We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Analysis of Acoustic Noise and its Suppression in Speech Recorded During Scanning in the Open-Air MRI

Jiří Přibil, Anna Přibilová and Ivan Frollo

Abstract

The paper focuses on describing three methods of noise reduction in the speech signal recorded in an open-air magnetic resonance imager (MRI) working in a weak magnetic field during human phonation for the vocal tract modelling. This paper also analyses and compares spectral properties of the acoustic noise produced by mechanical vibration of the MRI device gradient coils. Then, the experiment with mapping of noise sound pressure level (SPL) in the MRI neighbourhood is described. The changes in acoustic noise spectral properties caused by loading of the holder of the lower gradient coils by the weight of the examined person lying in the scanning area of the MRI device is evaluated too. The influence of setting of the basic scan parameters of the used MR sequence (TR and TE times) on the spectral properties of the generated acoustic noise is also analysed. The results achieved are used to create a database of initial MR scan parameters such as the filter bank for noise signal pre-processing and to design a correction filter for noise suppression in the speech signal recorded simultaneously with three-dimensional (3D) human vocal tract scanning.

Keywords: acoustic noise, spectral analysis, noise suppression in speech, statistical analysis, magnetic resonance imager

1. Introduction

The non-invasive magnetic resonance imaging (MRI) technique enables scanning of the human vocal tract showing the configuration of the supraglottal resonant cavities (pharynx, oral and nasal cavities) during vowel phonation. The primary volume models of these acoustic spaces created from the MR images can then be transformed into the three-dimensional (3D) finite element (FE) models [1]. The FE models facilitate new possibilities in simulating and under-

standing of speech production. Synchronicity between image and audio acquisition must be ensured if the speech signal is recorded in parallel with 3D vocal tract scanning [2]. Further speech analysis and vocal tract modelling is conceivable only under the condition of an adequate signal-to-noise ratio as the noise is an inherent part of imaging by magnetic resonance [3]. The MRI device usually consists of three gradient coils to produce three orthogonal linear fields for spatial encoding of a scanned object. The noise is produced by these gradient coils due to rapidly changing Lorentz forces during fast switching inside the large static magnetic field [4]. It results in a significant mechanical vibration that subsequently propagates in the air as a progressive sound wave perceived by the human auditory system as a noise. Due to harmonically related audio frequencies of the produced acoustic noise, it can be analysed using similar methods as those used in speech signal processing.

There exist several approaches to reduce the acoustic noise produced during MRI scanning [5–7]. In our case, the tested person is lying and articulating in the scanning area of the MRI equipment while the scanning MR sequence is running to obtain an MR image of the human vocal tract. The recorded speech signal with the superimposed noise is then subject to offline analysis and processing. We have successfully used the noise reduction method based on the idea that the acoustic noise produced by the gradient coils of the MRI machine is a periodic signal and its fundamental frequency is represented by a typical peak in the real cepstrum in a similar way as it is for the voiced speech [8]. Our first proposed noise reduction method is based on the limitation of the real cepstrum of the noisy speech and clipping the "wrong peaks" corresponding to the harmonic frequencies of the acoustic noise. This method works well when the basic pitch period of the human voice differs from the repeating period of the running MR scan sequence. However, if both fundamental periods are the same, then alternative techniques must be selected. Our next tested approach to noise suppression uses subtraction between the short-time spectra of the audio signals recorded by two microphones [9]: the first one recorded the speech together with the acoustic noise and the second one recorded only the acoustic noise. Another method of speech enhancement in the MRI environment is based on spectral subtraction of the estimated periodic noise [10]. Therefore, we must know the statistical parameters of spectral properties of the noise produced by the MRI device. This third proposed method uses only one microphone picking up the noise of the running scan sequence that is followed by recording of the speech signal of phonation with the same noise background. As in the case of two-channel spectral subtraction, the natural logarithm and IFFT are applied on the resulting spectrum after subtraction, whereby the limited real cepstrum is obtained [11]. The final clean signal is reconstructed by the cepstral speech synthesizer [12].

This chapter is focused on comparison of the mentioned three methods of acoustic noise suppression in a speech signal recorded simultaneously with MRI scanning for 3D modelling of the human vocal tract. The main motivation of the work was to perform optimization of a correction filter to increase suppression of the acoustic noise while preserving spectral properties of the reconstructed speech and thus ensuring its good quality. The successfulness of different noise reduction methods can be evaluated subjectively—by visual comparison of spectrograms, spectral envelopes, etc. or with the help of objective approaches—numerical matching of determined differences between spectral envelopes, comparison of spectral

distance values calculated between the obtained periodograms, etc. The special aspect of voice recording in the weak magnetic field environment is also discussed here. Next, the detailed analysis of basic and supplementary spectral features of the acoustic noise produced by the MRI device was carried out. The auxiliary measurement experiments also mentioned in this chapter include mapping of the acoustic noise pressure level in the MRI neighbourhood, analysis of the influence of different setting of scan sequence parameters as well as the influence of different body mass of the examined person situated in the scanning area on spectral properties of the generated noise signal.

2. Equipment and methods

The conventional whole-body MRI scanner with its specific acoustic noise makes speech recording in such an environment quite a challenging task that needs special solutions. One possible way of implementation is a construction of a special sound collector [13]; however, this realization requires a completely passive metal-free element without moving parts so that not to cause any artefacts in the scanning image due to induced magnetic field inhomogeneity. In practice, another solution was applied—speech signal collection by optical arrangements using special fibre optical microphones located in the scanning space [14]. In all cases, it allows for real-time processing of the speech signal simultaneously with obtaining of MR images. However, the practical implementation is significantly complicated (synchronization of both processes, special hardware for MRI device, etc.) and more expensive. In our case, when the tested open-air MRI device works with a weak magnetic field with the static magnetic induction up to 0.2 Tesla [15], an easier way to fulfil the condition about elimination of interaction with metal elements may be found. From this point of view, the most feasible solution for speech and noise recording is to use a condenser microphone located in the sufficient distance from the static magnetic field—see the overall view of the MRI equipment in **Figure 1a**.

In addition, due to the low basic magnetic field B_0 in the used MRI equipment, the scanning times are longer than those in the MRI devices working with 1.5 or 3 Tesla superconductive magnets. Prior to the start of each scan sequence, the quality factor (QF) parameter is auto-matically pre-calculated to obtain sufficient quality of a picture—see the obtained MR image of the human vocal tract in **Figure 1c**. The current setting of the scan parameters together with the chosen type of the sequence has also influence on the final scanning time. Typical scanning time is about 1 min, although for the 3D or Hi-Res scan sequences more than 2 min are necessary. However, phonation is performed in blocks with mean duration of 8, 15 or 30 s depending on physical condition of a tested person. The assumed physical position of the tested subject during the MRI session is typically the supine position, that is lying on the back with the neck placed inside the RF coil and between two gradient coils—see an arrangement photograph in **Figure 1b**. The previously performed experiments and measurements [8, 15] have shown that setting of the basic parameters of the scanning sequences—repetition time (RE) and echo time (TE)—has significant influence on the spectral properties of the acoustic noise produced by the MRI device. Values of these parameters result primarily from the chosen

type of the scanning sequence as well as from other settings (field of view, slice thickness, number of sagittal images, etc.). The value of TR time affects mainly the basic noise frequency $F0_N$ in a similar manner as the vocal chords vibration affects the fundamental frequency $F0$ of the speech signal. On the other hand, setting of TE time, number of averages, etc. have effect on noise signal higher frequencies manifested as the changes of the first two formant positions (formant frequencies in a speech signal). From another experimental measurement and analysis follows that the mass of the examined person lying on the lower plastic holder of the gradient coil and the permanent magnet in the scanning area of the MRI device also changes the basic spectral properties of the generated acoustic noise [16].
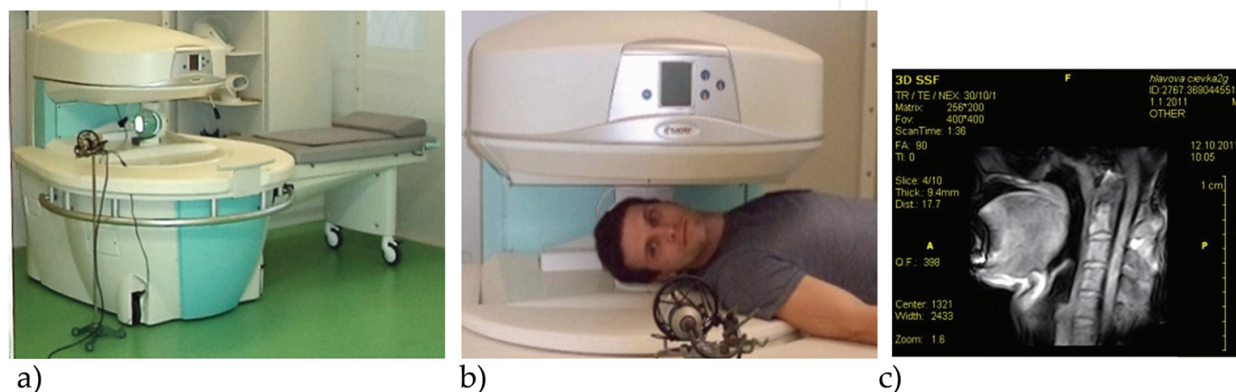

Figure 1. Overall view of the open-air MR imager with the tested water phantom inserted in the RF coil inside the scanning area including an installed pick-up microphone (a), one microphone arrangement of the speech and noise recording experiment with the lying person (b) and obtained MR image of the human voice tract in a sagittal plane after 90° rotation obtained using the MR scan sequence 3D SSF (c).

