# AKUSTICKÉ LISTY

České akustické společnosti
www.czakustika.cz

ročník 24, číslo 1–2 červen 2018

## Obsah



ČESKÁ AKUSTICKÁ SPOLEČNOST

# Measurement of Nonlinear Distortion of low-cost Microphones

## Měření nelineárního zkreslení low-cost mikrofonů

Jakub Kolář[a] and Petr Honzík[b]

[a]Czech Technical University in Prague, Faculty of Electrical Engineering,
Technická 2, 166 27 Praha 6, Czech Republic
[b]Czech Technical University in Prague, Faculty of Transportation Sciences,
Konviktská 20, 110 00 Praha 1, Czech Republic

Measuring of the nonlinear distortion of microphones is affected by nonlinear distortion of the loudspeaker that emits the measuring signal because of serial combination of these two nonlinear systems. This paper describes an application of a method of nonlinear system identification in case of the series combination of two nonlinear systems, where the first one represents the loudspeaker and the second one represents the microphone. The second nonlinear system (microphone) is modeled by Hammerstein model, the coefficients of this model being identified by the method. From these frequency dependent coefficients the frequency dependence of the total harmonic distortion (THD) of the microphone is calculated.

## 1. Introduction

Knowing the nonlinear distortion of a microphone (total harmonic distortion – THD) is useful in many applications. This work has been motivated by the use of low-cost electret microphones in wireless sensor networks for noise monitoring [1], where the nonlinear distortion can affect the accuracy of the measurements in the higher range of the noise level.

Measuring of the nonlinear distortion of microphones is affected by nonlinear distortion of the loudspeaker because of serial combination of these two nonlinear systems (see figure 1). This is the reason why we cannot consider that measuring of nonlinear distortion of the microphone is standard analysis of one nonlinear block with known input signal $x(t)$ a output signal $y(t)$ as it is presented in [2] and [3]. The application of nonlinear system identification in the case of the series combination of two nonlinear systems based on method presented in [4] and using swept sine method for identification presented in [2] and [3] is proposed herein. Using these methods we can describe the microphone (the second nonlinear system) by Hammerstein model and calculate coefficients of this model. Knowing this coefficients we can model the microphone output voltage for every harmonic component using Hammerstein coefficients and acoustic pressure at microphone input. From the voltage amplitude of harmonic components we can easily calculate frequency dependence of THD of the microphone, which is not affected by the nonlinear distortion of the loudspeaker [5, 6].

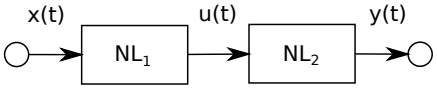



Figure 1: Two nonlinear systems in series

This Introduction is followed by section 2 where the method used herein to obtain the frequency dependent THD is briefly described. The section 3 presents the experimental setup and results. After the Conclusion the method for identification of the coefficient of the Hammerstein nonlinear model is shortly reviewed in the Appendix.

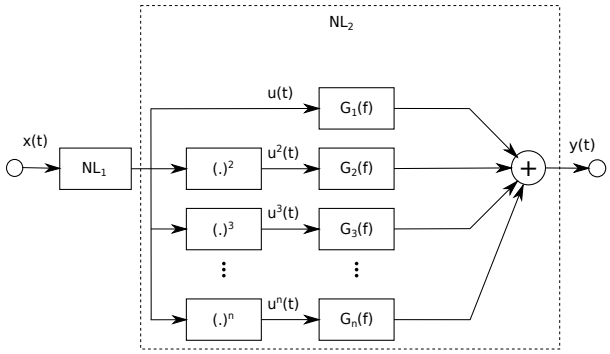



Figure 2: Two nonlinear systems in series, the second one represented by a generalized Hammerstein model

## 2. Theoretical background

In this paper we propose an application of the method of nonlinear system identification [4], which is shortly reviewed in Appendix, for identification of nonlinear distortion of microphone. In this case, we can also use general schema of nonlinear systems in series, as shown in figure 2, where $NL_1$ is the loudspeaker and $NL_2$ is the analyzed microphone described by the Hammerstein model of nonlinear system in this application.

Přijato 7. listopadu 2017, akceptováno 20. května 2018.

For creating the Hammerstein model of the microphone we need to know signals $x(t)$, $y(t)$ and $u(t)$, as shown in Appendix. Signal $x(t)$ (generated measuring signal supplying the loudspeaker) is generally known, signal $y(t)$ is the signal from the analyzed microphone. In order to capture the signal $u(t)$ we need to use a calibrated a reference microphone whose nonlinear distortion is negligible comparing to the one of the measured microphone. The reference microphone is placed as close as possible to the measured microphone. This arrangement is necessary because of minimization of acoustic pressure difference between microphones.

The exciting exponential swept-sine signal is used for supplying the loudspeaker. The coefficients of the Hammerstein model are calculated from the signals captured on both microphones using procedure described in the Appendix. However, these coefficients describe only transmission of harmonic components, so these coefficients can not be used for comparison of nonlinear distortion between microphones.

For comparison between the microphones we need to calculate THD frequency dependence of the measured microphone. If the input signal $x(t)$ is defined as $x(t) = P.\sin(\omega_0 t)$, where $P$ is amplitude of the acoustic pressure and $\omega_0$ is angular frequency, defined as $\omega_0 = 2\pi f$, then the Fourier image of the output signal $y(t)$ of measured microphone can be calculated using Hammerstein model as

$$Y(\omega) = G_1(\omega)\, P\, \mathcal{F}(\sin(\omega_0 t)) + G_2(\omega)\, P^2\, \mathcal{F}(\sin^2(\omega_0 t)) + G_3(\omega)\, P^3\, \mathcal{F}(\sin^3(\omega_0 t)) + \ldots, \quad (1)$$

where $\mathcal{F}(.)$ represent the Fourier transformation and $G_n(\omega)$ are the coefficients of Hammerstein model as described in the Appendix. The Fourier images of the sine function and its second and third power are given by [7]

$$\mathcal{F}(\sin(\omega_0 t)) = -\,\mathrm{j}\frac{\delta(\omega+\omega_0)}{2} + \mathrm{j}\frac{\delta(\omega-\omega_0)}{2}, \quad (2)$$

$$\mathcal{F}(\sin^2(\omega_0 t)) = -\,\frac{\delta(\omega+2\omega_0)}{4} + \frac{\delta(\omega)}{2} - \frac{\delta(\omega-2\omega_0)}{4}, \quad (3)$$

$$\mathcal{F}(\sin^3(\omega_0 t)) = \mathrm{j}\frac{1}{8}\delta(\omega+3\omega_0) - \mathrm{j}\frac{3\delta(\omega+\omega_0)}{8} + \mathrm{j}\frac{3\delta(\omega-\omega_0)}{8} - \mathrm{j}\frac{1}{8}\delta(\omega-3\omega_0). \quad (4)$$

The amplitudes of the DC component and the first, second and third harmonics can be expressed as follows

$$V_0 = \tfrac{1}{2}G_2(0)P^2, \quad (5)$$

$$V_1(\omega_0) = G_1(\omega_0)P + \tfrac{3}{4}G_3(\omega_0)P^3, \quad (6)$$

$$V_2(2\omega_0) = \tfrac{1}{2}\mathrm{j}G_2(2\omega_0)P^2, \quad (7)$$

$$V_3(3\omega_0) = -\tfrac{1}{4}G_3(3\omega_0)P^3. \quad (8)$$

Now, we can use the amplitudes calculated using formulas (5–8) for calculating THD, defined as

$$\mathrm{THD} = \frac{\sqrt{|V_2|^2 + |V_3|^2 + \cdots + |V_n|^2}}{|V_1|} \cdot 100, \quad (9)$$

respective the frequency dependence of THD, defined as

$$\mathrm{THD}(\omega) = \frac{\sqrt{|V_2(2\omega)|^2 + |V_3(3\omega)|^2 + \cdots + |V_n(n\omega)|^2}}{|V_1(\omega)|} \cdot 100 \quad (10)$$

in percentages [8].

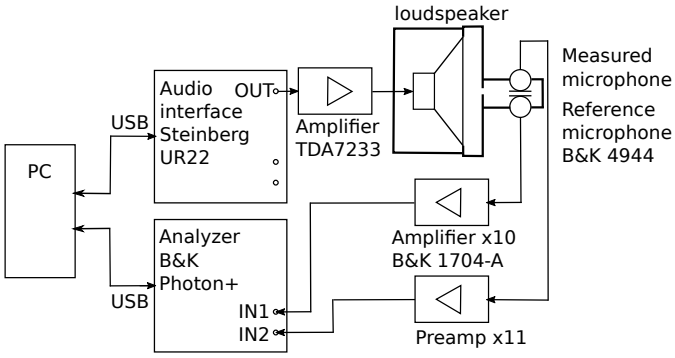



Figure 3: Scheme of measuring setup

## 3. Experimental results

Whole measurement was performed using setup shown in figure 3. The exponential swept-sine in range 15 Hz to 15 kHz, defined in [2, 4], generated in MATLAB software was used as the exciting signal. The loudspeaker Visaton K28WP 8 Ω was supplied by this signal through the Steinberg audio interface UR22 and the power amplified TDA7233. Both the analyzed and reference microphones were placed in a small cavity of volume $7.1{\cdot}10^{-6}$ m$^3$ connected with the acoustic output of the loudspeaker by a narrow hole in order to achieve spatially uniform and relatively high acoustic pressure at the input of the microphones even with the small low-cost loudspeaker. The measured microphone is the low-cost electret microphone MCA 2500 with the circular membrane of the dimensions similar to the standard 1/4" microphones. The signal from this microphone is amplified by a preamplifier containing the operational amplifier TS971 with gain of 11. Thus, the nonlinear distortion of the serial combination of the measured microphone and the preamplifier is measured using this setup. However, in most of applications the preamplifier is always present, therefore it is logical to measure the distortion of the whole system containing the microphone and the preamplifier. Moreover, the nonlinearities caused by the preamplifier are supposed to be negligible comparing to the nonlinearities originating in the measured microphone. The measuring calibrated microphone B&K DeltaTron Pressure-field 1/4" type 4944B with B&K amplifier 1704 (the gain being set to x10) was used as the reference microphone. Note that this type of microphone is well suitable for the measurements at high acoustic pressures having an upper limit of dynamic range of 169 dB with 3 % THD at this level. Therefore the distortion of the reference signal path (reference microphone and the

B&K amplifier) can be considered negligible comparing to the distortion originating in the measured microphone signal path. Signals of both microphones were captured by analyser Brüel Kjær Photon+ (whose contribution to the distortion of both signal paths is supposed to be negligible) and digitized using 96 kHz sampling frequency. Measured data, were imported to the MATLAB software and processed using proposed method.

The coefficients of the Hammerstein model $G_n(\omega)$ were calculated using the method described in Appendix. The frequency dependence of these coefficients in full range is shown in figure 4 (the physical meaning of the coefficients being equivalent to the acoustic pressure sensitivity, thus the unit of the coefficients being V/Pa). However, the coefficients out of the range 100 Hz – 5 kHz are significantly affected by noise. This can be caused by the following reasons: i) the small loudspeaker is not able to emit sufficiently high level of measuring signal at low frequencies, ii) the measurement has been performed in standard room with some level of background noise which is concentrated at low frequencies and iii) the narrow hole connecting the loudspeaker and the cavity containing both microphones act as a low-pass filter, which causes an important attenuation of the measurement signal at high frequencies. Figure 5 shows the spectrum of the signal on the reference microphone (black line) compared with the spectrum of the background noise (grey line). At the low frequencies the signal seems to be high enough, but the signal to background noise ratio is poor. The measurement signal is highly attenuated above approximately 2 kHz. Therefore it seems that the reason i) is not relevant here but the reasons ii) and iii) play an important role.

