RECORDING SPEECH DURING MRI: PART II

J. Malinen, P. Palo
Dept. of Mathematics and Systems Analysis, Helsinki Univ. Tech., Espoo, Finland

***Abstract*: We design and construct a recording arrangement for speech during an MRI scan of the speakers vocal tract. We concentrate on the acoustic environment around the test subject inside the MRI machine. The data thus obtained is used for construction and validation of a numerical model of the vocal tract.**
***Keywords*: Speech recording, MRI, acoustic wave guides**

I. Introduction

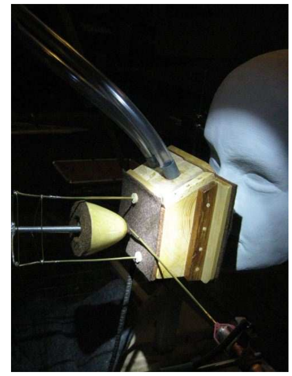
We model vowel production by the wave equation, see [5]. For the vocal tract, we use a boundary controlled wave equation and the corresponding resonance model (i.e., the Helmholtz equation) which we solve with FEM. A similar approach has been taken in, e.g., [9]. As a source for the wave equation, we use a flow mechanical glottis model introduced in [1], [2]. To validate and tune the vocal tract model, we need to extract formants and their bandwidths from speech and singing signals. These signals must be recorded in high quality during an MRI scan.

The present article is the latter part of a description of the sound recording arrangement based on an acoustic sound collector and wave guides; see [6] for the first part. We now concentrate on the acoustics around the test subject inside the MRI machine in the measurement configuration shown in Fig. 1b.

The acoustic signals are captured by the sound collector, see Fig. 1a, whose position with respect to the test subject is as in Fig. 1b. The collector is a two-channel device that operates by the same principle as a differential microphone. It has separate horns for speech and noise — one on each side — that lead to the wave guides. As explained in [6], both sound channels are acousticly transmitted by wave guides to a microphone assembly inside a Faraday cage. Because the noise cancellation is realized by analog electronics, the de-noised signal can be fed back into subject's earphones without delay.

The recording equipment has been designed for recording speech and singing signals for research purposes. Further engineering effort is required to make the device suitable for clinical practice.

(a)



(b)

Figure 1: (a) The sound collector and a reflector paraboloid suspended from a temporary measurement suspension (shown with a cm scale ruler), (b) Measurement arrangement for near field acoustics of the sound collector

II. Challenges and solutions

A. Engineering challenges

The MRI room is a quite challenging sound recording environment. There is acoustic noise of about 90 dB(SPL) during the imaging sequence. The noise arrives to the sound collector with different delays because of multi-way propagation. A Siemens Magnetom Avanto 1.5 T MRI machine produces a static 1.5 T magnetic field, and an imaging sequence produces an electromagnetic field at 81 MHz with a peak power of several kW. Because of safety and image quality considerations, no metal or electronics can be taken near the test subject. For speech naturality, comfortability of the test subject is important [4].

B. Technical solutions

The sound collector is completely passive, metal free, and without moving parts. The wave guides detach from the sound collector so as not to hinder taking the subject out from the machine. The collector is fully compatible with MR safety requirements and does not cause any artefacts in the images. It fits on the head coil of a Siemens Magnetom Avanto 1.5 T machine. The test subject lies in supine position inside the MRI machine. The sound collector is about 30 mm away from the lips.

The ambient noise is partly removed by the two-channel recording arrangement (as shown in [Fig. 5, 6]), and partly by attenuation material carefully positioned inside the MRI machine. For acoustic impedance adjustment, both horn surfaces of the collector are covered with attenuation material that also takes care of exhalation noise.

On top of the sound collector in Fig. 1a, there is a reflector paraboloid that widens the incoming noise beam by shadowing it in the middle. Without such a reflector, the noise sample gets collected in an undesirably narrow angle. The test subject affects the acoustic impedance (hence, the frequency response) of the speech channel, and the form and distance of the paraboloid are tuned to approximate the same effect on the noise channel side. The noise cancellation is succesful when the acoustic impedances are close to each other.

See [6] and [8] for details of the components not described here.

C. Measurement approximations.

To simplify analysis, we divide the acoustic space into two parts. We regard a sphere of radius 8 cm around the center point of the sound collector as the *near field*. The characteristic curves (such as Fig. 2) of the whole recording equipment are particularly sensitive to near field phenomena. In the *far field*, noise and its reflections from the MRI machine require most attention.

To further simplify analysis, we divide the frequency range according to the near field length scale: *low frequencies* under 2 kHz ($\lambda/2 > 8$ cm), *middle frequency range* $2 - 4.4$ kHz (4 cm $< \lambda/2 <$ 8 cm), and *high frequencies* above 4.4 kHz ($\lambda/2 < 4$ cm). At low frequencies, the two channel noise cancellation is most effective when the inconvenient longitudinal resonances of the wave guides are properly controlled by wave guides' impedance terminations. At high frequencies, propagation of noise can be understood by the ray optical approximation but severe complications in active noise cancellation are caused by phase differences due to multi-way propagation. The middle frequency range has none of the good and all of the bad qualities. In such case, the only reasonable solution is the placement of damping material around the test subject, based on trial and error.