2.1. Reduction in the gradient coil noise by cepstrum limitation and clipping

The first analysed approach of noise reduction in the recorded speech is based on clipping of the cepstrum peaks corresponding to the frequencies of the acoustic noise produced by the MRI gradient system. This method works in six steps—see the block diagram in **Figure 2**:

1. Calculation of the real cepstrum (including the energy) of the speech with the superimposed noise signal.

2. Detection of the pitch-period L_0 and the voiceness type of the speech signal.

3. Determination of the necessary minimum number of N_0 cepstral coefficients for sufficient log spectrum approximation.

4. Determination of the cepstral peaks corresponding to the frequencies of the acoustic noise in correspondence with estimated L_N value.

5. Limitation of the real cepstrum and cutting off the peaks.

6. Reconstruction of the cleaned speech signal by the pitch-synchronous cepstral speech synthesizer.
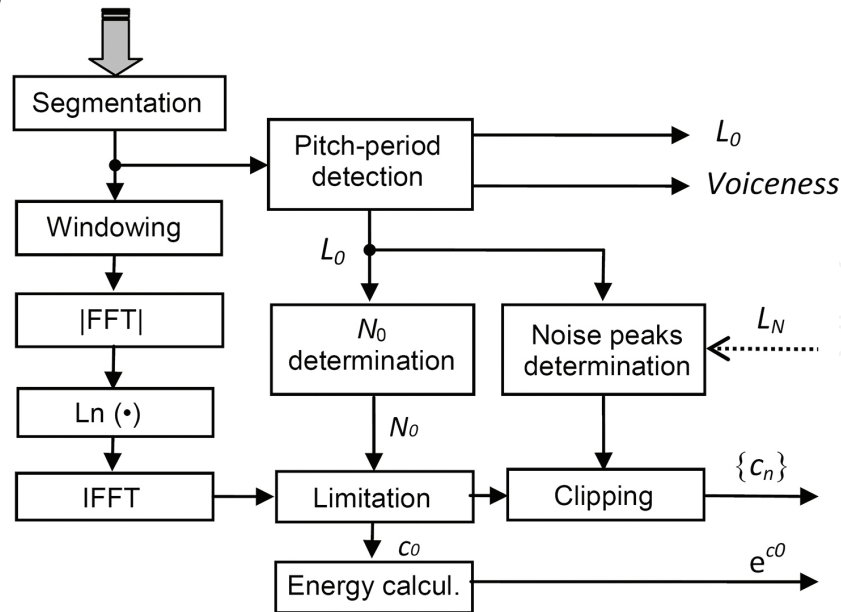
Speech with noise

Figure 2. Block diagram of the noise reduction method based on clipping of the real cepstrum peaks.

The period L_N of the superimposed noise part of the analysed signal must be estimated externally, usually from the known repetition period (given by the currently chosen TR parameter) of the used MR scanning sequence.
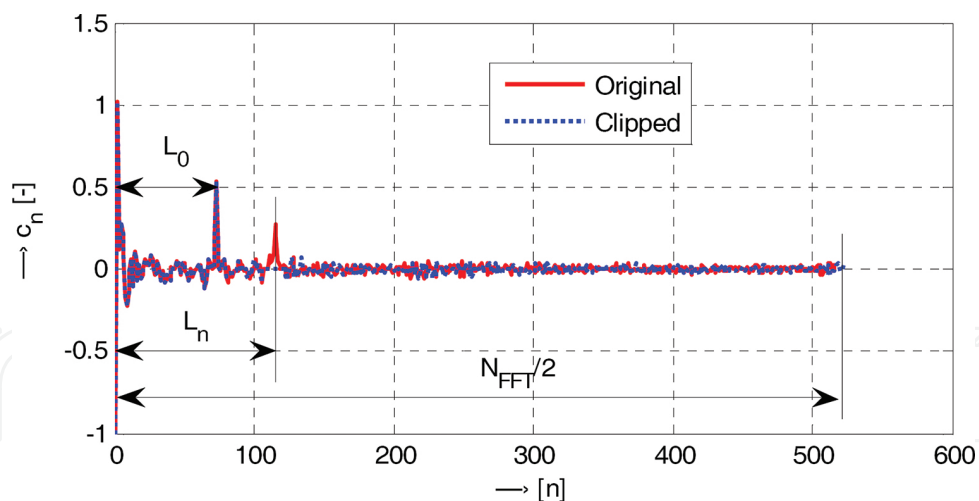

Figure 3. Graphic examples of noise reduction in the recorded speech by clipping and limitation of the real cepstrum; Lo = pitch-period of the speech signal (long vowel "a" articulated by a female person), Ln = period of the superimposed noise and N_{FFT} = 1024.

For cepstral analysis of the speech with the superimposed noise signal, the FFT of the Hamming-weighted signal frames is used for power spectrum computation and the natural logarithm followed by the inverse FFT is applied to get the symmetric real cepstrum—see an example in **Figure 3**. Limitation of the real cepstrum to its first $N_0 + 1$ coefficients can be

described by the Z-transform as $C(z) = c_0 + c_1 z^{-1} + c_2 z^{-2} + \ldots + c_{N0} z^{-N_0}$. The truncated cepstrum represents an approximation of the log spectrum envelope

$$E(f) = c_0 + 2\sum_{n=1}^{N_0} c_n \cos(n \cdot 2\pi f) \tag{1}$$

where the first cepstral coefficient c_0 corresponds to the signal energy.

The reconstruction of the speech signal is realized by a digital filter implementing approximate inverse cepstral transformation. The system transfer function is defined as

$$G_F(z) = e^{c_0} \cdot \prod_{i=1}^{N_0} G_i(z) \tag{2}$$

and it can be implemented as a cascade connection of N_0 elementary filter structures. The system transfer function of the elementary filter is given by an exponential relation

$$G(z) = e^{\hat{S}(z)} \tag{3}$$

where the exponent is the Z-transform of the truncated speech cepstrum

$$\hat{S}(z) = \sum_{n=0}^{N_0} \hat{s}_n z^{-n} \tag{4}$$

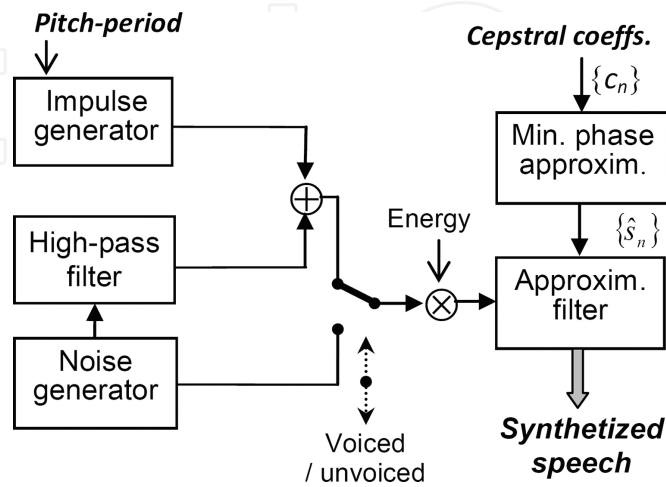and $\{\hat{s}_n\}$ represents the minimum phase approximation of the real cepstrum.

Figure 4. Block diagram of the cepstral speech synthesizer.

The transfer function of the vocal tract model is approximated by Padé approximation of the continued fraction expansion of the exponential function. For the voiced speech, the filter is excited by a combination of an impulse train and a high-pass filtered random noise, and for unvoiced speech, the excitation is formed by a random noise generator—see **Figure 4**. The error of the inverse cepstral approximation depends on the number and the values of the cepstral coefficients and the used approximation structure [12]. Still, the speech signal reconstruction must be carried out pitch synchronously to obtain correct proportionality between duration of voiced and unvoiced parts in the original and the resynthesized signal and to minimize the transient effect.

2.2. Two-channel noise reduction based on spectral subtraction

This method is based on the main idea that the noisy speech signal $x(n)$ is interpreted as an addition of a clean speech signal $q(n)$ and an additive noise $n(n)$.

$$x(n) = q(n) + n(n) \tag{5}$$

This noisy signal is segmented into the frames weighted by windows to obtain its short-time spectrum using the absolute value of the fast Fourier transform $|X(k)|$ of the signal $x(n)$

$$X(k) = \sum_{n=1}^{N_{FFT}} x(n) e^{-j\frac{2\pi}{N_{FFT}}nk} \tag{6}$$

where N_{FFT} represents the number of the points processed by FFT for calculation of the short-time spectrum $X(f, n)$. The enhanced speech spectrum can be obtained by subtracting the noise magnitude spectrum (recorded by the microphone Mic. 2) from the noisy speech magnitude spectrum (recorded by the first microphone Mic. 1)

$$S(f,n) = \left(|X(f,n)| - |N(f,n)| \right) \cdot e^{\phi_n(f,n)} \tag{7}$$

where $\varphi_n(f, n)$ represents the phase of the noisy speech spectrum and n is the index of the processed frame. After the subtraction, the natural logarithm and IFFT are applied on the resulting spectrum, whereby the limited real cepstrum is obtained. Finally, the clean signal is reconstructed by the cepstral speech synthesizer—see the block diagram in **Figure 5**.