From the coefficients of the Hammerstein model the THD has been calculated according to the procedure described in the section 2. Frequency dependence of THD, modelled for incident acoustic pressure of 117,5 dB SPL, is shown in figure 6 in the frequency range of 100 Hz to 5 kHz. Two curves in this figure represents the THD calculated using $G_2$ and $G_3$ (grey curve) and using $G_2$ only (black curve) owing to the fact that the coefficient corresponding to the third harmonic $G_3$ is more affected by the noise then the one corresponding to the second harmonic $G_2$ and that the second harmonic should theoretically predominate in the electrostatic transducers [5, 6]. Both curves are affected by the noise present in the estimated values of the coefficients $G_2$ and $G_3$, this noise being "amplified" by the multiplication with the second and the third power of the amplitude of the input signal $P$ in the equations (7) and (8). However, it can be seen that the estimated value of THD, regardless the noise, is slightly above 10 %, which is realistic value for this type of microphones. The method then shows a good potential to yield useful results provided that a better signal to noise ratio is reached. This can be achieved in future works by the free-field measurements in acoustically controlled conditions using the full-range powerful loudspeaker.

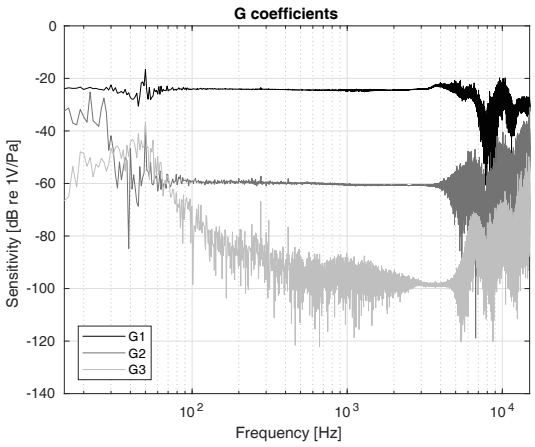



Figure 4: Frequency dependence of the estimated first three coefficients of the Hammerstein model $G_n(\omega)$

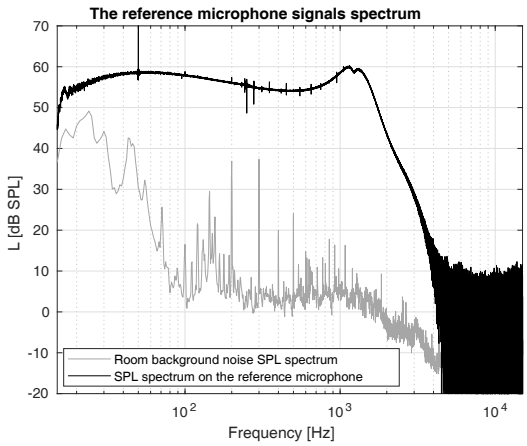



Figure 5: Spectrum of the reference microphone signal compared with the spectrum of the background noise

## 4. Conclusion

The application of the method of identification of nonlinear systems to the measurement of the nonlinear distortion of low-cost microphones is proposed herein. Using this method the parameters of the Hammerstein model of the microphone can be found. From this model the frequency dependent harmonic components of the output voltage and thus the frequency dependent THD of the microphone can be calculated for various input acoustic pressures. The requirement for high signal to noise ratio is the partial disadvantage of this method. The presence of the noise in the captured signal in the low and high frequency range is also the reason why in our measurement only the part between approximately 100 Hz and 2 kHz is correct.

Improvement of the noise immunity and extension of the frequency range of applicability of the method will be studied in the future work.

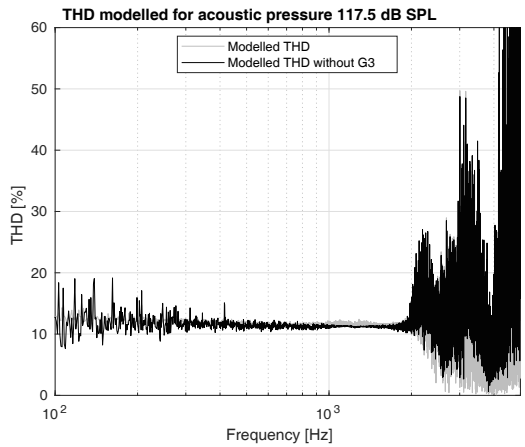



Figure 6: THD vs. frequency in range 100 Hz – 5 kHz

## Acknowledgement

This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS18/200/OHK2/3T/16. The authors are grateful to Dr. Antonín Novák from Laboratoire d'Acoustique de l'Université du Mans (LAUM) for helpful discussions.

## Appendix – A brief description of general method of identification of nonlinear systems in series [4]

Two nonlinear systems (NLS) connected in series are considered, as depicted in figure 1, where $x(t)$ represents the input signal of the system, $u(t)$ represents the output signal of the first NLS and $y(t)$ represents the output signal of whole system.

The method of nonlinear identification [4] allows the identification of the second NLS. The second NLS can be described by $N$-th order generalized Hammerstein model [2, 4], as depicted in figure 2. The identification of the system is then equivalent to estimating the linear filters transfer functions $G_n(f)$, which describe transmission of $n$-th harmonic, $n = 1, \ldots, N$. The estimating can be performed using the known signals $x(t)$, $u(t)$ and $y(t)$ and the formula

$$\begin{bmatrix} H_1^{(y,x)}(f) \\ H_2^{(y,x)}(f) \\ \vdots \\ H_L^{(y,x)}(f) \end{bmatrix} = \begin{bmatrix} H_1^{(u,x)}(f) & H_1^{(u^2,x)}(f) & \cdots & H_1^{(u^n,x)}(f) \\ H_2^{(u,x)}(f) & H_2^{(u^2,x)}(f) & \cdots & H_2^{(u^n,x)}(f) \\ \vdots & \vdots & \ddots & \vdots \\ H_L^{(u,x)}(f) & H_L^{(u^2,x)}(f) & \cdots & H_L^{(u^n,x)}(f) \end{bmatrix} \cdot \begin{bmatrix} G_1(f) \\ G_2(f) \\ \vdots \\ G_N(f) \end{bmatrix} . \tag{A.1}$$

The components $H_l^{(y,x)}$, $l = 1, \ldots, L$, are the Higher Harmonic Frequency Responses (HHFRs) [4] estimated between the signals $y(t)$ and $x(t)$ using Synchronized Swept-Sine Method [9], similarly components $H_l^{(u^n,x)}$ are the HHFRs [4] estimated between the signals $u(t)$ and $x(t)$. Matrix of components $G_n(f)$ is the matrix of the coefficients of the Hammerstein model. The equation A.1 can be solved using square matrix inversion under the condition that $N = L$.

For more information about the method refer to [4].

## References

[1] Brynda, P., Kopřiva, J., Horák, M.: *Trafficsensnet Sensor Network for Measuring Emissions from Transportation*, Procedia Engineering Special Issue Eurosensors 2015. Eurosensors 2015, Freiburg, 2015-09-06/2015-09-09. Oxford: Elsevier Ltd, 2015. p. 902–907. ISSN1877-7058.

[2] Novák, A., Simon, L., Kadlec, F.: *Nonlinear System Identification Using Exponential Swept-Sine Signal*, IEEE Transactions on Instrumentation and Measurement 59(8), 2010.

[3] Farina, A.: *Simultaneous measurement of impulse response and distortion with a swept-sine technique*, 108th AES Convention, Paris 18–22 February 2000.

[4] Novák, A., Maillou, B., Lotton, P., Simon, L.: *Nonparametric Identification of Nonlinear Systems in Series*, IEEE Transactions on Instrumentation and Measurement 63 (8), p. 2044–2051, 2014

[5] Škvor, Z: *Elektroakustika a akustika*, ČVUT, Praha, 2012.

[6] Beranek, L. L., Mellow, V. T.: *Acoustics: Sound Fields and Transducers*, Elsevier, UK, USA, 2012.

[7] Novák, A.: *Identification of Nonlinear Systems in Acoustics* (Doctoral dissertation), Czech Technical University in Prague/Université du Maine 2009. `http://cyberdoc.univ-lemans.fr/theses/2009/2009LEMA1009.pdf`

[8] Kolář, J.: *Měření nelineárního zkreslení mikrofonů*, bakalářská práce, České vysoké učení technické v Praze. 2016.

[9] Novák, A., Simon, L., Lotton, P.: *Analysis, synthesis, and classification of nonlinear systems using synchronized swept-sine method for audio effects*, EURASIP J. Adv. Signal Process, ID 793816, Feb. 2010.

# Numerical Techniques for the Assessment of the Environmental Noise Attenuation

## Numerické metody pro hodnocení útlumu hluku ve venkovním prostředí

Jan Šlechta and U. Peter Svensson

Acoustics Research Centre, Department of Electronic Systems, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway
e-mail: jan.slechta@ntnu.no

Environmental noise caused by road, railway and air traffic and industrial sources is an issue which is influencing the well-being of many inhabitants in the EU member states. Numerical methods are suitable for modelling the sound propagation in complex cases in which widely used empirical or semi-empirical methods fail. The boundary element method (BEM) promises to converge to the solution of the Helmholtz partial differential equation. On the other hand, the BEM is quite time demanding, especially in real-scale 3D cases. In this paper, the convenience of two alternative methods, the fast multipole boundary element method (FMBEM) and the edge source integral equation (ESIE) for the assessment of environmental noise levels is presented. Two test cases are studied with different levels of discretization: a cube and a noise barrier. The computation time, the relative error convergence and the method comparison are plotted for each test case. The paper demonstrates that while an accurate solution is obtained easily for a small test case (a cube) with an ordinary computer and all methods agree well for this test case, obtaining the solution of a large-scale test case (a noise barrier) with desired accuracy is excessively time-demanding.

## 1. Introduction

Noise is considered to be a challenging problem both by the European Parliament and local authorities of the Czech Republic. An important legal document for the noise mapping is the Directive 2002/49/EC [1] which started the process of strategic noise mapping and consequent preparation of action plans in member states of the European Union. Directive 2002/49/EC is dealing with the noise caused by road, railway and air traffic and industrial noise sources and recommends interim methods of noise prediction.

Strategic noise mapping raised a question of choosing a common method for predicting the sound pressure levels caused by noise sources described in the Directive 2002/49/EC. The first developed method for this purpose was Harmonoise [2]. Harmonoise was verified both by measurements and a combination of numerical methods (Harmonoise Reference model [3]). The Harmonoise Reference model used the parabolic equation for modelling the atmospheric refraction, a ray model for modelling the sound propagation and the boundary element method for analyzing the sound propagation in cases of complex obstacles.

Nevertheless, Harmonoise was not chosen as a common method for the strategic noise mapping, probably because of its excessive complexity. For this purpose, the method CNOSSOS-EU was developed [4]. However, the strategic noise mapping itself does not provide any noise mitigation. Particular measures should be stipulated by action plans. Ref. [5] claimed that noise barriers belong among the most effective ways to decrease the influence of traffic noise. Ref. [5] also categorized noise barriers among the least cost-efficient measures. Methods described in this paper are suitable for optimized noise barrier design as it is demonstrated on a noise barrier case study.

## 2. Fundamental Theory

An overview of the fundamental theory for three numerical methods (BEM, FMBEM and ESIE) is given in this section. All described methods are valid for homogeneous conditions of the sound propagation, as governed by the Helmholtz partial differential equation (PDE), which is a second order PDE [6]:

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} + k^2 p = 0, \tag{1}$$

where $k$ ($\mathrm{m}^{-1}$) is the wave number, $p$ (Pa) is the acoustic pressure and $x$, $y$, $z$ (m) are axis coordinates.

