goerz Unigor 226221) at the output of the de-noising amplifier. At all measured frequencies, readings were taken both with and without the inverted channel coupled.

B. Sound Source

Consider the measurements described above. An ideal sound source for such measurements should be able to produce two sine wave sound signals of the same amplitude and without phase difference. Both the channels should be acoustically uncoupled, and their acoustic impedances should be the same. All this should be accomplished without distortion, over a wide range of frequencies and sound pressure amplitudes.

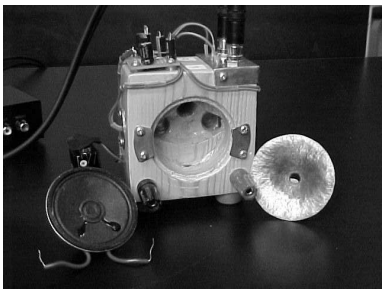

Figure 6: Disassembled sound source

Our design (see Fig. 6) consists of a loudspeaker ($\varnothing$ 50 mm, impedance 8 Ω), together with a symmetric cavity that divides the pressure field to two channels. There is a reference microphone of type Panasonic WM-62 embedded in the walls of each channel.

The sound source also includes a rudimentary acoustic impedance matching of the same type as used in the microphone assembly. Its purpose is to mimic qualitatively the impedance of the real sound collector that will be used inside the MRI equipment. We remark that the acoustic impedances of the sound collector and the sound source are different, which has a quantitative effect on a frequency response curve like Fig. 2.

The sound source suffers from the resonances of both the cavity and the loudspeaker itself. Near such a resonance, the produced sound signals are out of phase, and the results of the CMRR measurement are worse than the true CMRR would be. To reduce a particularly inconvenient resonance at ≈ 1.7 kHz, a horn made of copper plate, on the right in Fig. 6, had to be placed between the loudspeaker and the cavity. We could not obtain the CMRR data for high frequencies, since the cavity becomes resonant at ≈ 3.5 kHz. On the other hand, frequencies under 0.4 kHz must be produced without the horn in place, since the horn distorts the signal at lower frequencies.

The peaks at 1.7 kHz, 2.85 kHz, and 3.3 kHz in the upmost CMRR curve in Fig. 5 are at least partly explained by a phase difference of the sound source channels. These phase differences were verified by an oscilloscope Lissajous measurement. However, the peak at 1.95 kHz is not due to phase difference.

Above 2 kHz, the channels of the sound source begin to drift out of balance because the loudspeaker is not symmetric. When this lack of balance was compensated by readjusting the de-noising amplifier, we obtained a much better CMRR curve for 2.6–3.3 kHz that has been plotted in Fig. 5, too.

We conclude that the true CMRR for the wave guides is significantly better for high frequencies than what Fig. 5 would indicate. The design and construction of a good quality, multi-channel sound source remains a challenging exercise in acoustic engineering.

IV. Conclusions

We have described noise cancellation, sound transmission, and recording techniques through acoustic wave guides in difficult environments such as the MRI room.

The acoustic wave guides change sound quality; speech becomes somewhat crisp or even hoarse. However, speech remains easily understandable without numerical compensation of the wave guide resonances. As a conclusion, we expect to obtain good quality recordings of many types of speech signals from which, e.g., successful formant extraction should be possible.

Acknowledgment: We wish to thank Dr. A. Laakso and Dr. K. Rytsölä (Laboratory of Physics, Helsinki University of Technology) for valuable discussions and providing laboratory facilities for the measurements.

Mr. Lukkari has received support from Academy of Finland.

References

[1] Baer, T., Gore, J. C., Boyce, S., and Nye, P. W. (**1987**). "Application of MRI to the analysis of speech production," Magnetic Resonance Imaging **5**, 1 – 7.

[2] Dedouch, K., Horáček, J., Vampola, T., and Černý, L. (**2002**). "Finite element modelling of a male vocal tract with consideration of cleft palate," in "Forum Acusticum," Sevilla, Spain.

[3] Engwall, O. (**2004**). "Speaker adaptation of a three-dimensional tongue model," in S. H. Kim and D. H. Youn., eds., "ICSLP 2004," vol. I, Jeju Island, Korea, 465 – 468.

[4] Hannukainen, A., Lukkari, T., Malinen, J., and Palo, P. (**2007**). "Vowel formants from the wave equation," Journal of the Acoustical Society of America Express Letters **122**, EL1–EL7.

[5] Lu, C., Nakai, T., and Suzuki, H. (**1993**). "Finite element simulation of sound transmission in vocal tract," J. Acoust. Soc. Jpn. (E) **92**, 2577 – 2585.

[6] Nishimoto, H., Akagi, M., Kitamura, T., and Suzuki, N. (**2004**). "Estimation of transfer function of vocal tract extracted from MRI data by FEM," in "The 18th International Congress on Acoustics," vol. II, Kyoto, Japan, 1473 –1476.

[7] Sondhi, M. M. (**1986**). "Resonances of a bent vocal tract," J. Acoust. Soc. Am. **79**, 1113–1116.

[8] Švancara, P. and Horáček, J. (**2006**). "Numerical modelling of effect of tonsillectomy on production of Czech vowels," Acta Acustica united with Acustica **92**, 681 – 688.

[9] Švancara, P., Horáček, J., and Pešek, L. (**2004**). "Numerical modelling of production of Czech wovel /a/ based on FE model of the vocal tract," in "Proceedings of International Conference on Voice Physiology and Biomechanics," .