2.3. One-channel noise reduction by spectral subtraction

The alternative proposed method uses only one microphone picking up the noise of the running scan sequence without phonation that is followed by recording of the speech signal during phonation with the noise produced by gradient coils of the MRI device with the same setting of the used MR scanning sequence.
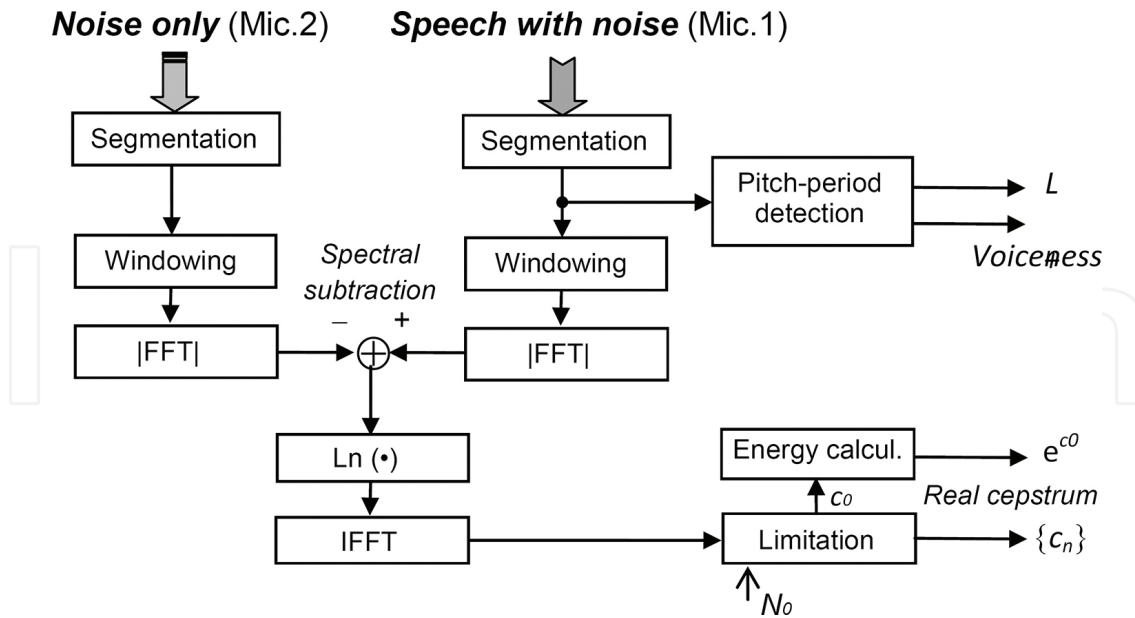
Figure 5. Block diagram of two-channel noise reduction by spectral subtraction and cepstral speech parameterization.
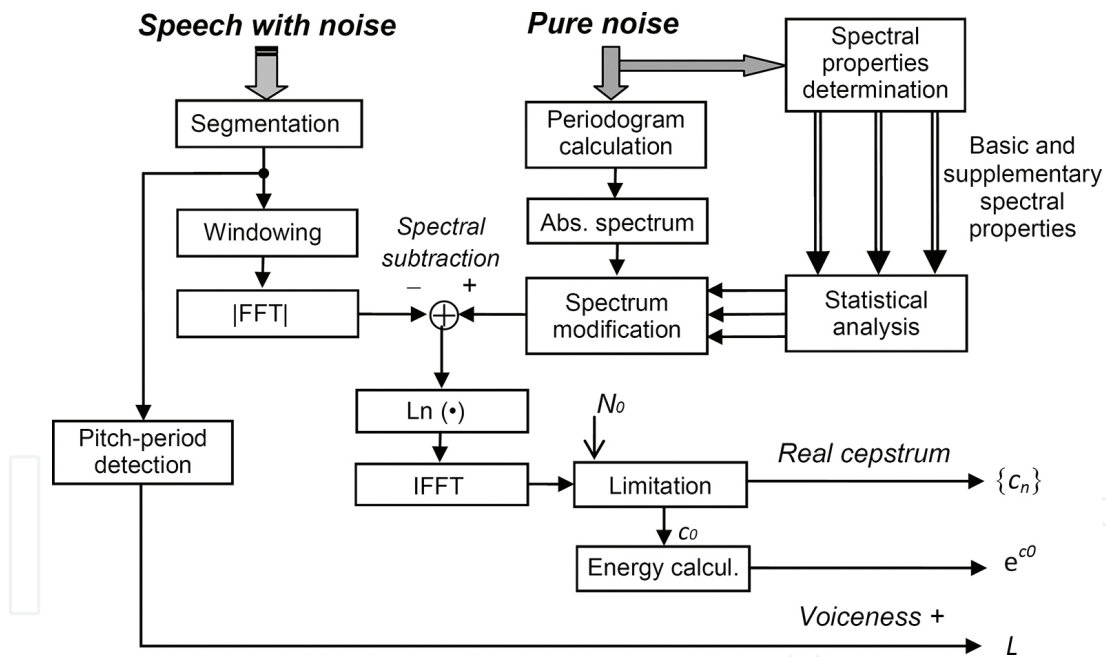
Figure 6. Block diagram of one-channel spectral subtraction method based on statistical analysis of noise spectral properties.

The algorithm contains practically the same processing of the speech signal with noise and the final cepstral analysis after spectral subtraction as it was in the previous two-channel method. However, totally different is processing of the pure noise signal part (typically shorter than the recoded speech signal of the long vowel articulation) is used to calculate the basic mean periodogram using the Welch method. After determination of the absolute spectrum $|S(k)|$,

the basic spectral envelope is obtained. The pure noise signal is next segmentally analysed to determine the basic and supplementary spectral properties. The obtained spectral features are subsequently statistically processed, and the achieved values are used to modify the basic spectral envelope of the noise signal that is further subjected to spectral subtraction as described by the block diagram of the proposed method in **Figure 6**.
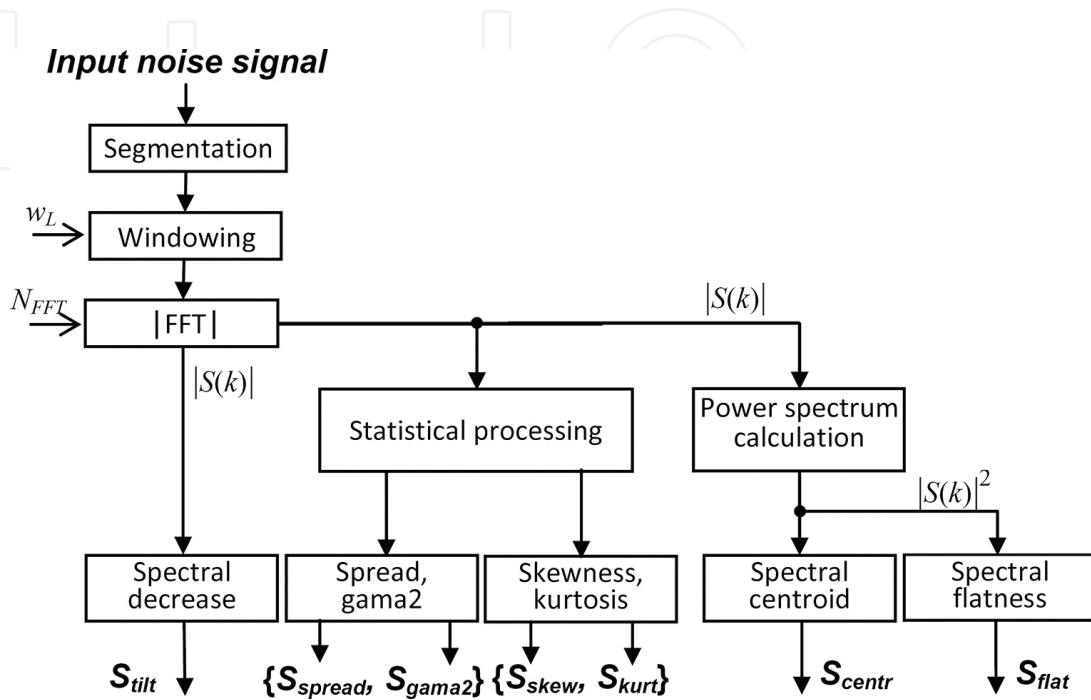

Figure 7. Block diagram of calculation and statistical processing of basic and complementary spectral properties of the noise signal.

The cepstral analysis (see the left part of the block diagram in **Figure 2**) can also be used for determination of basic and additional noise parameters. The basic as well as the supplementary spectral properties is usually determined from the frames (after segmentation and widowing —see the computation scheme in **Figure 7**). These noise spectral properties can be described with the help of the statistical parameters:

–   The spectral spread (S_{spread}) parameter represents the dispersion of the power spectrum around its mean value

$$S_{\mathrm{spread}} = E\left(x - \mu\right)^2 = \sigma^2 \tag{8}$$

where μ is the mean value or the first central moment and σ is the standard deviation (std) or the second moment of the spectrum values, and $E(t)$ represents the expected value of the quantity t in the processed data vector.

–   The spectral skewness (S_{skew}) is a measure of the asymmetry of the data around the sample mean, and it can be determined as the third moment

$$S_{\text{skew}} = \frac{\mu^3}{\sigma^3} \tag{9}$$

– The spectral kurtosis (S_{kurt}) expressed by the fourth central moment represents a measure of peakedness of the shape of the spectrum relative to the normal distribution for which it is 3 (or 0 after subtraction of 3).

$$S_{\text{kurt}} = \frac{\mu^4}{\sigma^4} - 3 \tag{10}$$

– The measure S_{gama2} can be also applied to describe the statistical properties of the spectrum. It is defined using the spectral kurtosis and the spectral spread as follows

$$S_{\text{gama2}} = S_{\text{kurt}} \big/ \big(S_{\text{spread}} \big)^2 \tag{11}$$

– The basic spectral properties include also the spectral decrease (tilt$-S_{\text{tilt}}$) representing the degree of fall of the power spectrum. It can be calculated by a linear regression using the mean square method.