This PDE can be reformulated into the Kirchhoff-Helmholtz (K-H) integral, using Green's theorem. This K-H integral gives the sound pressure in a point $P$ as an integral over contributions by the sound pressure, $p(Q)$ and the normal velocity, $v(Q)$, on an enclosed boundary $S$ [7]:

$$C(P)\,p(P) = \int_S \left[\frac{\partial G(P,Q)}{\partial \mathbf{n}} p(Q) + \mathrm{i}kz_0 v(Q)\,G(P,Q)\right] \mathrm{d}S + 4\pi p^I(P), \tag{2}$$

Přijato 3. března 2016, akceptováno 13. června 2018.

where $C(P)$ (–) is a geometrical constant, $P$ is a point in the sound field, $Q$ is a point on the boundary, $G(P,Q)$ is the Green's function relating points $P$ and $Q$, $\mathbf{n}$ is the normal unit vector on the boundary, i is the imaginary unit, $z_0$ (Pa.s.m$^{-1}$) is the characteristic impedance of the medium and $p^I$ (Pa) is the incident sound pressure.

The Green's function, $G(R)$, describes sound radiation from a point source in a free field [7]:

$$G(R) = \frac{\mathrm{e}^{-\mathrm{i}kR}}{R}, \tag{3}$$

where $R$ (m) is the distance between the point source and the receiver.

The boundary element method (BEM) and the fast multipole method (FMBEM) belong to the category of wave-based methods, which solve Eq. (2), while the edge source integral equation (ESIE) is an extension of geometrical acoustics, decomposing the sound field into direct sound, reflected sound and diffraction of first and higher orders.

The FMBEM is based on the BEM approach but it adds an element clustering algorithm and calculates the interaction between element clusters. Speeding up the calculation is also achieved by adding the expansion of the Green's function for the far field interaction, iterative solvers, collocation points, etc. The ESIE uses the concept of edge sources and studies the sound scattering for convex 3D bodies. Analyzing concave bodies is problematic due to a phenomenon called the slope diffraction [8].

A broadband sound spectrum can be decomposed into a sum of pure tones. All the following methods are valid for pure tones. The Fourier transform is written [9]:

$$F(k_x) = \int_{-\infty}^{\infty} f(x)\, \mathrm{e}^{-\mathrm{i}k_x x}\mathrm{d}x, \tag{4}$$

where $F(k_x)$ is the Fourier transform of a function $f(x)$, $k_x$ (m$^{-1}$) is the wave number and $x$ (m) is the distance of the travelling sound wave.

### 2.1. Boundary Element Method

The BEM solves the Kirchhoff-Helmholtz integral for scattering or radiating problems, Eq. (2), in two steps. Acoustic variables are found on the domain boundary in the first step and in field points in the second step. The boundary domain is discretized into elements and the sound pressure is calculated in element nodes and integrated across the element. In both steps, the Green's function, Eq. (3), is employed to calculate the incident sound pressure.

Eq. (2) expresses that the normal velocity and the sound pressure on the boundary are related and therefore, only one of them or their ratio can be independently defined.

After prescribing Eqs. (2) and (3), the acoustic variables on the boundary are found as a solution of a matrix equation [10]:

$$\mathbf{Cp} = \mathbf{Ap} + \mathrm{i}kz_0\mathbf{Bv} + 4\pi\mathbf{p^I}, \tag{5}$$

where $\mathbf{C}$ is a coefficient matrix containing the geometrical constant $C(P)$, $\mathbf{A}$ and $\mathbf{B}$ are coefficient matrices resulting from kernel Eqs. (2) and (3). $\mathbf{p}$ and $\mathbf{v}$ are matrices containing the sound pressure and the normal velocity and $\mathbf{p^I}$ is a matrix containing the incident sound pressure.

The second step of the BEM calculation, in which the sound pressure in field points is found, is performed by integrating Eq. (2) again because Eq. (2) already defines the sound pressure in field points when the acoustic variables on the boundary are known.

In this paper, the implementation of the boundary element method OpenBEM [11] is used. OpenBEM is a collection of open-source Matlab codes for solving acoustical problems in 2D, 3D or axi-symmetric settings. OpenBEM makes use of the direct collocation method which is operating directly with the acoustic variables (acoustic pressure and velocity) [10]. Another option is the indirect variational method which substitutes the acoustic variables with parameters storing the variation of acoustic variables across the boundary [10].

OpenBEM provides a mesh generator only for axi-symmetric or 2D problems but a mesh for a 3D problem can be generated by the open-source software GMSH [12].

OpenBEM is a practical tool as it was demonstrated, besides other examples, by usage of the OpenBEM for the design of the Brüel & Kjær sound intensity calibrator Type 4297 [13].

### 2.2. Fast Multipole Boundary Element Method

The FMBEM is implemented as a part of the OpenBEM, even though it is still an experimental package which is not advertised on the OpenBEM webpage.

The fast multipole method is a relatively recent method [14] which is speeding up obtaining the solution of the Kirchhoff-Helmholtz integral. The FMBEM is based on an assumption that the solution is needed only with certain accuracy and it is intended for large scale problems. The FMBEM has its usage in electrostatics as well as in acoustics.

Unlike the conventional BEM, the FMBEM demands a lot of pre-processing before the sound pressure in collocation points is calculated. It is crucial for the pre-processing to be performed in an efficient way [15].

The FMBEM calculation process might be split into these steps [15]:

- The surface (i.e. the boundary) is divided into elements. This step is done in the same way as in the boundary element method.
- Clustering algorithm – surface elements are grouped together to constitute clusters.
- The multipole expansion is used to model the acoustic interaction of clusters which are considered to be far from each other.

- If the clusters are considered close to each other, the acoustic interaction is modelled with the ordinary BEM.
- The system of equations is solved with an iterative solver which is another tool for speeding up the FMBEM calculation. The sound pressure is calculated in the element collocation points and not in the nodes as by the ordinary BEM.
- The sound pressure in field points is obtained.

The fast multipole method is based on the multipole expansion which is related to the Green's function. This function comes out from the Gegenbauer's addition theorem which can be written as [16]:

$$\frac{\mathrm{e}^{-\mathrm{i}k|\vec{a}+\vec{d}|}}{\left|\vec{a}+\vec{d}\right|} = -\mathrm{i}k\sum_{l=0}^{\infty}(2l+1)(-1)^l\,\mathrm{h}_l^2(k|\vec{a}|)\,\mathrm{j}_l(k|\vec{d}|)\,\mathrm{P}_l(\hat{a}\cdot\hat{d}), \tag{6}$$

where $|\vec{a}+\vec{d}|$ is the distance between two points whose acoustic variables are put into a relation, $k$ ($\mathrm{m}^{-1}$) is the wave number, $\hat{a}$ and $\hat{d}$ are unit vectors of $\vec{a}$ and $\vec{d}$, $\mathrm{h}_l^2$ is the spherical Hankel function of the second kind and order $l$, $\mathrm{j}_l$ is the spherical Bessel function of the first kind and order $l$, $\mathrm{P}_l$ is a Legendre polynomial of order $l$.

Eq. (6) is a mathematically correct infinite sum and works well when $|\vec{a}| > |\vec{d}|$. Unfortunately, this infinite sum must be truncated in real calculations. Hence, an approximation in the calculation occurs and usage of the fast multipole expansion means that the result is always approximated.

The infinite sum is truncated after a specific number of terms $M$ which determines the speed of algorithm and precision. When the number of terms is low, the calculation is running fast but the precision suffers.

Quite a popular semi-empirical equation was developed to estimate the necessary number of evaluation terms $M_d$ [14]:

$$M_d = 2k\left|\vec{d}_{\max}\right| + \frac{\eta}{1.6}\ln\left(2k\cdot\left|\vec{d}_{\max}\right| + \pi\right), \tag{7}$$

where $\vec{d}_{\max}$ is the length of the largest $\vec{d}$ and $\eta$ is the demanded precision.

While the ordinary BEM makes use of classical Gauss elimination, the FMBEM does not explicitly express matrices which could be converted in a triangular shape. The FMBEM is also intended to work with high speed and that is why an iterative solver is needed.

Matlab contains an inbuilt function for the generalized minimum residual method (abbrev. GMRes) which is a suitable iterative solver for this task. The GMRes tries to solve a system of linear equations in the shape $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{A}$ is a square $n$ by $n$ matrix, $\mathbf{b}$ is a column vector of length $n$ and $\mathbf{x}$ is the result vector.

GMRes accepts these input parameters (except others): the maximum number of iterations before the restart, the maximum number of restarts and the predefined calculation tolerance. GMRes is restarted after an input number of iterations.

The temporary result after the predefined number of iterations is used in a new calculation cycle. A low number of iterations before the restart means that the solver is converging slowly but, on the other hand, a high number of iterations demands a large amount of memory.

### 2.3. Edge Source Integral Equation

The ESIE is a numerical method and Matlab toolbox published under the terms of GNU General Public License (also noted as the edge diffraction toolbox [17]).

The ESIE is based on geometrical acoustics and does not have non-uniqueness problems for the exterior domain as the BEM in which the BEM demands so-called CHIEF points (for further details see Ref. [18]). On the contrary, the ESIE has problems at certain numerically challenging receiver positions [19]. The ESIE also does not take into account the phenomenon called the slope diffraction which makes the ESIE inaccurate for non-convex shapes [8].

The edge source integral equation is based on the sound field decomposition into geometrical acoustic components (direct sound $p_{\mathrm{dir}}$ (Pa) and specular reflections $p_{\mathrm{ref}}$ (Pa)), the first order diffraction $p_{\mathrm{diff},1}$ (Pa) and the sum of $n$ higher orders of diffraction $p_{\mathrm{diff},n}$ (Pa):

$$p_{\mathrm{sum}} = p_{\mathrm{dir}} + p_{\mathrm{ref}} + p_{\mathrm{diff},1} + \sum_{i=2}^{n} p_{\mathrm{diff},i}, \tag{8}$$

where $p_{\mathrm{sum}}$ (Pa) is the resulting sound pressure in a receiver.

The direct sound is calculated with the Green's function (see Eq. (3)), with an additional visibility factor, which is 0 or 1. Specular reflections are modelled with mirror sources, that are also subject to visibility factors. It is possible to calculate scenarios with a Neumann boundary condition.

The first order diffracted sound is specified by a line integral over the set of edges $\Gamma$ [20]:

$$p_{\mathrm{diff},1} = -\frac{1}{4\pi}\int_{\Gamma}\nu_z\,V_{\mathrm{R},z}\,V_{z,\mathrm{S}}\,\frac{\mathrm{e}^{-\mathrm{i}kr_{\mathrm{R},z}}}{r_{\mathrm{R},z}}\cdot\frac{\mathrm{e}^{-\mathrm{i}kr_{z,\mathrm{S}}}}{r_{z,\mathrm{S}}}\cdot \beta(\mathrm{R},z,\mathrm{S})\,\mathrm{d}s_z, \tag{9}$$

where $\nu_z$ (–) is the wedge index of a wedge containing the point $z$, R is the receiver, S is the source, $r_{a,b}$ is the distance between points $a$ and $b$, $V_{a,b}$ is a visibility factor between points $a$ and $b$ and $\beta(\mathrm{R},z,\mathrm{S})$ is the directivity function.