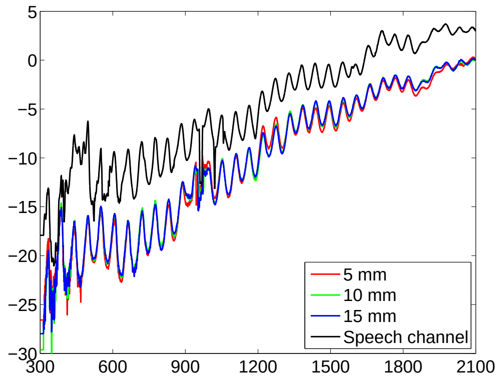


Figure 2: Frequency responses of the speech channel (above) and the noise channel with various reflector positions in mm with attenuation in dB and frequency in Hz. Note that the amplifications of the speech and noise channels are different to improve readability.

III. Characteristic curves and tuning

In the measurements we have a two fold objective: First, to understand the near field behaviour in terms of acoustic impedances and frequency responses. Second, to understand the constraints posed by near field engineering on the technical solutions of the far field problems.

To tune the equipment and to obtain necessary frequency responses for sound post-processing by DSP, we carried out several near field measurements in an anechoic chamber from physical models, see Fig. 1b. We return to acoustic measurements with physical models when weeding out artefacts from acoustic data measured from a test subject.

A. Frequency responses

The frequency responses of the speech and noise channels are given in Fig. 2. The noise channel response is measured with several paraboloid positions. The frequency responses of both channels are very similar as expected from the symmetrical construction of the channels. This is a prerequisite for the noise cancellation to work by analogue signal subtraction.

The data in Fig. 2 has been measured using the experimental setting of Fig. 1b. The tip of the reference microphone probe, see Fig. 4b, is placed at the distance of 5 mm above the surface level of the sound collector at the center of the corresponding horn. The frequency responses of the channels have been determined with respect to the sound pressure at these reference points. Note that the speech channel is measured using the point source in Fig. 4a, and the noise channel using a plane wave source that resembles the

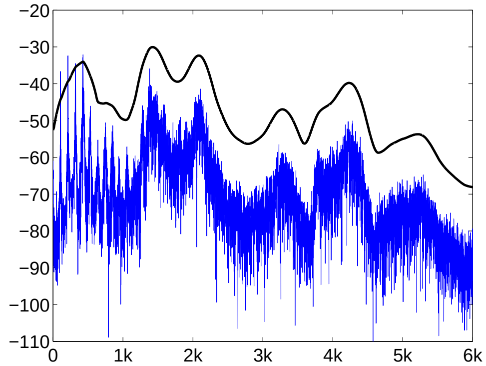
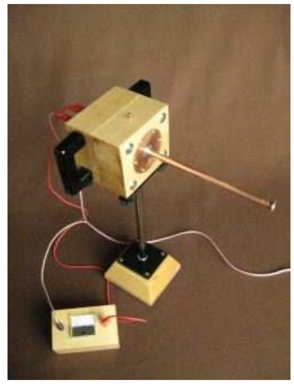


Figure 3: Original and averaged (raised by 20 dB) spectrum of a long production of [øː] recorded within an anechoic chamber



(a)

(b)

Figure 4: (a) Acoustic point source, (b) Reference microphone probe (right), microphone units of type Panasonic WM-62 (middle)

interior surface of the MRI machine.

B. *Tuning the reflector paraboloid*

As can be seen from Fig. 2, the paraboloid position does not significantly affect the low frequency response of the noise channel. Thus, the paraboloid position and size can be optimised according to the requirements of the middle frequency range $2 - 4.4$ kHz without compromising the noise cancellation properties in the low frequencies. The optimisation objectives in the middle frequency range are both the incoming noise beam shape and noise cancellation (to the extent it is feasible).

IV. Speech recording experiment

We recorded a full set of Finnish vowels produced by both authors. We did not use any model of the MRI device or its noise. The formant peaks at 0.42 kHz, 1.41 kHz, 2.09 kHz, 3.28 kHz, 4.23 kHz, and 5.22 kHz are very clearly visible in spectrograms even without any DSP compensation, see Fig. 3. The glottis pulse is easily recoverable, too.

A typical MRI device produces a sparse and spiky noise spectrum — i.e., the sound energy is restricted to few fairly narrow frequency bands and their superharmonics. Given the linearity of the recording equipment (the microphones are the dominating source of nonlinearity), it is possible to separate speech from the residual noise in frequency domain. This is carried out by recording *a priori* noise spectrum data when the silent test subject lies in the MRI machine during an imaging sequence.