The supplementary spectral properties describe the shape of the power spectrum $|S(k)|^2$ calculated from the noise signal:

– The spectral centroid (S_{centr}) is defined as a centre of gravity of the spectrum, that is the average frequency weighted by the values of the normalized energy of each frequency component in the spectrum. The S_{centr} in [Hz] can be calculated as

$$S_{\text{centr}} = \frac{\sum\limits_{k=1}^{N_{FFT}/2} k \left| S(k) \right|^2}{\sum\limits_{k=1}^{N_{FFT}/2} \left| S(k) \right|^2} \cdot \frac{f_s}{N_{FFT}} \tag{12}$$

where f_s is the sampling frequency.

– The spectral flatness (S_{flat}) determined as a ratio of the geometric and the arithmetic mean values of the power spectrum

$$S_{\text{flat}} = \frac{\left[\prod\limits_{k=1}^{N_{FFT}} \left| S(k) \right|^2 \right]^{\frac{2}{N_{FFT}}}}{\frac{2}{N_{FFT}} \sum\limits_{k=1}^{N_{FFT}} \left| S(k) \right|^2} \tag{13}$$

can also be used for determination of a degree of periodicity in the signal. The S_{flat} values lie in the range of <0 ~ 1>: the zero value represents the signal that is totally periodical (e.g. a pure sinusoidal signal); in the case of $S_{flat} = 1$, the totally noisy signal is classified (e.g. the white noise signal).

2.4. Comparison of noise reduction methods by spectrograms and spectral envelopes

Several methods can be utilized for visual comparison of effectiveness of different approaches for noise reduction in the speech. A spectrogram is a typical representative graph for comparison in the time/frequency domain. The approach of visual comparison of the whole spectrograms has a disadvantage in the subjectivity (strong dependence on the person who makes the matching). To decrease this subjectivity, the detailed analysis of only the selected part–the nearest region of interest (ROI) area—is necessary. In this manner, more precise matching of differences can be obtained—see an example in **Figure 8**.
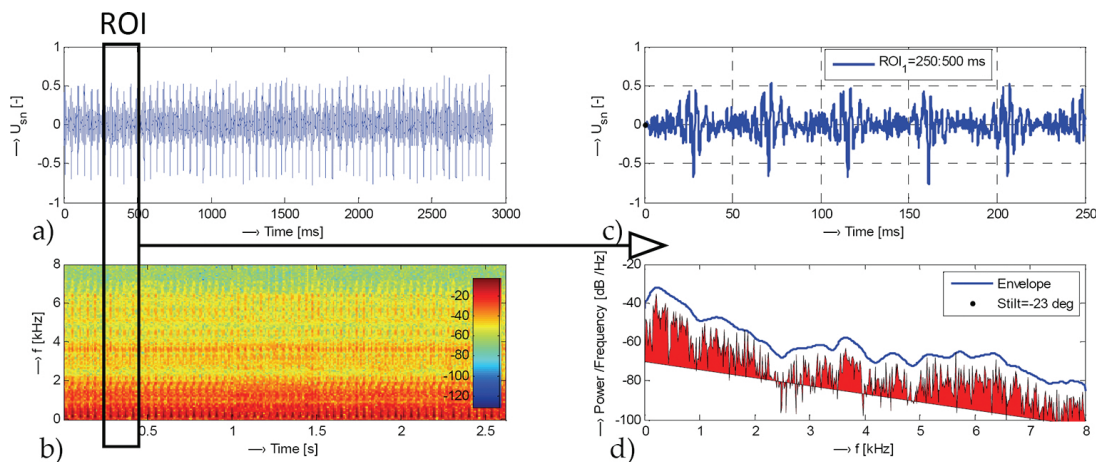

Figure 8. Example of speech and noise signal processing (f_s = 16 kHz): the whole recorded signal (a), the corresponding spectrogram (b), the selected ROI in 250-ms time interval (c) and the calculated spectral density together with its envelope and a tilt in degrees (d).

Generally, it can be said that the periodogram represents an estimate of the power spectral density of the input signal. When the N_{FFT}-point FFT algorithm is used for computing of the PSD as $S(e^{j\omega})/f_s$ (with sampling frequency f_s in [Hz]), we can obtain the resulting spectral density in logarithmic scale expressed in [dB/Hz]. These PSD graphs can be used for visual comparison as well as for numerical matching with the help of the calculated spectral difference (ΔP) or using other additional spectral parameters (spectral tilt). To obtain the smoothed spectral envelope, the Welch mean periodogram in [dB] can be used, too, as documented by the right graph in **Figure 9d**. The resulting spectral envelopes can be used for subsequent comparison. For exact numerical comparison (objective matching method), it is possible to calculate the difference between the spectral envelopes, for example between the original noisy speech signal and the signal after noise suppression. It holds also in this case of comparison that evaluation in selected frequency sub-ranges must be applied for obtaining more precise results. Therefore, the comparison of spectral envelopes was divided to three ranges: the full frequency

range up to $f_s/2$ (0–8k), the low-frequency sub-band up to 2.5 kHz (0–2k) and the middle frequency sub-band of 2~6 kHz (2–6k)–when the sampling frequency is f_s = 16 kHz. Then, the spectral distances D_{RMS} (by the RMS method) between these envelopes were calculated to compare the noise suppression methods for all three frequency ranges. In the case of the low-frequency band of 0–2k, the frequency F_{max} corresponding to the maximum difference ΔP_{max} was localized and determined, as documented by an example in **Figure 9**.
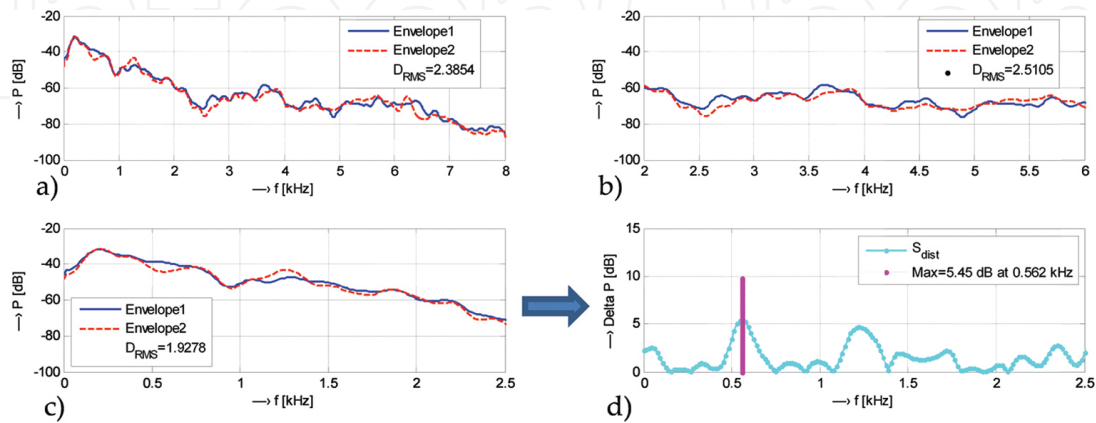

Figure 9. Example of visualization and calculation of differences between two spectral envelopes of the speech signals with the acoustic noise in the full frequency range 0~f_s/2 (a), in the middle frequency sub-band of 2~6 kHz (b), in the low-frequency band 0~2.5 kHz (c) and the spectral distance between envelopes in the low-frequency band (d).

3. Performed noise measurements and speech recording experiments

The open-air MRI equipment E-scan OPERA working with the weak magnetic field of 0.178 Tesla [17] was used for realization of all the experiments and measurements. This type of the MRI device includes also an adjustable bed, which can be positioned in the range from 0° (at the left corner near the temperature stabilizer) to 180°—see the overview photograph in **Figure 1a** and a principal angle diagram of the MRI scanning area in **Figure 10c**. The temperature stabilizer produces an additional acoustic noise, but its sound pressure level L_0 is almost constant so it can be easily subtracted as a background. As the noise depends on the position of the measuring microphone, the directional pattern of the noise source in the MRI scanning area must be measured first. On the basis of the obtained results, the design of the final arrangement of the recording experiment is carried out. It means that the optimal recording microphone position and parameters (the distance from the central point of the MRI scanning area, the direction angle, the working height and the type of the microphone pickup pattern) must be chosen. Therefore, the performed analysis of the acoustic noise produced by the gradient system of the MRI device consists of two phases. The first step is the auxiliary measurement of the acoustic noise distribution in the MRI scanning area and the mapping of the noise sound pressure level (SPL) in the MRI neighbourhood (see the arrangement photograph and principal angle diagram of the MRI scanning area in **Figure 10**):

- Directional pattern of the acoustic noise in the MRI neighbourhood in the angle intervals of 12.5° at the distances of {45, 60 and 75 cm} from the central point of the scanning area, at the height of 85 cm from the floor.

- Directional pattern of the acoustic noise in the angle intervals of 12.5° at the distance of 60 cm in three measurement heights of {75, 85 and 95 cm} from the floor corresponding to the sound level meter location on the level of the bottom gradient coil, in the middle between both coils and finally on the level of the upper coil.

- Noise SPLs at the directions {30°, 90° and 150°} at the distance of 60 cm in the height of 85 cm with different male/female testing persons lying in the scanning area of the MRI device without phonation; comparison with values obtained by the measurement using the water phantom.
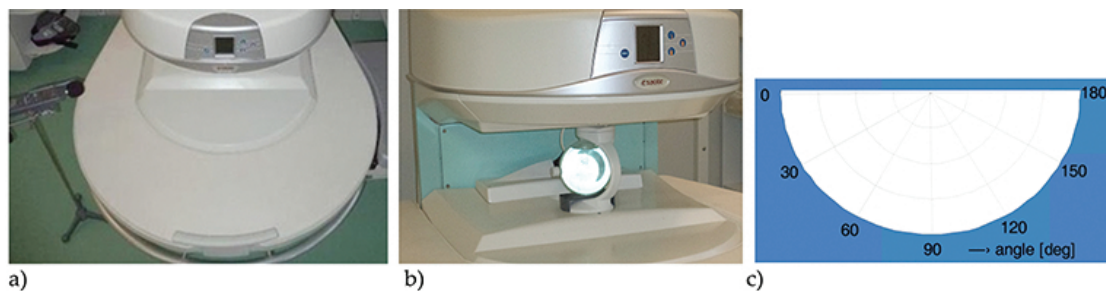

Figure 10. Arrangement photograph of the SPL distribution measurement in the MRI Opera neighbourhood with the sound level meter at the position of 30° in the 45 cm distance from the middle point of the scanner (a), detailed photograph of the MRI scanning area with a testing water phantom inserted in the RF coil (b) and a principal angle diagram of the MRI scanning area (c).