Higher orders of diffraction are defined as [20]:

$$p_{\mathrm{diff},n} = -\frac{1}{4\pi}\int_{\Gamma}\int_{\Gamma} q(z_1,z_2)\,\frac{\mathrm{e}^{-\mathrm{i}kr_{z_1,z_2}}}{r_{z_1,z_2}}\cdot\frac{\mathrm{e}^{-\mathrm{i}kr_{\mathrm{R},z_1}}}{r_{\mathrm{R},z_1}}\cdot\nu_1\cdot \beta(\mathrm{R},z_1,z_2)\,\mathrm{d}s_{z_2}\,\mathrm{d}s_{z_1}, \tag{10}$$

where $q(z_1, z_2)$ is the edge source signal which can be calculated from an integral equation [20]:

$$q(z_1, z_2) = q_0(z_1, z_2) - \frac{1}{4\pi} \int_\Gamma q(z_2, z) \frac{\mathrm{e}^{-\mathrm{i}kr_{z_2,z}}}{r_{z_2,z}} \cdot \nu_2 \cdot \beta(z_1, z_2, z)\, \mathrm{d}s_z, \quad (11)$$

where $q_0(z_1, z_2)$ is the term expressing a contribution from the sound source [20]:

$$q_0(z_1, z_2) = -\frac{\nu_2}{4\pi} \frac{\mathrm{e}^{-\mathrm{i}kr_{z,\mathrm{S}}}}{r_{z,\mathrm{S}}} \beta(z_1, z_2, \mathrm{S}). \quad (12)$$

A sound source amplitude has been assumed in Eqs. (9) and (12).

## 3. Test cases

Two test cases are demonstrated to analyze properties of three used numerical methods: a cube and a noise barrier. Test case geometry, time convergence, relative-error convergence and method comparison is plotted for each test case.

Computation time is measured by Matlab stopwatch functions tic and toc. The time is displayed as a function of the number of elements to show how the computation time increases. The computation time is also presented in two tables.

Each method's relative error is presented as a function of the computation time. The relative error is calculated in two receivers for both cases. The reference result is always the result computed with the same method, with the highest number of mesh elements for the BEM and FMBEM, or the highest number of edge points for the ESIE. The relative error, $\epsilon_{\mathrm{rel}}$ (–), is defined as:

$$\epsilon_{\mathrm{rel}} = \left| \frac{p - p_{\mathrm{ref}}}{p_{\mathrm{ref}}} \right|, \quad (13)$$

where $p$ (Pa) is the sound pressure in the receiver for the relevant mesh and $p_{\mathrm{ref}}$ (Pa) is the reference sound pressure.

Methods are compared in the last figure for both cases. The method comparison for the cube test case is supposed to show that there are not significant differences in the insertion loss obtained by the used methods. The insertion loss is defined as the sound pressure level in the free field relative to the sound pressure level with the obstacle.

The comparison for the noise barrier test case is supposed to show that a convergence to the result with the demanded precision in the large-scale test case needs high amount of calculation time. The comparison between methods is done with the mesh of highest density (10204 elements for the cube and 19092 elements for the noise barrier) or the highest number of integration points (70 integration points per edge for the cube and 140 integration points per longest edge for the noise barrier).

All calculations were carried out on a desktop computer with an operating system Windows 10 and a processor Intel(R) Xeon(R) 3.4 GHz and 8.0 GB RAM.

### 3.1. Cube

The modelled cube together with receivers is shown in Fig. 1. The receivers are placed along a circle around the cube with the diameter 2.0 meters in the plane $z = 0$. The source is placed at coordinates $10^6 \cdot [1, 1, 1]$. The source is situated far from the cube in order to approximate the sound wave as a plane wave. The cube edge is 1.0 meter. The frequency of the source is 500 Hz and the sound speed is 344 m/s.

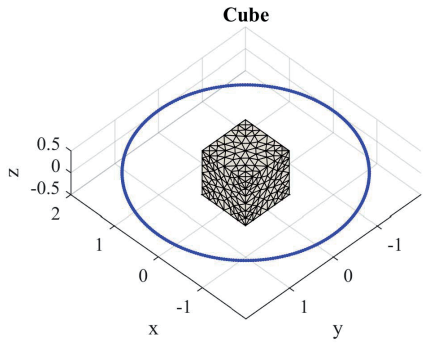



Figure 1: Geometry of the cube test case. Receivers are placed in a circle and marked with blue color. The plotted mesh has 624 triangular elements

The computation time which is needed for finding the solution in field points for the BEM and FMBEM is shown in Tab. 1 and plotted in Fig. 2.

| Test case | Mesh elements | Calc. time (s) FMBEM | BEM | Saved time (%) |
|---|---|---|---|---|
| Cube | 624 | 27.2 | 37.3 | 27.0 |
| | 744 | 35.1 | 47.1 | 25.6 |
| | 1096 | 51.2 | 82.5 | 38.0 |
| | 1480 | 74.5 | 130.0 | 42.7 |
| | 3124 | 143.9 | 473.3 | 69.6 |
| | 6260 | 308.4 | 1880.8 | 83.6 |
| | 10204 | 545.5 | 4625.2 | 88.2 |

Table 1: Number of mesh elements for the cube test case and corresponding calculation time for the BEM and FMBEM and time savings for the FMBEM

The ESIE divides the edges into integration points which are not uniformly distributed, like the nodes in the BEM mesh, but rather follow a Gauss-Legendre discretization scheme. That is why it is convenient to display the ESIE convergence in a different graph with the number of integration points on the $x$-axis. The calculation time for the ESIE is shown in Tab. 2 and plotted in Fig. 3.

The relative error convergence (relative error as a function of the calculation time) is displayed in Fig. 4 for the BEM and FMBEM and in Fig. 5 for the ESIE. The receiver R1 has coordinates [1.97, 0.36, 0] and the receiver R2 has coordinates [−0.65, −1.89, 0].

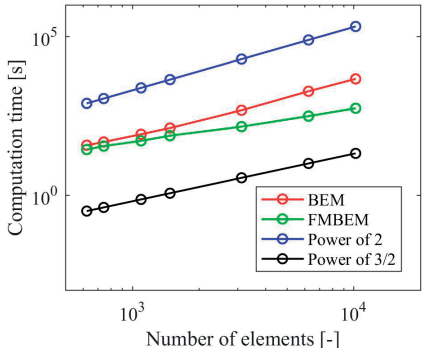



Figure 2: Computation time as a function of the number of elements for the BEM and FMBEM for the cube test case. The number of elements to the power of 2 and 3/2

| Test case | Edge points | Calc. time (s): ESIE |
|---|---|---|
| | 10 | 6.9 |
| | 20 | 10.3 |
| | 30 | 17.2 |
| Cube | 40 | 29.1 |
| | 50 | 58.4 |
| | 60 | 102.8 |
| | 70 | 163.6 |

Table 2: Number of edge points for the cube test case and corresponding calculation time

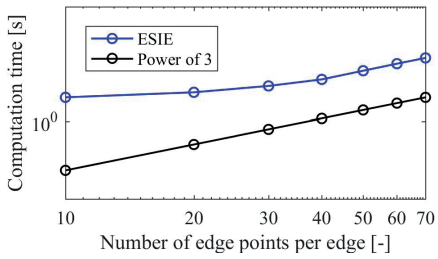



Figure 3: Computation time as a function of the number of edge points per edge for the ESIE for the cube test case. The number of edge points per edge to the power of 3

Fig. 4 compares the BEM and FMBEM relative error and shows that the BEM needs more time for converging with demanded precision. The ESIE lowest relative error in Fig. 5 is of order $10^{-5}$ while the BEM and FMBEM lowest relative error in Fig. 4 is of order $10^{-3}$. It is presented in graphs and tables that the ESIE converges significantly faster than the other methods. The last relative error value (for 10204 mesh elements or 70 edge integration points) is used as a reference result in Fig. 4 and Fig. 5 and it is not presented in graphs. It is worth to mention that an analytical solution of the Helmholtz equation for the cube case is not available.

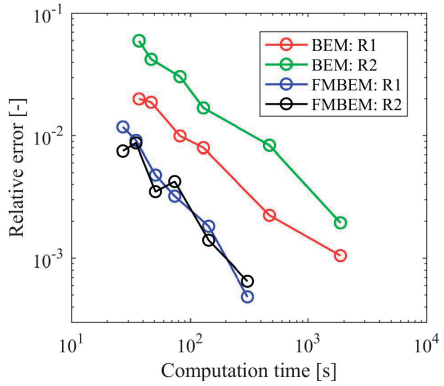



Figure 4: The relative error as a function of the computation time for the BEM and FMBEM and the cube test case. Reference result is the sound pressure calculated with the highest number of mesh elements

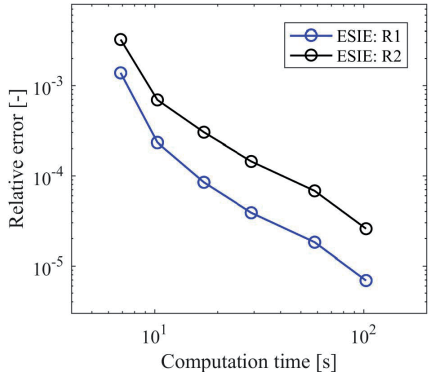



Figure 5: The relative error as a function of the computation time for the ESIE and the cube test case. Reference result is the sound pressure calculated with the highest number of edge points

The method comparison is presented in Fig. 6. The insertion loss is displayed as a function of the receiver num-

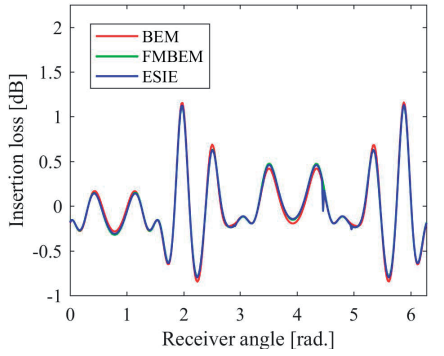



Figure 6: The method comparison: the insertion loss as a function of the receiver angle for the cube

ber (receivers are located along a circle with a step of 0.02 radians). The comparison is shown in Fig. 6 for the densest mesh and the highest number of edge integration points. Fig. 6 shows that there are minor differences in the insertion loss obtained by the used methods for a small scale test case.

### 3.2. Noise Barrier

The noise barrier is a very typical test case, even though with the high source frequency and high dimensions, it is time demanding to perform calculations for this test case.

The modelled noise barrier is shown in Fig. 7. Receivers are placed in positions $[X, 0, 0]$ where $X$ goes from 0.15 to 20.15 with a step of 0.1 (i.e. in front of the noise barrier). The sound source is placed in $[-5, 0, 0]$ (i.e. behind the noise barrier). Dimensions of the noise barrier are 0.25 m (width), 10.0 m (length) and 2.0 m (height above the ground). The modelled noise barrier height and the sound source power is intentionally doubled to simulate a rigid ground. The frequency of the source is 500 Hz and the sound speed is 344 m/s.