V. Construction of laboratory equipment

In addition to the equipment described below, we use a loudspeaker assembly for simulating the ambient (noise) field. Custom Matlab 7.4 code was written to generate weighted sweeps, and to estimate and compensate frequency responses.

A. *Sound source and face model*

We constructed an acoustic point source and a natural size face model shown in Fig. 4a and Fig. 1b, respectively. The horn of the point source can be placed at the mouth of the face model.

A practically constant, sufficiently high amplitude sound pressure can be obtained above 300 Hz when the point source is fed by a properly weighted sinusoid sweep signal. Then the virtual source point is at the center of the exponential horn opening on the right in Fig. 4a. Because of the dimensions of the source, acceptable signal cannot be produced under 300 Hz. Fortunately, the recording equipment requires little attention at these low frequencies.

For validation, we measured the polar patterns of the source at the distance of 35 mm from the virtual source point with and without the face model. The patterns are presented in Fig. 5 at frequencies 0.5, 1, 2, and 4 kHz. Even with the face model, the amplitude variation stays within an acceptable 3 dB range.

B. *Reference probe*

In the near field measurements, even the small ∅ 9 mm reference microphone requires a special probe to avoid considerable distortion in the results. The probe and some microphone units are shown

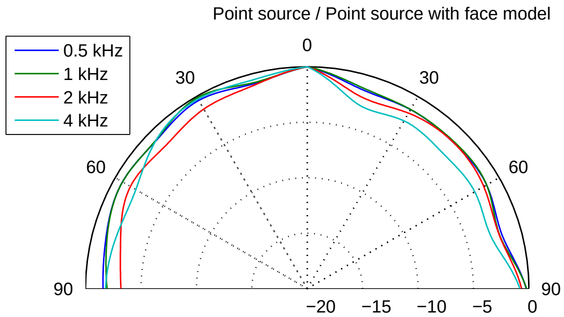


Figure 5: Polar pattern of the point source at 35 mm with (right) and without (left) the face model for angles between 0° and 90° and attenuation in dB

in Fig. 4b. The probe can be seen extending from the lower right corner in Fig. 1b. It is 150 mm long and of ∅ 2 mm. Its frequency response was measured using the sound source shown in [Fig. 6, 6].

VI. Conclusions

We have described equipment and characteristic curve measurements for a recording arrangement of speech during an MRI scan. Up to 4.4 kHz, the presented equipment performs essentially like a pair of microphones in dipole configuration. For a detailed discussion, see [8].

The space around the subject in the MRI machine contains reflecting surfaces, and the sound collector receives strong echoes of high frequency noise in different phases. For this reason, the active dipole based noise cancellation cannot be expected to work well for frequencies over 2 kHz. Instead, passive attenuation material must be placed inside the MRI machine. The position of the damping material is most easily determined empirically in the imaging situation, rather than by using mathematical or physical models.

There are comparable systems that are based on fiber optics (e.g., [3], [7]). Some of the challenges in acoustic and optical solutions are the same, such as the multi-way propagation of noise. Acoustic equipment is larger than optical, and additional complications from various acoustic impedances require more attention. With acoustic equipment, however, linearity is always guaranteed if the microphones are used within their operational limits; the non-microphonic sound collector is immune to vibrations; and the recording arrangement can be easily modified to meet a great variety of practical situations.

Acknowledgment: We wish to thank Prof. P. Alku (Dept. of Signal Processing and Acoustics, TKK) for valuable discussions and for providing laboratory facilities, and Chief Eng. V.-A. Hakala (Testing Hall of Structural Engineering, TKK) for providing facilities for construction of the equipment.

Mr. P. Palo has received support from the Instrumentarium Science Foundation and the Finnish Cultural Foundation.

References

[1] Aalto, A. (**2009**). A low-order glottis model with nonturbulent flow and mechanically coupled acoustic load, Master's thesis, TKK, Helsinki. available at `http://math.tkk.fi/research/sysnum/`.

[2] Aalto, A., Alku, P., and Malinen, J. (**2009**). "A LF-pulse from a simple glottal flow model," in "MAVEBA 2009," Florence, Italy.

[3] Branderud, P. *et al.* (**2009**). Personal communication.

[4] Engwall, O. (**2006**). Speech production: Models, Phonetic Processes and Techniques, chap. Assessing MRI measurements: Effects of sustenation, gravitation and coarticulation., Psychology Press, New York, 301 – 314.

[5] 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.

[6] Lukkari, T., Malinen, J., and Palo, P. (**2007**). "Recording speech during magnetic resonance imaging," in "MAVEBA 2007," Florence, Italy, 163 – 166.

[7] Optoacoustics, Ltd., accessed Aug. 25th, (**2009**). http://www.optoacoustics.com/

[8] Palo, P. (**2009**). A wave equation model for vowels: Validation and parameter estimation, Manuscript for Licentiate thesis, xii + 50 pp., TKK, Institute of Mathematics.

[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," .