The sound level meter of the multi-function environment meter Lafayette DT 8820 (with the range set to 35–100 dB) was used for the noise SPL measurement and mapping. In the first two experiments, the spherical phantom weighing about 0.75 kg filled with doped water solution was placed in the RF knee coil (see the photograph in **Figure 10b**), and the measurement was carried out during the MR scan sequence SSF 3D execution (the patient bed was at the position of 180).

Afterwards, three basic types of noise and speech signal recordings were realized:

1. The noise produced by the gradient coils recorded by one pick-up microphone only—the tested person in the scanning area of the MRI device without phonation—the Mic. 1 located at 150.

2. The signals recording by the two pick-up microphones during execution of a scan MR sequence and phonation of the person lying in the scanning area of the MRI device (the noise and speech by Mic. 1 located at 90 and the noise only by Mic. 2 at 180 behind the lying person to minimize crosstalk of the speech signal).

3. The phonation signal with an additive noise recording by one pick-up microphone (Mic. 1 at 90—see the middle photograph in **Figure 1b**).

The demonstration examples of typical speech and noise waveforms for all three above-mentioned noise reduction methods are shown in **Figure 11**.

The electrical signals from the pick-up 1 'Behringer dual diaphragm condenser microphone B-2 PRO (as Mic. 1) and the RØDE NTK 1' condenser microphone (Mic. 2–in the case of two-way noise reduction method, both with a cardioid directional pattern) were recorded by means of the Behringer XENYX 502 analogue mixing console connected to a notebook by the USB interface UCA 202. Synchronization of the noise and/or speech recording and the MR scan process was realized manually by the console operator. The noise as well as speech signals was originally recorded at 32 kHz and then resampled to 16 kHz. A typical duration of the recorded signal was about 30 s, and for further signal processing, the stationary parts lasting 15 s were selected using the sound editor program Sound Forge 8.0.
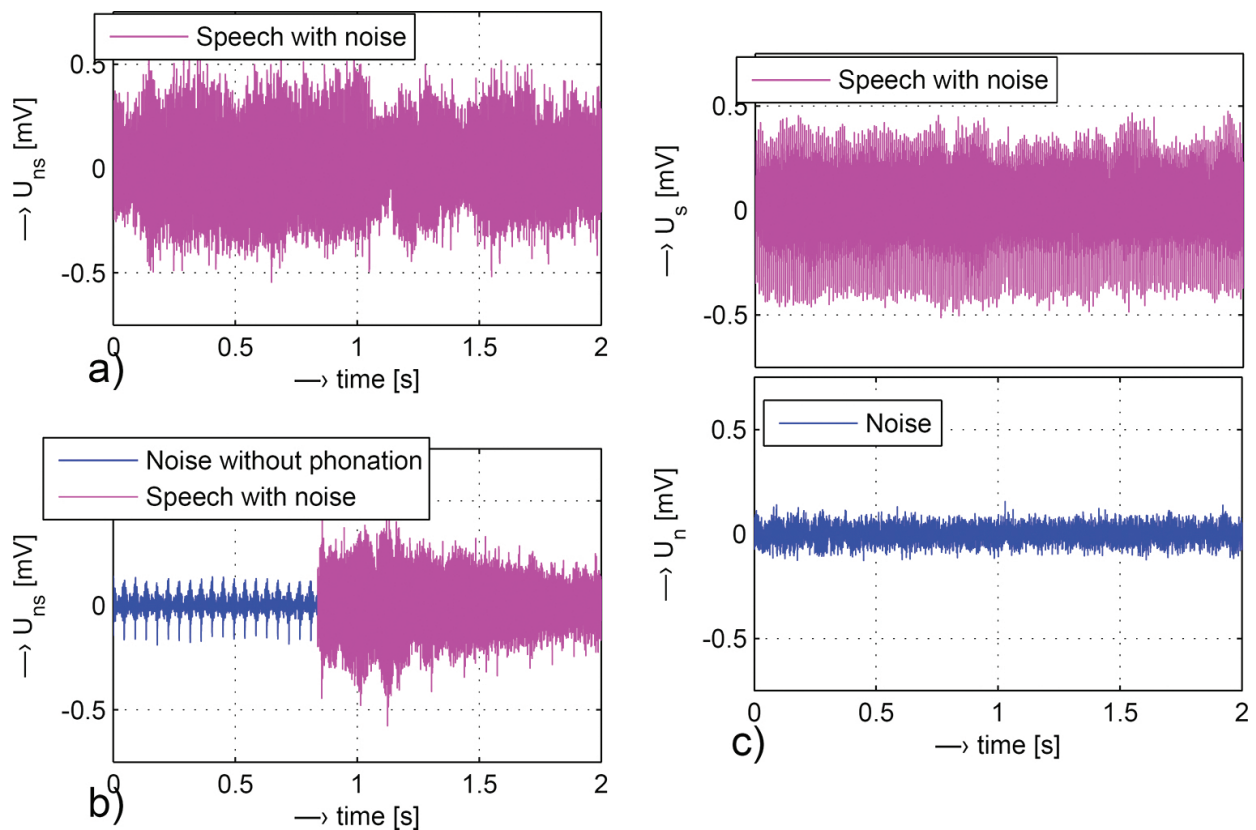

Figure 11. Examples of typical speech and noise waveforms for three analysed noise reduction methods: one-channel recording of the speech with noise (a), one-channel consecutive recording of an acoustic noise (without phonation) and the superimposed speech signal with noise (b) and the speech with noise and the noise signals recorded by two micro-phones in two separate channels during phonation of a vowel "a" by a male person (c), the used scan MR sequence 3D SSF.

Two corpora of the recorded signals were created by the described methods. The first one consists of the recorded noise signals only. These signals were used in two auxiliary experiments:

1. Analysis of noise supplementary spectral properties for different weights of the examined persons lying in the MRI device scanning area.

2. Analysis of the influence of the basic scan parameters (TR and TE times of the used MR sequence) on the spectral properties of the generated acoustic noise.

The noise signals brought about by one type of the MR scan sequence with different settings of TR = {10, 20, 30 and 40 ms} and then TE = {10, 14, 18 and 22 ms} were processed. From the recorded pure noise signals, the spectral envelopes and supplementary spectral properties were calculated and compared. The obtained values were next analysed statistically, and their histograms were calculated.

Person	$M1_{(JP)}$	$M2_{(TD)}$	$M3_{(LV)}$	$F1_{(AP)}$	$F2_{(BB)}$	$F3_{(ZS)}$
$F0_{mean}$ [Hz]	133	127	98	228	177	207
Weight [kg]	78	75	80	53	50	55

Table 1. The mean $F0$ values and approximate weights of the persons participated in the experiment.

The second corpus consists of the speech signal collected in dependence on the tested noise reduction method: one or/and two channels—using only one or/and two pick-up microphones for signal recoding. They all originate from the recordings of six voluntary persons lying in the MRI scanning area (3 + 3 male/female) with different mean $F0$ values and different approximate weights—see **Table 1**. The main database of long Czech and Slovak vowels was consequently used for creation of the second one consisting of manually selected ROIs corresponding to the stationary parts of the vowels 'a', 'e', 'i', 'o' and 'u' (typically three to five parts with the mean duration of 3 s). Due to different $F0$ values of the persons participating in our experiment, different parameters of the window length w_L and the window overlapping w_O must be set—we use 24-ms frames for processing of the male voice speech signal and 20-ms frames for the female one. The parameter N_0 for limitation of the real cepstrum was chosen in correspondence with the period of the noise part of the signal L_N equal to 256 (when $N_{FFT} = 1024$) for both types of voices. The whole comparison process consists of three steps:

1. Visual comparison of the calculated spectrograms and periodograms using different frequency ranges: up to 2.5 kHz or in the whole bandwidth up to half the sampling frequency.

2. Numerical matching of the spectral distances as well as the signal RMS values.

3. Localization and determination of the maximum difference ΔP_{max} between spectral envelopes of the original noisy speech signal and the signal after noise suppression (in the low-frequency band) for final comparison of effectiveness of all three described noise reduction methods.