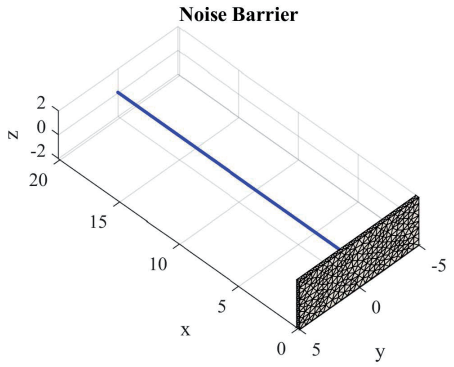



Figure 7: Geometry of the noise barrier test case. Receivers are placed along a line and marked with blue color. The noise barrier mesh has 1068 triangular elements

The number of elements in used meshes is shown in Tab. 3 together with the calculation time for the BEM and FMBEM. Time savings are increasing with the mesh size. Fig. 8 demonstrates that the computation time of the BEM is increasing with the power of 2 while the computation time of the FMBEM is rather increasing with the power of 3/2 of the element number.

| Test case | Mesh elements | Calc. time (s) FMBEM | Calc. time (s) BEM | Saved time [%] |
|---|---|---|---|---|
| Noise barrier | 1784 | 96.7 | 159.6 | 39.4 |
| | 2312 | 114.0 | 247.9 | 54.0 |
| | 3340 | 164.2 | 480.3 | 65.8 |
| | 4704 | 252.4 | 904.9 | 72.1 |
| | 7764 | 420.6 | 2367.3 | 82.2 |
| | 14428 | 1000.2 | 8336.5 | 88.0 |
| | 19092 | 1511.5 | 14083.7 | 89.3 |

Table 3: Number of mesh elements for the noise barrier test case and corresponding calculation time for the BEM and FMBEM and time savings for the FMBEM

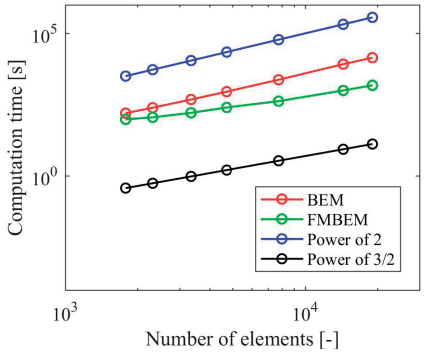



Figure 8: Computation time as a function of the number of elements for the BEM and FMBEM for the noise barrier. The number of elements to the power of 2 and 3/2

The ESIE calculation time is described in Tab. 4 together with the number of edge points. The computation time as a function of the number of edge points for the ESIE method is shown in Fig. 9. The calculation time for the ESIE and this test case is increasing with the 3rd power of the edge points number.

| Test case | Edge points | Calc. time (s): ESIE |
|---|---|---|
| Noise barrier | 20 | 13.1 |
| | 40 | 28.8 |
| | 60 | 97.5 |
| | 80 | 225.5 |
| | 100 | 428.0 |
| | 120 | 707.9 |
| | 140 | 1077.4 |

Table 4: Number of edge points for the noise barrier test case and corresponding calculation time

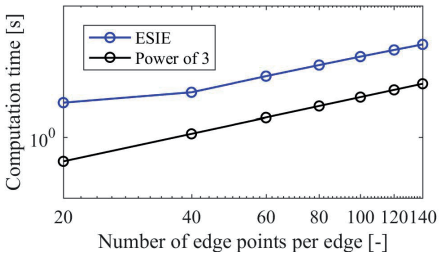



Figure 9: Computation time as a function of the number of edge points per edge for the ESIE and the noise barrier. The number of edge points per edge to the power of 3

The relative error is calculated with Eq. (13). The reference value is obtained with 19092 mesh elements for the

BEM and FMBEM and 140 edge integration points for the ESIE. The relative error convergence for the BEM and FMBEM is shown in Fig. 10 and the relative error for the ESIE is shown in Fig. 11. The receiver R1 has coordinates [10.05, 0, 0] and the receiver R2 has coordinates [20.05, 0, 0].

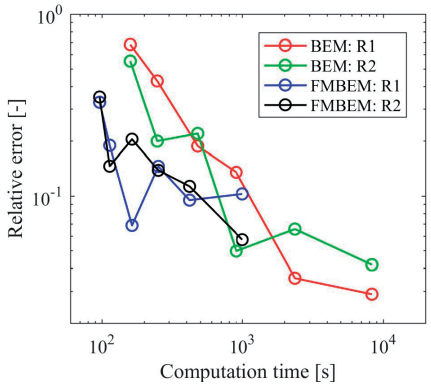



Figure 10: The relative error as a function of the computation time for the BEM and FMBEM and the noise barrier. Reference result is the sound pressure calculated with the highest number of mesh elements

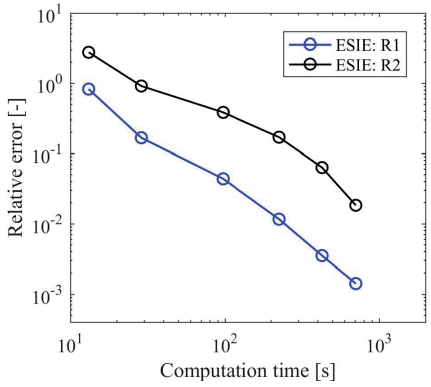



Figure 11: The relative error as a function of the computation time for the ESIE and the noise barrier. Reference result is the sound pressure calculated with the highest number of edge points

Fig. 10 shows excessively high calculation time for the BEM. The last relative error seems to be reasonable since it is lower than 0.05 for both BEM receivers. Convergence of the FMBEM is quite scattered as it is also shown later in the method comparison.

Both Fig. 5 and Fig. 11 show that the relative error convergence of the ESIE is straightforward. The BEM for a small test case converges also continuously even though geometrical singularities which increase the relative error can be found. The trickiest method is the FMBEM which converges dependently on the clustering process. The FMBEM also needs more input parameters, e.g. the number of iterations before truncating the infinite series (Eqs. (6) and (7)) and input parameters for the GMRes solver.

The method comparison is shown in Fig. 12 for 19092 mesh elements for the BEM and FMBEM or 140 edge integration points for the ESIE. Maximum difference between the ESIE and BEM is 0.86 dB and the maximum difference between the FMBEM and BEM is 4.55 dB. Maximum calculation time for the BEM is 3.9 hours compared to 25.2 minutes for the FMBEM and 18.0 minutes for the ESIE.

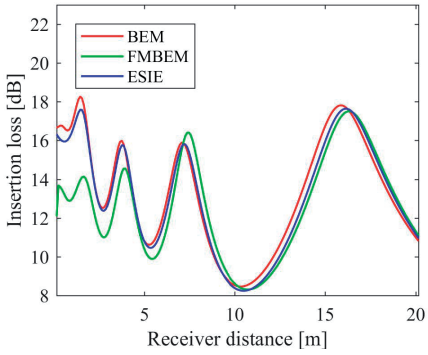



Figure 12: The method comparison: the insertion loss as a function of the receiver distance for the noise barrier

While the BEM and ESIE in closed proximity to the noise barrier work well, the insertion loss obtained by the FMBEM does not agree with other methods near to the noise barrier and rapidly decreases very close to the surface. Fig. 12 shows that the used FMBEM implementation lacks the near-singular integration. This behaviour occurs when the receiver is very close to the calculated element. The characteristic element length is 0.11 meters which means that 6.25 characteristic element lengths per wavelength are used.

## 4. Conclusions

Three numerical techniques have been demonstrated in the paper in order to assess the environmental noise attenuation. The first is an ordinary BEM which is very well tested and implemented as the open-source Matlab codes OpenBEM.

The second method is an experimental extension of the OpenBEM for the FMBEM. Setting of the FMBEM is very individual for each test case compared to the ordinary BEM. It is an important and not a trivial issue to carefully choose the right number of evaluation terms of an infinite sum (see Eqs. (6) and (7)) and also parameters of GMRes solver (i.e. the number of iterations before the restart, the maximum number of restarts and the calculation tolerance).

The last used method is the ESIE which is implemented as the “Edge diffraction toolbox”. The paper has shown

that it is the ESIE which has quite ambitious perspectives. The presented research has indicated that the calculation time necessary for the convergence of the ESIE is shorter than the time of the FMBEM and significantly shorter that the time of the BEM.

The ESIE or FMBEM does not have potential to replace the boundary element method. The reason is that the BEM claims to converge to the solution of the Helmholtz differential equation. The ESIE also shows problematic behaviour in several receiver positions as described in Ref. [19]) and has issues for solving non-convex bodies [8]. The FMBEM is using several algorithms which are significantly increasing the computational speed. On the other hand, these algorithms are by their definitions approximate and the FMBEM converges to the result only with certain accuracy.

## Acknowledgement

This work was carried out during the tenure of an ERCIM 'Alain Bensoussan' Fellowship Programme.

## References

[1] *Directive 2002/49/EC of the European Parliament and of the Council relating to the assessment a management of environmental noise*, 25. 6. 2002.

[2] Vos, P., Beuving, M., Verheijen, E.: *Harmonoise: Final Technical Report*, Revision Number: 04, 2005.

[3] Salomons, E., Heimann, D.: *Harmonoise: Reference Model. Description of the Reference model*, Revision number: 00.10, 2004.

[4] Kephalopoulos, S., Paviotti, M., Anfossolédée, F.: *Common Noise Assessment Methods in Europe (CNOSSOS-EU)*, Luxembourg, Publications Office of the European Union, 2012, ISBN 978-92-79-25281-5.

[5] Pigasse, G.: *Optimised Noise Barriers*, A State-of-the-Art Report, Vejdirektoratet, 2011, ISBN: 978-87-92094-77-3.

[6] Jacobsen, F., Poulsen, T., Rindel, J. H., Gade, A. C., Ohlrich, M.: *Fundamentals of Acoustics and Noise Control*, Lyngby, Department of Electrical Engineering, Technical University of Denmark, 2011.

[7] Juhl, M. P.: *The Boundary Element Method for Sound Field Calculations*, 1993, Lyngby, Dissertation thesis, Technical University of Denmark, The Acoustics Laboratory.

[8] Summers, J. E.: *Inaccuracy in the treatment of multiple-order diffraction by secondary-edge-source methods*, J. Acoust. Soc. Am. **133**, 3673–3676 (2013).

[9] Williams, E. G.: *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, San Diego, Academic Press, 1999. ISBN 978-01-27539-60-7.

[10] Henríquez, V. C., Juhl, P. M.: *OpenBEM – An open source Boundary Element Method software in Acoustics*, INTER-NOISE 2010, Lisbon, 15.–16. 6. (2010).

[11] Henríquez, V. C., Juhl, P. M.: *OpenBEM – "Open source Matlab codes for the Boundary Element Method"*, The Maersk Mc-Kinney Møller Institute, University of Southern Denmark, `http://www.openbem.dk/`.

[12] Geuzanine, C., Remacle, J. F.: *Gmsh: a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities*, International Journal for Numerical Methods in Engineering 79(**11**), 1309–1331 (2009).

[13] Olsen, E. S., Cutanda, V., Gramtorp, J., Eriksen, A.: *Calculating the Sound Field in an Acoustic Intensity Probe Calibrator – A Practical Utilization of Boundary Element Modelling*, Proceedings of Eighth International Congress on Sound and Vibration, Hong Kong, China, 2313–2321 (2001).

[14] Rokhlin, V.: *Rapid Solution of Integral Equations of Classic Potential Theory*, Journal of Computational Physics, Vol. **60**, 187–207 (1985).

[15] Ellegaard, S. G.: *Implementation of the Fast Multipole Boundary Element Method (FMBEM) for sound field calculations*, Odense, Diploma thesis, University of Southern Denmark, Faculty of Engineering.

[16] Abramowitz, M., Stegun, I.: *Handbook of Mathematical Functions*, Dover, New York, 1974. ISBN 0-486-61272-4.

[17] Svensson, U. P.: *Edge diffraction toolbox for Matlab*, Acoustics Research Centre, Norwegian University of Science and Technology, `http://www.iet.ntnu.no/~svensson/software/`.

[18] Schenck, H. A.: *Improved integral formulation for acoustic radiation problems*, J. Acoust. Soc. Am. **44**, 41–58 (1968).

[19] Svensson, U. P., Haike, B., Forssén, J.: *Benchmark cases in 3D diffraction with different methods*, Forum Acusticum, 7.–12. 9. 2014, Krakow.

[20] Asheim, A., Svensson, U. P.: *An integral equation formulation for the diffraction from convex plates and polyhedral*, J. Acoust. Soc. Am. **133**, 3681–3691 (2013).

# Alofonická variabilita v češtině z pohledu řečové syntézy

## Allophonic Variability in Czech from the Perspective of Speech Synthesis

Radek Skarnitzl

Univerzita Karlova, Filozofická fakulta, Fonetický ústav – nám. Jana Palacha 2, 116 38 Praha 1

This study examines the allophonic variability of the consonant system of Czech from the perspective of concatenative speech synthesis. Systematic variants exist for several Czech phonemes; these concern the place of articulation (such as the alveolar and velar variant of /n/), voicing (the voiced and voiceless variant of the fricative trill /ř/), or the syllabicity of /r/ and /l/. So far, these variants were identified in the inventory of the ARTIC synthesis system as separate units, but pooled before unit selection. In this study, an incorrect variant was force-synthesized into target words (e.g., an alveolar [n] into the word *banka* instead of the velar [ŋ], a voiceless /ř/ in *řeka*). The resulting synthetic phrases were auditorily analyzed to see whether this process would yield intrusive artefacts. Recommendations were formulated and already implemented in the ARTIC system: in the end, most of the systematic variants had to be split.