4. Discussion of results

From the performed auxiliary measurements of the acoustic noise SPL, directional pattern in the three distances follows that the maximum level of about 72 dB was achieved for the nearest location of the SPL meter (D_L = 45 cm), while the background noise SPL_0 originated from the temperature stabilizer reached approximately 52 dB (measured in the time instant when the no scan sequence is executed)—see the left graph in **Figure 12**. To prevent any interaction between the ferromagnetic pieces of the measuring devices (it holds mainly for the recording microphones) and the stationary magnetic field of the MRI equipment, we eliminated this close distance for future use in our experiments. On the other hand, differences between the measured SPLs in 60- and 75-cm distances were small; therefore, we finally chose the middle distance of D_L = 60 cm as the basic one for next measurements and recording experiments. Obtained results of the second auxiliary experiment are in the correspondence with our expectation that the maximum SPL should be achieved at the height h = 85 cm (in the middle between the upper and the lower gradient coils and the permanent magnets of the MRI device) where the noise from two components is superimposed, while for the SPL meter located at the heights of 75 and 95 cm, the noise distribution is practically the same and about 3 dB lower than at the middle height—compare the right directional pattern in the **Figure 12**.
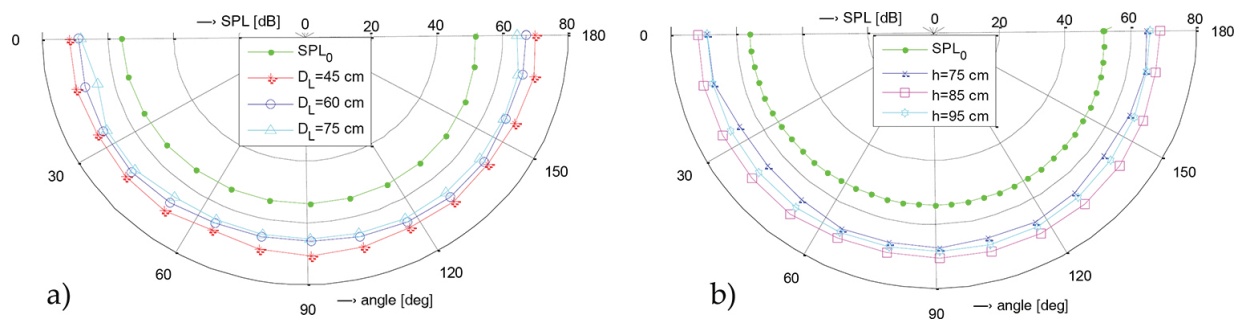

Figure 12. Measured directional patterns of the MRI gradient coils noise together with the background noise SPL_0: for three SPL meter distances at the height of 85 cm (a) and for three SPL meter heights at the distance of D_L = 60 cm (b).

Within the next auxiliary measurement, the influence of a weight of a lying person/water phantom in the scanning area of the MRI device on the vibration and subsequently generated acoustic noise SPL was analysed. The performed comparison of SPL measurement first of all shows that the maximum values are obtained at the position of 30° where the noise produced by the temperature stabilizer contributes for the most part of the resulting acoustic noise, and the minimum SPLs are achieved for the position of 15° with the maximum distance from the stabilizer. For all three measured directions, the noise SPL values measured with the examined person lying in the MRI scanning area were about 10 dB lower in comparison with the situation when the water phantom was used. When male and female testing persons lay on the bottom plastic holder of the permanent magnet and gradient coils, the obtained noise SPL values were roughly inversely proportional to their weights: the group of males with approximate weights of 78 kg and the group of females with the mean weight of 53 kg—as documented by the bar graph in **Figure 13**.
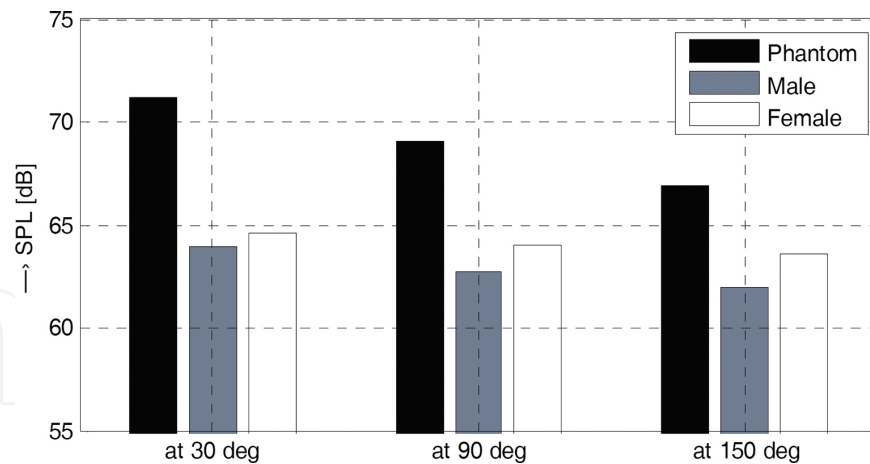
Figure 13. Comparison of the noise SPL values measured in the directions of 30, 90 and 150° using the water phantom, male and female persons lying in the scanning area of the MRI device without phonation, the SPL meter in the height of $h = 85$ cm, at the distance of $D_L = 60$ cm.
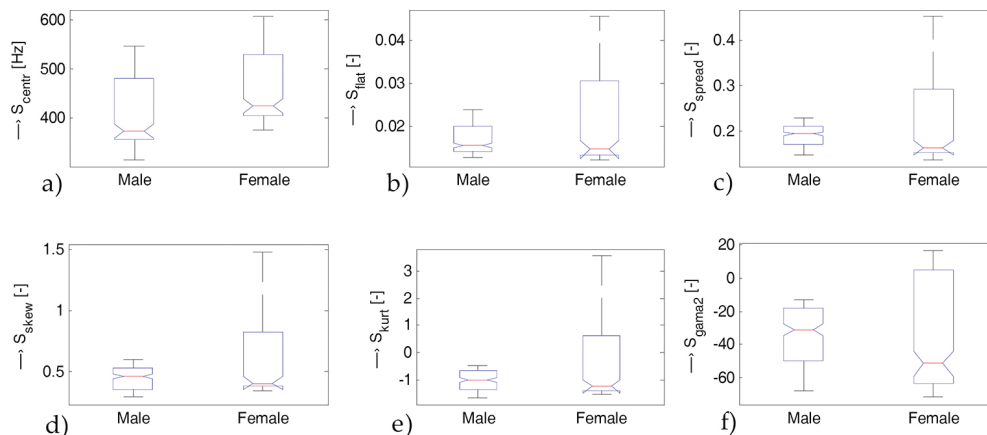

Figure 14. Box-plot of basic statistical properties of the basic and supplementary noise spectral parameters summarized for the groups of male/female persons lying in the scanning area of the MRI device: centroid (a), flatness (b), spread (c), skewness (d), kurtosis (e) and gama2 (f), signals recorded at the position of 150°.

Also another effect can be observed when the examined person lies in the scanning area of the open-air MRI and the holder of the lower gradient coils is loaded with his/her weight. The changed mass of the whole mechanical system affects its mechanical resonance modes, and the altered spectrum of the generated acoustic noise is apparent. To minimize the acoustic noise component caused by the temperature stabilizer, for further analysis of this phenomenon, the recording microphone was placed at 150° in the main experiment. The performed evaluation comprises comparison of the supplementary spectral properties for the weight pairs of male and female person groups—the box-plot of the basic statistical parameters (minimum, maximum, mean value and standard deviation) is shown in **Figure 14** and calculated histograms of the noise spectral properties is shown in **Figure 15**. The differences are most significant in the case of the spectral centroid feature with typical peaks of the histogram located around 375 Hz (for the male person group) and near 425 Hz for the female testing

person group. In addition, the deviation of the obtained values is higher for the female person group than for the male one. It can be explained by greater differences among the weights within this group (compare the approximate values in **Table 1**).
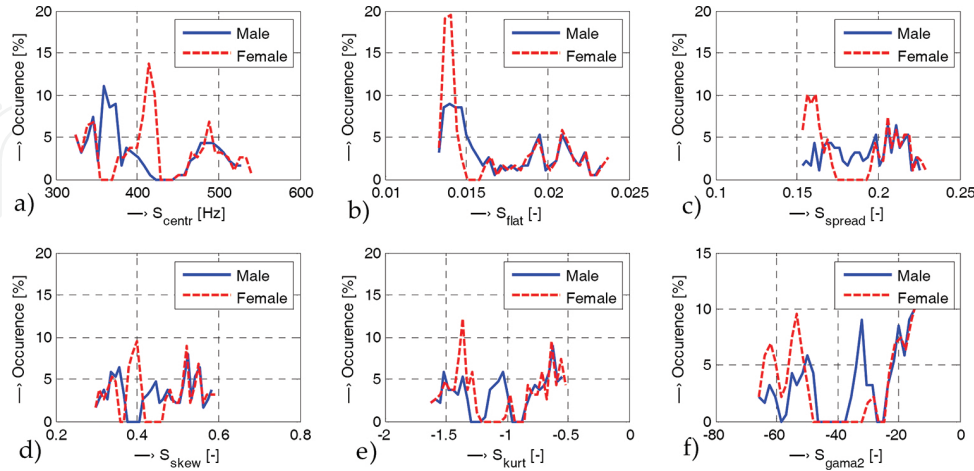

Figure 15. Histograms of the noise spectral properties summarized for the groups of male/female persons lying in the scanning area of the MRI device in correspondence with **Figure 14**: centroid (a), flatness (b), spread (c), skewness (d), kurtosis (e) and gama2 (f).
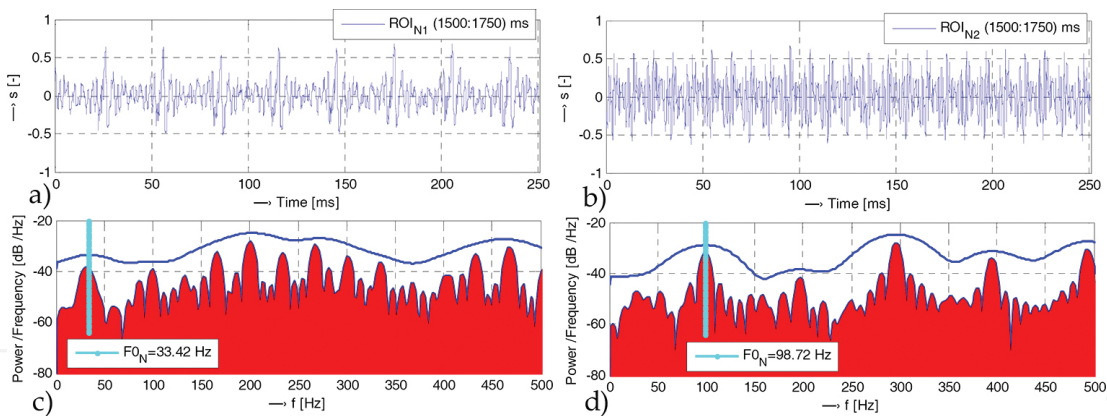

Figure 16. Detailed analysis of the influence of different TR parameter settings on changes in the basic spectral properties of generated noise signals: examples of vibration signals in the selected ROI in 250-ms time interval for TR = 33 ms (a) and TR = 10 ms (b), corresponding spectral densities including the spectral envelopes depicted for the frequency range up to 500 Hz together with the determined basic $F0_N$ frequencies (c and d), the used scan sequence of 3D SSF type with setting TR = 10 ms.

The aim of the last auxiliary experiment was to analyse how the setting of the basic scan parameters of the used MR sequence (TR and TE times) impacts the spectral properties of the generated acoustic noise. The performed detailed analysis shows that different settings of TR and TE parameters are manifested by changes of the noise spectral properties as follows:

- The repetition time has much influence on the basic $F0_N$ frequency of the generated periodical noise as documented by the graphs of the spectral density together with the

smoothed spectral envelopes as well as by the results of the statistical analysis of the $F0_N$ frequencies for different TR time settings (see **Figures 16** and **17**).
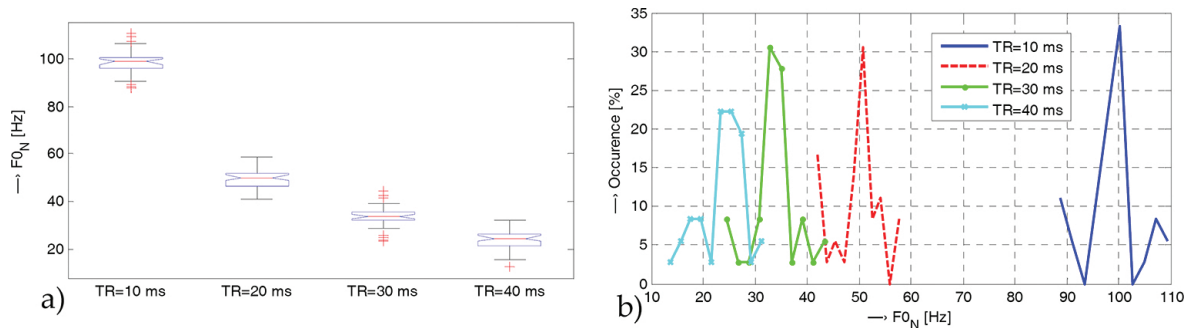

Figure 17. Summary comparison of influence of different TR parameter settings on the basic spectral properties of generated noise signals: box-plot of basic statistical properties of the determined basic $F0_N$ frequencies for TR = {10, 20, 30 and 40 ms} (a) and corresponding histograms (b), the processed scan sequence of 3D SSF type with TE = 10 ms.
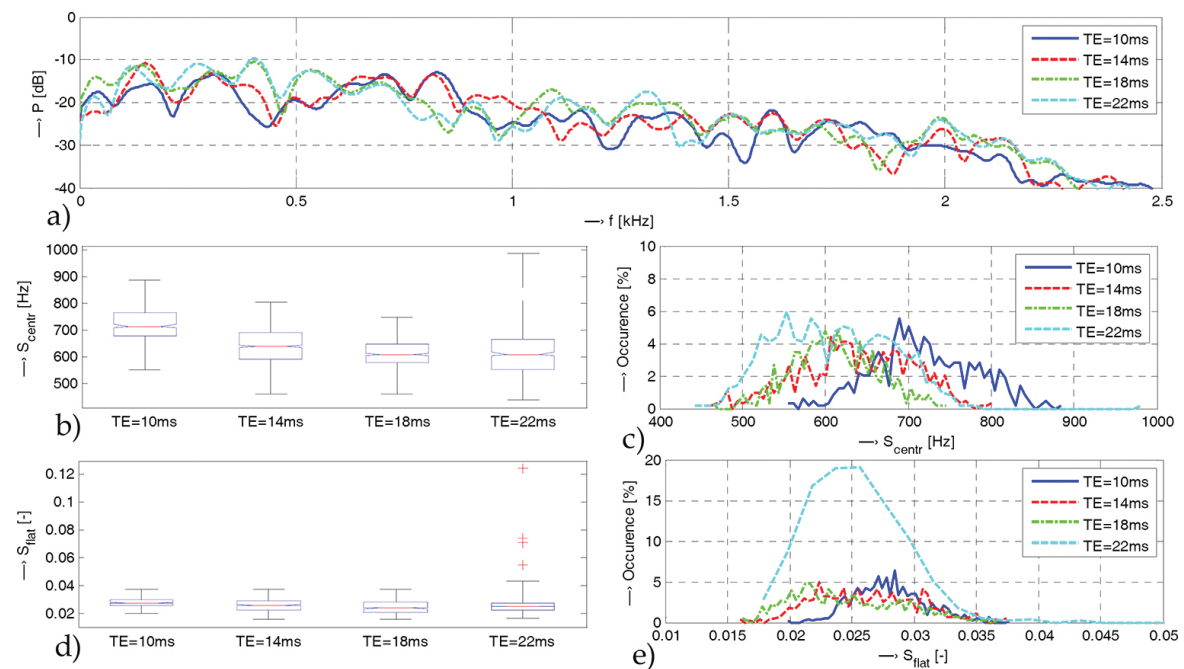

Figure 18. Influence of different TE parameter settings on the spectral properties of the generated noise: spectral envelopes depicted for the low-frequency band up to 2.5 kHz (a), box-plot of statistical parameters of S_{centr} (b), histograms of distribution of S_{centr} (c), box-plot of statistical parameters of S_{flat} (d) and histograms of distribution of S_{flat} (e), the used scan sequence of 3D SSF type with setting of TR = 30 ms and TE = {10, 14, 18 and 22 ms}.

- Echo time settings have effect on higher frequencies of the noise signal, as it is well documented by its supplementary spectral properties—see the box-plot of the basic statistical parameters of the centroid and the flatness together with their histograms in **Figure 18**.

The performed main comparison experiment confirms usability of all three applied noise reduction methods based on cepstral limitation and clipping, or using one-channel/two-

channel spectral subtraction. The significant differences between the noisy and the cleaned speech signals were observed for all processed samples as documented by the graphs in **Figures 19–21**, and by the resulting mean, spectral distances were calculated between spectral envelopes of vowels recorded in MRI noisy environment and cleaned by both noise reduction methods summarized in **Table 2**, as well as the calculated mean spectral differences ΔP at the basic noise frequency $F0_N$ for all analysed stationary parts of vowels listed in **Table 3** (for both speaker genders). Comparison of calculated spectral distances between the original noisy and the cleaned speech signals shows that spectral changes are higher when using the second noise reduction method. However, in practical realization, it is very difficult to fulfil the basic condition—to find a suitable arrangement of the Mic. 2 for the pure noise recoding. It means that there are some recording imperfections resulting in a crosstalk of the speech signal into the Mic. 2 intended only for noise recording. Next disadvantage of the spectral envelope subtraction method consists in the fact that this approach brings also undesirable effects (e.g. aliasing), and therefore, some frequency filtering must be subsequently carried out. This shortcoming can be solved also in the cepstral domain—by greater limitation of the real cepstrum expansion, but it would be contradictory to the requested higher quality of the finally reconstructed speech signal.
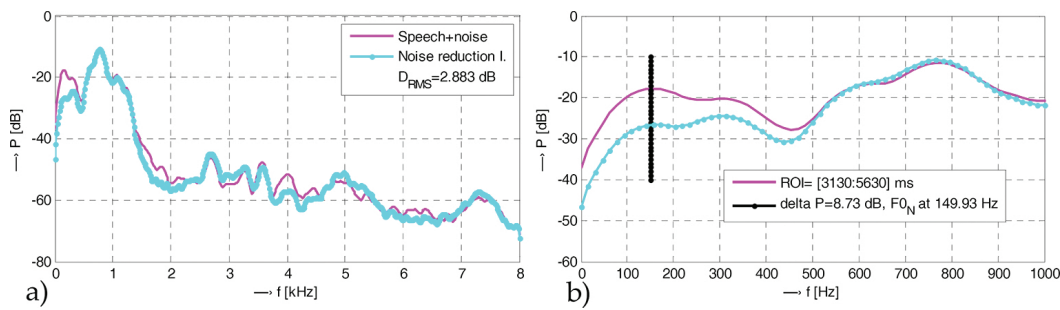

Figure 19. Comparison of spectral envelopes for noise reduction method I: summary view in the full frequency range up to $f_s/2$ (a), detailed part for the frequency range up to 1 kHz with the determined spectral difference ΔP at the frequency $F0_N$ (b), processed speech signal of the long vowel "o" (female voice) in the selected ROI of 2.5-s time interval, the used scan sequence of 3D SSF type, TR = 8 ms, f_s = 16 kHz.