## 1. Úvod

Ačkoli během posledních přibližně deseti let narůstá obliba řečové syntézy založené na skrytých Markovových modelech (HMM) [1] nebo hlubokých neuronových sítích (DNN) [2], případně hybridních systémů [3], reálným aplikacím prozatím stále dominují systémy konkatenační řečové syntézy založené na dynamickém výběru jednotek (*dynamic unit selection*) [4]. Je dobře známé, že konkatenační syntéza je preferována především díky celkově vyšší přirozenosti [5]; stejně tak je však známý jejich hlavní problém, jímž je výskyt nepředvídatelných slyšitelných artefaktů ve výsledné řeči.

Konkatenační syntéza řeči využívá řetězení (konkatenace) řečových jednotek nacházejících se v rozsáhlém korpusu, z něhož jsou pro konkrétní syntetizovaný text vybírány. Základními jednotkami jsou většinou difony a ze sady potenciálních difonů je vybrán takový kandidát, aby byly minimalizovány dvě hodnoty nazvané *cena cíle* a *cena řetězení* [4]. *Cena cíle* (*target cost*) odpovídá rozdílu mezi vlastnostmi kandidáta a ideálního cíle; každá jednotka je popsána pomocí příznaků, které vyjadřují například její poziční vlastnosti. Oproti tomu *cena řetězení* (*concatenation cost* nebo *join cost*) se týká hladkosti řetězení dvou na sebe navazujících difonů; vychází se tedy z rozdílu jejich akustických vlastností.

Současné výzkumy v českém i světovém měřítku se převážně věnují hledání efektivnějších způsobů navazování jednotek, aby se pokud možno zabránilo výskytu výše zmíněných artefaktů, které kvalitu výsledné řečové syntézy snižují. Snahy jsou tedy upírány na přesnější vyjádření *ceny řetězení*. Jak ukázal Matoušek a kolegové [6], artefakty mohou vznikat kvůli chybnému označení řečových jednotek v databázi; kvůli tomu, že zmíněné ceny zcela neodpovídají lidské percepci; a proto, že algoritmy preferují nižší celkovou cenu, což však může vést k lokálně vyšším cenám v konkrétních konkatenačních bodech. Mnoho výzkumů ukazuje, že nejrušivější artefakty jsou způsobeny nespojitostí v základní hlasové frekvenci (F0) [7] a ve spektrální oblasti [8].

V této studii se od téměř výhradního zaměření na *cenu řetězení* odchylujeme a věnujeme se *ceně cíle*, nikoli však z hlediska pozičních či jiných příznaků, ale z hlediska samotného inventáře řečových jednotek. V řečové syntéze nebývá důvod inventář jednotek zpochybňovat: jednotky se stanoví na základě převážně fonologických vlastností daného jazyka a dále se s nimi v řečové syntéze pracuje. Detailní analýza řeči syntetizované pomocí českého systému ARTIC [9] však ukázala, že by přehodnocení inventáře užívaných jednotek, založené na fonetické analýze reálných syntetizovaných promluv, mohlo vést k efektivnějšímu výběru jednotek z databáze a ke kvalitativně lepší syntetizované řeči. Navržené a implementované změny se týkají výhradně konsonantického systému; v dalším popisu se proto věnujeme pouze souhláskám.

## 2. Konsonantický systém češtiny

Základním konstruktem pro popis zvukové stránky jazyka je *foném*, nejmenší jednotka, která má v daném jazyce distinktivní platnost (tj. může měnit významy slov). Souhlásky /t/ a /k/ jsou tedy bezpochyby fonémy, protože jejich záměna vede ke změně významu (např. *trám – krám*), a přepisujeme je v lomených závorkách. Inventář českých konsonantů představuje tabulka 1. Pro dnešní češtinu uvažujeme o 26 souhláskových fonémech; ty jsou v tabulce vyznačeny černě. Konsonanty tradičně dělíme podle *způsobu artikulace* (v tabulce jako jednotlivé řady), *místa artikulace* (sloupce) a podle *znělosti*; v rámci jednotlivých buněk jsou neznělé hlásky (nedoprovázené kmitáním hlasivek) vyznačeny nalevo, zatímco hlásky znělé (při jejichž produkci hlasivky kmitají) jsou vyznačeny napravo. Je patrné, že znělost je distinktivní pouze u prvních čtyř způsobů artikulace: tyto hlásky souhrnně nazýváme *obstruenty*. Hlásky ve spodních čtyřech řadách se nazývají

Přijato 23. května 2017, akceptováno 18. dubna 2018.

| | bilabiální | labiodent. | alveolární | postalveol. | palatální | velární | laryngální |
|---|---|---|---|---|---|---|---|
| explozivy | p b | | t d | | ť ď | k g | ʔ |
| frikativy | | f v | s z | š ž | | x ɣ | h |
| afrikáty | | | t͡s d͡z | t͡š d͡ž | | | |
| frikativní vibranty | | | | r̝̊ ř | | | |
| nazály | m | ɱ | n | | ň | ŋ | |
| aproximativní vibranta | | | r | | | | |
| laterální aproximanta | | | l | | | | |
| aproximanta | | | | | j | | |

Tabulka 1: Inventář českých konsonantů

*sonory* a v základní podobě jsou pouze znělé. Detailnější popis českých konsonantů poskytuje například [10].

Pro tuto studii je klíčové, že fonémy se v souvislé řeči mohou systematicky měnit; takové systematické varianty fonémů nazýváme *alofony* a v tabulce 1 jsou vyznačeny šedou barvou. Jedná se o varianty poziční, protože se vyskytují v konkrétních pozicích, v konkrétních hláskových kontextech. Klasickým příkladem jsou varianty fonému /n/: ve slově *ven* vyslovíme jeho základní, alveolární variantu [n], avšak ve slově *venku* dojde vlivem následující hlásky, velárního /k/, k asimilaci místa artikulace a vyslovíme velární [ŋ]: [veŋku]. Všimněme si, že z hlediska fonetického je vztah mezi /t/ a /k/ na jedné straně a [n] a [ŋ] na straně druhé shodný: jedná se o dvojice hlásek, které odlišuje pouze místo artikulace (alveolární *vs.* velární). Z hlediska fonologického je však situace odlišná: [ŋ] je variantou fonému /n/, která se vyskytuje pouze v kontextu před velárními explozivami.

Rozdíl mezi [n] a [ŋ] spočíval v místě artikulace. Podobný princip můžeme uplatnit i v dalším případě: foném /m/, bilabiální nazála, má základní variantu [m], ale v kontextech před labiodentálními frikativami se vlivem asimilace často mění na nazálu labiodentální, [ɱ], například ve slovech *nymfa* [nɪɱfa] nebo *tramvaj* [traɱvaj].

Zdrojem alofonické variability může dále být v češtině znělost. Jak jsme zmínili, obstruenty obvykle vystupují ve dvojicích, které se odlišují právě znělostí. V některých případech se však nejedná o „plnohodnotné“, fonémické protiklady, ale o varianty jednoho fonému. Z tabulky 1 je patrné, že se to týká tří hlásek – [ɣ], [d͡z] a [r̝̊]. První dvě vznikají v kontextech regresivní asimilace znělosti, tedy před znělým obstruentem, jako znělé varianty neznělého /x/ (*Mach dal* [maɣ dal]) a /t͡s/ (*moc dlouho* [mod͡z dlouho]). Spojení *Mach dal* lze rovněž vyslovit s laryngální frikativou: [mah dal]. V případě českého /ř/ je naopak základní variantou znělé [ř], které se v asimilačních kontextech mění na variantu neznělou, [r̝̊]: *keř* [keř], *tři* [tr̝̊ɪ], *nářky* [naːřkɪ]. Poznamenejme, že některé zdroje české /ř/ definují jako hlásku alveolární, v tabulce je uvedeno jako postalveolární; otázka přesného místa artikulace této pro češtinu specifické hlásky doposud nebyla uspokojivě zodpovězena.

Kromě místa artikulace a znělosti je dalším jevem, který rozlišuje systematické varianty českých souhlásek, slabičnost či slabikotvornost u /l/ a /r/. Ve slově *lněný* tak vyslovíme běžné [l], [lňeniː], zatímco ve slově *vlněný* bude /l/ vrcholem slabiky, [vl̩ňeniː]; podobně i *rty* [rtɪ] a *vrty* [vr̩tɪ]. Slabičné varianty /l/ a /r/ nejsou uvedeny v tabulce 1, protože o nich běžně neuvažujeme jako o alofonických variantách, nicméně pro účely této studie se jedná o kontrast významný. Se slabičností se někdy setkáme i při ortoepické výslovnosti číslovek *sedm* a *osm* ([sedm̩], [osm̩]) a jim příbuzných slov jako *sedmdesát* či *osmnáct*.

Pro úplnost zmiňme poslední hlásku, která v češtině nemá fonémický status a která je v tabulce 1 definována jako neznělá laryngální exploziva, [ʔ]; jedná se o tzv. *ráz* vyskytující se před samohláskami na začátku slova (např. ve spojení *stál u okna* [staːl ʔu ʔokna]). Ačkoli ráz nepatří mezi cílové hlásky přímo analyzované v tomto výzkumu, bude v něm hrát důležitou roli.

## 3. Systém ARTIC a jeho konsonantický inventář

Systém řečové syntézy češtiny ARTIC (Artificial Talker in Czech) je vyvíjen na Západočeské univerzitě v Plzni [9]. Pro tuto studii byly využity čtyři hlasy, označované podle křestního jména mluvčích: dva ženské (Iva, Kateřina) a dva mužské (Jan, Stanislav). Základní informace o databázi řečových jednotek našich čtyř mluvčích shrnuje tabulka 2.

Systém ARTIC ve své dosavadní verzi pracuje s kompletním konsonantickým inventářem, jak byl popsán v předcházejícím oddílu. Jedná se o všech 32 hlásek uvedených v tabulce 1 spolu se třemi slabičnými konsonanty. Z našeho pohledu systém na úrovni databáze jednotek disponoval maximální možnou mírou informace. Během procesu syntézy, konkrétně během samotného výběru jednotek, však doposud byly jednotlivé alofonické varianty shlu-

| mluvčí | vět | trvání | průměrné tempo |
|---|---|---|---|
| Jan | 12 277 | 17:39 | 13,6 hl/s |
| Stanislav | 12 306 | 15:40 | 12,3 hl/s |
| Iva | 12 151 | 15:04 | 12,2 hl/s |
| Kateřina | 12 707 | 12:53 | 10,5 hl/s |

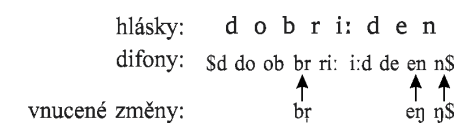
Tabulka 2: Základní údaje o databázi čtyř zdrojových mluvčích systému ARTIC

kovány. To znamená, že se ve slovním spojení *dobrý obr* teoreticky mohl vyskytnout tentýž difon <br>; podobně i ve spojeních *já bych byl* a *já bych pil* by se mohl vyskytnout stejný difon <ɪx>, ve slovech *příchod* a *bříza* stejný difon <řiː> atd.

Je zřejmé, že při vytváření systému konkatenační syntézy řeči je třeba brát v potaz kompromis mezi fonetickou přesností popisu a reálnými možnostmi hláskového inventáře. Prosté rozdělení popsaných alofonických variant i pro výběr jednotek nemusí být optimálním řešením. Cílem této studie proto bylo na základě analýzy zmíněných čtyř hlasů užívaných v řečové syntéze ARTIC určit, zda shlukování alofonických variant (resp. difonů z těchto variant odvozených) způsobuje slyšitelné a rušivé artefakty ve výsledné syntetizované řeči. Pokud by záměna některé z variant častěji vedla k přítomnosti percepčně rušivých jevů, bylo by vhodné je udržet oddělené i během výběru jednotek pro syntézu.

## 4. Metoda

Aby bylo možné ověřit percepční rušivost umístění nesprávné alofonické varianty, bylo třeba pro účely této studie vysyntetizovat v systému ARTIC speciálně sestavené věty a do cílových kontextů vnutit nesprávnou variantu. Např. pro vytvoření spojení *dobrý den* by byla nejprve provedena syntéza pomocí dosavadního systému a v druhém kroku by do konkrétních kontextů byly vnuceny nesprávné varianty, jak ukazuje obrázek 1; do slova *dobrý* by tak například byl vložen difon <br̥> ze slova *obr* a do slova *den* difon [eŋ] ze slova *tenký*.




Obrázek 1: Ukázka vnucení slabičného [r̥] do neslabičného kontextu a velárního [ŋ] do alveolárního kontextu ve spojení *dobrý den*

Takto bylo pro každou cílovou dvojici alofonických variant vysyntetizováno několik vět; příklady ukazuje tabulka 3. Při sestavování vět a slovních spojení jsme se snažili o reprezentativnost z hlediska segmentálního okolí cílových alofonů. Například znělá varianta /x/ se tak kromě slov *líh* a *hroch* objevila ve slovech *abych*, *nech*, *Bach* a *puch*, tedy v sousedství všech vokalických kvalit češtiny. Podobně slabičné [r̥] se vyskytovalo i po dalších konsonantech, než jsou ty uvedené v tabulce 3 (např. ve slovech *kapr*, *bagr*, *brzdy*, *frťan* či *zrcadlo*).

Jak již bylo zmíněno, cílem samotné percepční analýzy bylo zjistit, jestli popsaným způsobem vnucení nesprávné alofonické varianty – k čemuž při jejich shluknutí před výběrem jednotek před samotným syntetizováním výstupu může dojít – budou vznikat rušivé artefakty. Percepční analýzu provedli nezávisle na sobě dva fonetici; jedním z nich je autor této studie. Vzhledem k tomu, že se v případě vzniku artefaktů jedná o zcela zřejmé vady výsledné řeči (viz výsledky v následujícím oddíle), ověření větším počtem posluchačů jsme nepovažovali za účelné. Zároveň nás zajímalo, jestli se případné artefakty budou vyskytovat systematicky u všech čtyř mluvčích, nebo jestli se bude jednat o jevy náhodné.

| cílový jev | příklady syntetizovaných spojení |
|---|---|
| slabičnost R | Pet**r**ovi svet**r** sluší. |
| | Ud**r**žel pud**r**u h**r**st. |
| | t**r**sy t**r**ávy |
| | ch**r**t ch**r**ápe |
| slabičnost L | Zmok**l** na k**l**adině. |
| | Uhád**l** ned**l**ouhou hádanku. |
| | v**l**ci v**l**áčí |
| | s**l**zy s**l**abosti |
| alveolární a velární ŋ | Sr**n**ka ví**n**ko vypije. |
| | Sr**n**a pá**n**ovi uteče. |
| | Vl**n**ka ve**n**ku poteče. |
| | La**n**em vl**n**y nesvážeš. |
| znělé varianty x | Lí**h** by shořel snadno. |
| | Lí**h** snadno hoří. |
| | Hro**ch** běží rychle. |
| | Hro**ch** utíká rychle. |
| znělostní varianty Ř | ho**ř**ící ke**ř** |
| | zu**ř**ivá bou**ř**ka |
| | d**ř**evěné k**ř**ídlo |
| | úho**ř** d**ř**ímá |

Tabulka 3: Příklady syntetizovaných spojení

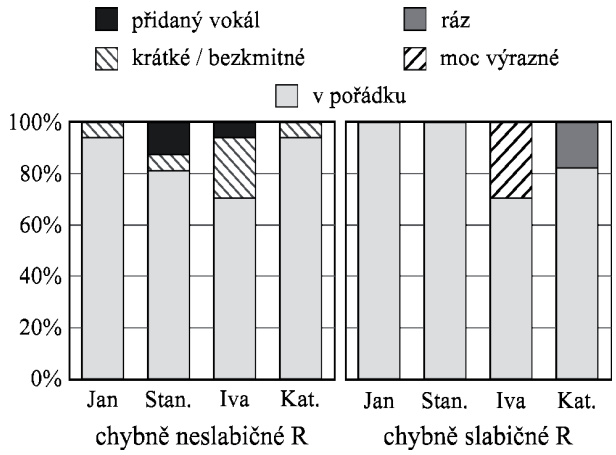
## 5. Výsledky

Výsledky poslechové analýzy budou prezentovány zvlášť pro jednotlivé alofonické varianty, ačkoli se některé problematické aspekty mohou opakovat.

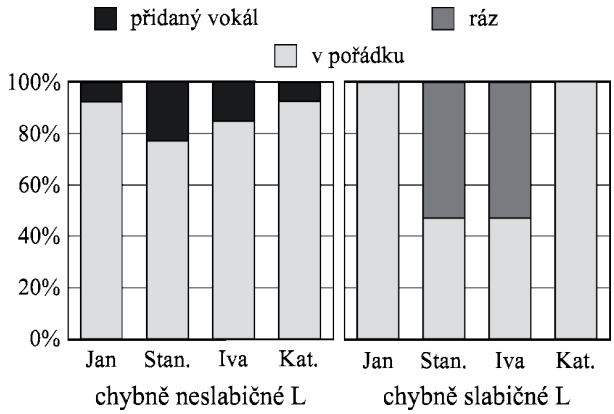
### 5.1. Slabičnost /r/

Pro rozdíl mezi běžným [r] a slabičným [r̥] bylo vytvořeno 17 spojení, přičemž každé z nich kanonicky obsahovalo minimálně dva zástupce /r/ (viz příklady v tabulce 3). Již u tohoto prvního jevu bylo zřejmé, že se syntetizovaná spojení s vnucenou chybnou variantou u čtyř zkoumaných mluvčích neprojevují stejně: výsledky ukazuje obrázek 2.

Vnucení neslabičného [r] do slabičného kontextu (např. do slova *srnka*) v 85 % případů nevedlo ke slyšitelnému problému. V 10 % případů byl výsledný syntetizovaný konsonant velmi krátký, příp. [r] dokonce nemělo vlivem konkatenace dvou difonů žádný kmit. Někdy byl konsonant tak krátký, že by takto v podstatě ani nebylo možné jej vyslovit. Slabičné [r̥] má v běžné řeči kmit vždy a jeho absence působí rušivým dojmem. Jak naznačuje obrázek 2,




Obrázek 2: Zastoupení správných a chybných realizací po vnucení neslabičného a neslabičného /r/ do opačného kontextu u čtyř mluvčích




Obrázek 3: Zastoupení správných a chybných realizací po vnucení neslabičného a neslabičného /l/ do opačného kontextu u čtyř mluvčích

krátká či bezkmitná varianta se alespoň jednou vyskytla u všech čtyř mluvčích. Méně častá chybná varianta je do jisté míry podobná: krátká či bezkmitná povaha výsledného konsonantu může budit dojem, že se jedná o [r] s vokálem (slovo *chrt* u mluvčí Iva znělo jako [xrut]) nebo o samotný vokál (slovo *drzý* u mluvčího Stanislav znělo jako [duzi:]).

Opačná situace, kdy do neslabičného kontextu byla vnucena slabičná varianta, způsobila slyšitelné artefakty jen u dvou mluvčích. U mluvčí Iva se jednalo o problém menší, který nepůsobí výrazně rušivě: u 30 % jejích položek mělo výsledné [r] (konkrétně jeho vokalický prvek) příliš silnou amplitudu. Z percepčního hlediska daleko závažnější jsou položky mluvčí Kateřina, u nichž došlo k chybnému vložení rázu; to se objevilo například ve slovech *Petrovi* [petr̩ʔovɪ] či *gril* [gr̩ʔɪl].

### 5.2. Slabičnost /l/

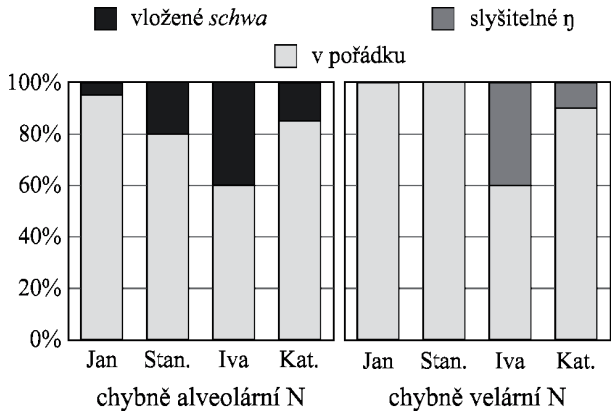
Pro rozdíl mezi běžným a slabičným [l̩] bylo vytvořeno celkem 13 spojení, přičemž každé z nich kanonicky obsahovalo jedno běžné [l] a jedno slabičné [l̩] (viz příklady v tabulce 3). Výsledky poslechové analýzy ukazuje obrázek 3; jak je z obrázku patrné, jsou problematické aspekty spojené s vnucením špatné varianty podobné jako v případě /r/.

Při vnucení neslabičného [l] do slabičného kontextu vznikl ve více než čtvrtině případů dojem spojení [l] a vokálu, případně samotného vokálu; z obrázku 3 je patrné, že se tato chyba objevuje u všech mluvčích, ačkoli v odlišné míře. U mluvčího Jan tak slovo *slzy* zní jako [suzɪ], u obou ženských mluvčích zní slovo *slízl* jako [sli:zlo], u Stanislava zní slovo *sezobl* jako [sezoblə], a slovo *odmítl* má dokonce dvě vkladná schwa [ʔodmi:tələ].

Při vnucení slabičného [l̩] do neslabičného kontextu dochází u dvou mluvčích k chybnému vložení rázu, podobně jako v případě slabičného [r̩], a to dokonce u nadpoloviního počtu položek. Jako příklady uveďme slova *kladině* [kl̩ʔaďiňe] či *potlach* [potl̩ʔax], která byla s rázem syntetizována u obou zmíněných mluvčích.

### 5.3. Alveolární a velární varianta /n/

Pro rozdíl mezi základní alveolární variantou [n] a asimilací místa artikulace vzniklým velárním [ŋ] bylo vytvořeno celkem 10 spojení, přičemž každé z nich obsahovalo dvě položky /n/ (srov. tabulku 3). Jak ukazuje obrázek 4, slyšitelné artefakty vznikají v případě vynucení obou chybných variant. Při vynucení alveolárního [n] do kontextu, kde v češtině vlivem asimilace vyslovujeme velární [ŋ], se ve 20 % případů vyskytuje výrazné vkladné (epentetické) *schwa*. Rušivost epentetického *schwa* stoupá zejména tehdy, pokud jeho přítomnost vyvolává dojem přidané slabiky [11]. Krátký vokalický prvek např. ve slově *mamka* [mamᵊka] je tedy relativně přirozeným důsledkem přechodu z jedné hlásky na hlásku následující, avšak tříslabičné slovo [mamәka] již přirozeně nepůsobí. Ve zde zmíně-




Obrázek 4: Zastoupení správných a chybných realizací po vnucení alveolárního [n] a velárního [ŋ] do opačného kontextu u čtyř mluvčích

ných příkladech působí vkladné *schwa* spíše rušivě, např. ve slovech *Manka* [manᵊka] či *vlnka* [vl̩nᵊka] u mluvčího Stanislav.