Figure 20. Comparison of spectral envelopes for noise reduction method II: summary view in the full frequency range up to $f_s/2$ (a), detailed part for the frequency range up to 1 kHz with the determined spectral difference ΔP at the frequency $F0_N$ (b), processed speech signal of the long vowel "a" (male voice) in the selected ROI of 2.5-s time interval, the used scan sequence of 3D SSF type, TR = 10 ms, f_s = 16 kHz.
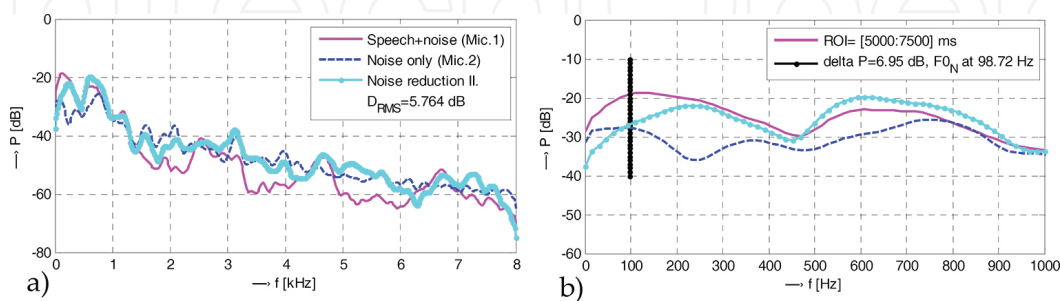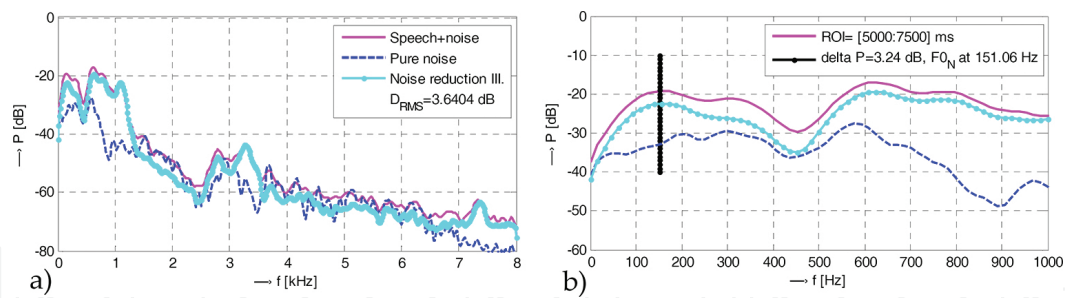
Figure 21. Comparison of spectral envelopes for noise reduction method III: summary view in the full frequency range up to $f_s/2$ (a), detailed part for the frequency range up to 1 kHz with the determined spectral difference ΔP at the frequency $F0_N$ (b), processed speech signal of the long vowel "o" (female voice) in the selected ROI of 2.5-s time interval, the used scan sequence of 3D SSF type, TR = 8 ms, f_s = 16 kHz.

Stationary part of vowel	Male voice			Female voice		
	Method I	Method II	Method III	Method I	Method II	Method III
a: D_{RMS} [dB]	7.32	5.58	4.21	5.45	5.33	3.87
e: D_{RMS} [dB]	6.49	5.24	3.97	4.86	5.01	3.82
i: D_{RMS} [dB]	7.36	5.63	4.14	5.27	5.14	3.10
o: D_{RMS} [dB]	5.66	4.98	4.02	6.11	4.86	3.19
u: D_{RMS} [dB]	4.73	4.89	3.81	5.46	4.78	4.42

Table 2. Comparison of the mean values of spectral distances of the analysed vowels (D_{RMS} are calculated between spectral envelopes of the original noisy and the cleaned speech signals) for male and female voices.

Stationary part of vowel	Male voice			Female voice		
	Method I	Method II	Method III	Method I	Method II	Method III
a: ΔP [dB]	6.38	7.22	3.26	4.89	6.72	2.84
e: ΔP [dB]	5.79	6.59	2.95	5.12	5.84	2.86
i: ΔP [dB]	6.11	7.23	3.45	5.64	6.18	3.28
o: ΔP [dB]	5.07	5.86	3.08	4.85	5.68	3.11
u: ΔP [dB]	5.58	4.97	2.86	5.34	5.02	2.85

Table 3. Results of mean spectral differences ΔP at the basic noise frequency $F0_N$ (determined between the spectral envelopes of the original noisy and the cleaned speech signals) for male and female voices.

Noise suppression based on cepstral peaks clipping can also cause suppression in the spectrum at the fundamental speech frequency $F0$ or at the first formant frequency $F1$ if they are close to the clipped noise frequency $F0_N$. Generally, it holds that the values of these frequencies $F0$ and $F1$ depend merely on the speaker vocal tract characteristics and they may not be affected by noise suppression in the speech signal. The third noise suppression method based on the noise estimation techniques using statistical evaluation of the spectral properties has a

principal handicap—this approach is not able to track real variations in the noise and its application in noise suppression results in an artificial residual fluctuating noise and a distorted speech.

5. Conclusions

The achieved results will serve to create databases of initial parameters (such as the filter bank for noise signal pre-processing) used to design a filter for noise suppression in the speech signal recorded simultaneously with 3D human vocal tract scanning. It will be useful in experimental practice when it often occurs that the basic parameter setting of the used scanning sequence as well as the other scanning parameters must be changed depending on the person that is currently examined. The advantage of the system is in the design method of the device noise elimination by specially developed software which enables to measure the acoustic characteristics of the subject during phonation in the MRI device with good accuracy. Finally, these three described algorithms have much a wider range of applications than indicated by the particular experimental arrangement in this paper.

For better knowledge of the acoustic noise conditions in the scanning area and in the vicinity of the MRI device, it is necessary to carry out additional measurement and experiments. The results obtained in this way will help to describe the process of the gradient coil electric excitation, the subsequent mechanical vibration and the resulting acoustic noise generation in the MRI device scanning area and its neighbourhood. There is also need for the knowledge about the contribution of the upper gradient coil (and its plastic holder) to the resulting acoustic noise. For this reason, the parallel measurement of the vibration signal on the surface of both plastic holders as well as the analysis of reflective properties of the metal shielding cage and its influence on the progression of the acoustic wave should be performed.

Acknowledgements

Author details

Jiří Přibil[1*], Anna Přibilová[2] and Ivan Frollo[1]

*Address all correspondence to: umerprib@savba.sk

1 Institute of Measurement Science, SAS, Bratislava, Slovakia

2 Faculty of Electrical Engineering and Information Technology, SUT, Bratislava, Slovakia

References

[1] Vampola T, Horáček J, Laukkanen AM, Švec JG. Human vocal tract resonances and the corresponding mode shapes investigated by three-dimensional finite-element modelling based on CT measurement. Logopedics Phoniatrics. 2015; 40:14–23.

[2] Zhu Y, Kim YC, Proctor MI, Narayanan SS, Navak KS. Dynamic 3-D visualization of vocal tract shape during speech. IEEE Transactions on Medical Imaging. 2013; 32:838–848.

[3] Burdumy M. et al. Acceleration of MRI of the vocal tract provides additional insight into articulator modifications. Journal of Magnetic Resonance Imaging. 2015; 42:925–935.

[4] Moelker A, Wielopolski PA, Pattynama MT. Relationship between magnetic field strength and magnetic-resonance-related acoustic noise levels. Magnetic Resonance Materials in Physics Biology and Medicine. 2003; 16:52–55.

[5] Kannan G, Milani AA, Panahi IMS, Briggs RW. An efficient feedback active noise control algorithm based on reduced-order linear predictive modeling of fMRI acoustic noise. IEEE Transactions on Biomedical Engineering. 2011; 53:3303–3309.

[6] Shou X. et al. The suppression of selected acoustic frequencies in MRI. Applied Acoustics. 2010; 71:191–200.

[7] Wu Z, Kim YC, Khoo MCK, Nayak KS. Evaluation of an independent linear model for acoustic noise on a conventional MRI scanner and implications for acoustic noise reduction. Magnetic Resonance in Medicine. 2014; 71:1613–1620.

[8] Přibil J, Horáček J, Horák P. Two methods of mechanical noise reduction of recorded speech during phonation in an MRI device. Measurement Science Review. 2011; 11:92–98.

[9] Montazeri V, Pathak N, Panahi I. Two-channel multi-stage speech enhancement for noisy fMRI environment. Canadian Journal of Electrical and Computer Engineering. 2013; 36:60–67.

[10] Sun G, Li M, Rudd BW, Lim TC, Osterhage J, Fugate EM, Lee JH. Adaptive speech enhancement using directional microphone in a 4-T scanner. Magnetic Resonance Materials in Physics, Biology and Medicine. 2015; 28:473–484.

[11] Wang J, Liu H, Zheng C, Li X. Spectral subtraction based on two-stage spectral estimation and modified cepstrum thresholding. Applied Acoustics. 2013; 74:450–458.

[12] Vích R, Přibil J, Smékal Z. New cepstral zero-pole vocal tract models for TTS synthesis. In: IEEE Region 8 EUROCON'2001; Bratislava, Slovakia, 2001. p. 458–462.

[13] Aalto D. et al. Large scale data acquisition of simultaneous MRI and speech. Applied Acoustics. 2014; 83:64–75.

[14] Narayanan S. et al. Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research. Journal Acoustical Society of America. 2014; 136:1307–1311.

[15] Přibil J, Přibilová A, Frollo I. Analysis of spectral properties of acoustic noise produced during magnetic resonance imaging. Applied Acoustics. 2012; 3:687–697. doi:10.1016/j.apacoust.2012.01.007

[16] Přibil J, Přibilová A, Frollo I. Influence of Person Weight on Spectral Properties of Acoustic Noise in the Open-air MRI. In: Holub J, editor. XXI IMEKO World Congress "Measurement in Research and Industry" – Full Papers; August 30–September 4, 2015; Prague, Czech Republic. Issue 1, 2015. 4 p.

[17] Esaote E-scan Opera. Image Quality and Sequences Manual. 830023522 Rev. A. Esaote S.p.A., Genoa, April 2008.