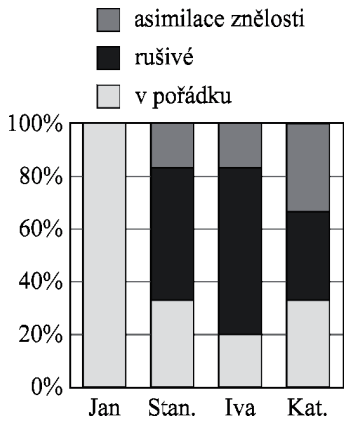
Vnucení velární nazály do alveolárního kontextu způsobilo slyšitelné změny pouze u dvou mluvčích, u Ivy však ve 40 % všech položek. V těchto slovech je pak skutečně zřetelně slyšet velární hláska, tedy například *vlny* jako [vl̩ŋɪ], *poleno* jako [poleŋo] a podobně.

### 5.4. Neznělé a znělé varianty /x/

V tomto oddílu zmíníme záměnu neznělého [x] za znělou velární frikativu [ɣ] i za znělou laryngální frikativu [h] (viz oddíl 2) a opačně. Pro tento kontext bylo vytvořeno 12 spojení, vždy s jedním /x/, jak ukazují příklady v tabulce 3.

Důležité je, že záměna [x] a [ɣ] nevedla ani v jednom případě ke vzniku slyšitelného artefaktu, většinou proto, že ve skutečnosti výsledná syntetizovaná hláska nebyla stoprocentně znělá ani neznělá.

Jako zajímavější se ukazuje vnucení znělého [h] do kontextu neznělého [x]; tato změna vedla ke slyšitelným a rušivým jevům celkově přesně u poloviny položek. Jak je však patrné z obrázku 5, žádný z nich se neobjevil u mluvčího Jan; naopak u ostatních tří mluvčích byla syntetizována slyšitelně chybná varianta ve více než 60 % jejich položek. Rušivost skutečně znělého [h] ve spojení jako *nech Petra* [neh petra] nebo *Bach se hraje* [bah se hraje] jistě není překvapivá. Obrázek 5 ukazuje, že u další skupiny případů způsobil vynucený výběr znělého [h] ve výsledné syntetizované řeči asimilaci znělosti. Spojení *líh snadno* tak zní [li:h znadno].



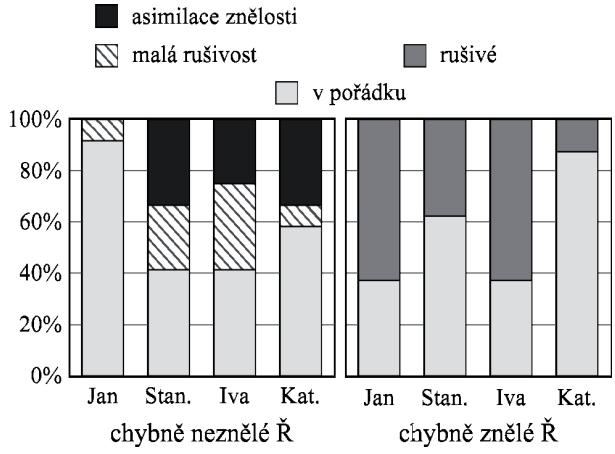

Obrázek 5: Zastoupení správných a chybných realizací po vnucení znělého laryngálního [h] do kontextu neznělého velárního [x] u čtyř mluvčích

### 5.5. Varianty /ř/

Posledním jevem, na který se tato studie zaměřuje, je znělost u alofonických variant /ř/. Pro tento jev bylo vytvořeno 10 spojení, přičemž každé z nich obsahovalo dvě hlásky (viz příklady v tabulce 3). Výsledky poslechové analýzy ukazuje obrázek 6.

Již na první pohled je zřejmé, že vynucení nesprávné varianty /ř/ vedlo k vysokému počtu slyšitelných artefaktů. Co se týče umístění neznělého [r̝̊] do znělého kontextu, v 25 % případů byl neznělý charakter /ř/ slyšitelný (např. ve slově *dveře* [dveř̊e]), ale tato záměna zároveň nebyla výrazně rušivá. Vzhledem k tomu, že [r̝] je hláskou, která i ve znělém kontextu nejčastěji ztrácí znělost [12: kapitola 9], je nízká rušivost neznělé varianty očekávatelná. Ve 23 % položek došlo k závažnější záměně, která zahrnovala asimilaci znělosti: vynucení neznělé varianty ve výsledku způsobilo chybnou regresivní asimilaci, takže např. první slabika ve slově *dřevěné* byla u tří ze čtyř mluvčích *tře-* [tr̝̊evjene:]. Podobný výsledek byl zaznamenán například u slov *zadřená*, které zní jako *zatřená*, či *dřímá*, kde záměna dokonce vede ke změně významu slova na *třímá*.

Umístění znělé varianty do neznělého kontextu bylo výrazně rušivé v 43 % všech položek: zde většinou nedošlo současně k asimilační změně (tedy z *křídla* se nestalo [gři:dlo]), ale přítomnost skutečně znělého [r̝] (tedy výslovnost [kři:dlo], která se vyskytla u všech čtyř mluvčích) rušivě působila. Častá chybná záměna byla i ve slovech *skřípou* [skři:pou] či *tříska* [tři:ska].




Obrázek 6: Zastoupení správných a chybných realizací po vnucení neznělého a znělého /ř/ do opačného kontextu u čtyř mluvčích

## 6. Diskuse a závěr

Cílem této studie bylo zjistit, jestli systematické varianty českých fonémů mohou být nadále shlukovány před výběrem jednotky pro konkatenační syntézu, nebo jestli je vhodné některé z nich držet i pro finální krok syntézy pomocí dynamického výběru jednotek zvlášť. Protože oddělení jednotek (resp. odpovídajících difonů) může výrazně omezit inventář, z něhož je daný difon možné vybírat, nejedná se o rozhodnutí triviální a bylo nutné jej důkladně podložit.

Rozhodnutí se zdá být jednoznačné v případě slabičné a neslabičné varianty /r/ a /l/. Důvodem chyb je zřejmě nedokonalá anotace řečového korpusu, která byla v úvodu zmíněna jako jeden z hlavních faktorů přítomnosti artefaktů v syntéze řeči [6]: informace o rázu není optimální. Pokud tedy vnutíme do slova *Petrovi* slabičné [r̩], použije se například difon [r̩o] ze spojení *Petr odpověděl*, které však obsahuje ráz: [r̩ʔo]. Domníváme se, že pokud je možné, že bude vložen ráz do kontextu, do nějž vůbec nepatří, je vhodné slabičné a neslabičné varianty /r/ a /l/ oddělit i pro samotnou syntézu.

Hlavním problémem u záměny alveolárního a velárního alofonu /n/ bylo vkladné (epentetické) *schwa* (např. [lenᵊka]). Pokud vložením *schwa* vzniká nová, dodatečná slabika, jedná se o jev rušivý, ale k tomu zdaleka nedošlo ve všech případech. Naše analýza proto nevedla k jednoznačnému závěru: pokud by rozdělení obou variant pro syntézu vedlo k přílišnému omezení inventáře, je možné obě varianty (resp. jim odpovídající difony) ponechat shluknuté.

Oddíl 5.4 se věnoval znělým variantám velární frikativy /x/. Za základní znělou variantu neznělého /x/ považujeme velární [ɣ] a při záměně těchto dvou hlásek se žádné artefakty nevyskytly; obr. 5 se týkal pouze vynucené záměny [x] a [h], k níž by však v reálném kontextu syntézy pomocí dynamického výběru jednotek nejspíš nedošlo. Bylo tedy možné uzavřít, že alofonické varianty /x/ mohou být ponechány shluknuté; nakonec byl systém ARTIC modifikován tak, že znělé velární [ɣ] je mapováno na laryngální [h].

Záměna znělostních variant /ř/ vedla k relativně velkému počtu rušivých položek; jejich oddělení pro fázi výběru jednotek se proto ukázalo jako jednoznačné doporučení.

Všechna doporučení založená na tomto výzkumu byla přijata a již implementována do všech hlasů v systému ARTIC, ačkoli z prezentovaných výsledků vyplývá, že slyšitelné artefakty nevznikají u všech mluvčích stejnou měrou. Zároveň však výsledky nenaznačily jednoznačné idiosynkratické tendence – nelze tedy říct, že by například mužské hlasy měly k chybám větší sklon než ženské. Jedinou výjimkou je mluvčí Jan, u nějž vzniklo rušivých chyb nejméně.

## Poděkování

Tento výzkum byl podpořen projektem Grantové agentury ČR, číslo GAČR 16-04420S.

## Reference

[1] Yoshimura, T., Tokuda, K., Masuko, T., Kobayashi, T., Kitamura, T.: Simultaneous modeling of spectrum, pitch and duration in HMM-based speech synthesis, *Proc. Eurospeech 1999*, p. 2347–2350, 1999.

[2] Zen, H., Senior, A., Schuster, M.: Statistical parametric speech synthesis using deep neural networks, *Proc. ICASSP 2013*, p. 7962–7966, 2013.

[3] Tiomkin, S., Malah, D., Shechtman, S., Kons, Z.: A hybrid text-to-speech system that combines concatenative and statistical synthesis units, *IEEE Transactions on Audio, Speech, and Language Processing*, 19(**5**), p. 1278–1288, 2011.

[4] Dutoit, T.: Corpus-based speech synthesis, in Benesty, J., Sondhi, M., Huang Y. (Eds.), *Springer Handbook of Speech Processing*, p. 437–455. Springer, Dordrecht, 2008.

[5] King, S.: Measuring a decade of progress in Text-to-Speech, *Loquens*, 1(**1**), 2014.

[6] Matoušek, J., Tihelka, D., Legát, M.: Is unit selection aware of audible artifacts? *Proc. 8th ISCA Speech Synthesis Workshop 2013*, p. 267–271.

[7] Legát, M., Matoušek, J.: Pitch contours as predictors of audible concatenation artifacts, *Proc. World Congress on Engineering and Computer Science 2011*, p. 525–529, 2011.

[8] Klabber, E., Veldhuis, R.: Reducing audible spectral discontinuities, *IEEE Transactions on Speech and Audio Processing*, 9(**1**), p. 39–51, 2001.

[9] Matoušek, J. Tihelka, D., Romportl, J.: Current state of Czech text-to-speech system ARTIC, *Proc. of the 9th International Conference TSD 2006, Lecture Notes in Artificial Intelligence, vol. 4188*, p. 439–446. Springer, Berlin/Heidelberg, 2006.

[10] Skarnitzl, R., Šturm, P., Volín, J.: *Zvuková báze řečové komunikace: Fonetický a fonologický popis řeči*, Karolinum, Praha, 2016.

[11] Matoušek, J., Skarnitzl, R., Machač, P., Trmal, J.: Identification and automatic detection of parasitic speech sounds. *Proc. Interspeech 2009*, p. 876–879, 2009.

[12] Skarnitzl, R.: *Znělostní kontrast nejen v češtině*, Nakladatelství Epocha, Praha, 2011.

Akustické listy: ročník 24, číslo 1–2 červen 2018 ISSN: 1212-4702
Vydavatel: Česká akustická společnost, z. s., Technická 2, 166 27 Praha 6 Vytisklo: Nakladatelství ČVUT, výroba
Počet stran: 24 Počet výtisků: 200
Redakční rada: M. Brothánek, O. Jiříček, J. Kozák, R. Čmejla, J. Volín
Jazyková úprava: R. Svobodová, M. Tharp

Uzávěrka příštího čísla Akustických listů je 31. října 2018.