An Experimental Study on the Phase Importance in Digital Processing of Speech Signal

Lazar Tesic, Boban Bondzulic, Milenko Andric, Boban Pavlovic

Military Academy, University of Defence in Belgrade, Generala Pavla Jurisica Sturma 33, 11000 Belgrade, Serbia
e-mails: lazar.tesic@vs.rs, boban.bondzulic@va.mod.gov.rs, milenko.andric@va.mod.gov.rs, boban.pavlovic@va.mod.gov.rs

Abstract: This paper presents the results of a late large-scale subjective study of phase importance of speech quality. Present study includes a collection of speech sentences distorted by limited transmission bandwidth and phase degradations. A detailed statistical analysis of the collected subjective judgments is presented. Among the used signal distortions, subjects preferred modification with signal phase preservation. Mean opinion scores of such modification are the closest to the original, undistorted sentences. This subjective study contributes to improving the algorithms of the speech processing, and in addition provide valuable data to develop objective or automatic methods of speech quality assessment, as well as to estimate their performance.

Keywords: phase spectrum; phase importance; speech signal; subjective evaluation

1 Introduction

Over the past decade listening devices and speech communication devices are more frequently used. Users of such devices expect their devices to provide good quality and intelligibility anywhere and at any time [5].

Most of the used digital processing approaches of speech signals exploit a short-time Fourier transform (FT). In this domain, signal is represented with complex-valued coefficients, which can, therefore, be observed by their magnitude and their phase. It is well known that for one dimensional and two dimensional signals, the magnitude and the phase of the FT play different roles in the signal reconstruction. Many authors emphasize that the phase of the FT is more important than the magnitude [2, 7, 11, 15-17]. The confirmation of this notion they mainly verified through signal modification in such way that phase is preserved, but the magnitude of all the spectral components is set to unity, i.e.

they did not consider end users (listeners or auditory) opinion. For example, authors in [2, 7, 15-17], illustrated the phase importance through phase-only image reconstruction. Moreover, authors of [11] employed objective speech quality measures for perceptual estimation of speech quality – signal-to-noise ratio (SNR) and perceptual evaluation of speech quality (PESQ) [8].

Justification of phase importance from a statistical viewpoint has been presented in [14]. It was shown that a random distortion of the phases can dramatically distort the reconstructed signal, while a random magnitudes distortion will not.

Most of the research papers, which study digital processing of speech signals through subjective quality assessment consider speech signal coding algorithms [9] and speech enhancement under noisy conditions [6, 19, 20]. International Telecommunication Union (ITU) coded-speech database [9] contains coded and source speech material used in the ITU-T 8 kbit/s codec. Authors in [20] have presented a Czech language speech database in a car environment (database contains signals in a quiet car without background noise, background noise in running car without speech and speech signals in running car). Paper [19] presented a database designed to evaluate the performance of speech recognition algorithms in noisy conditions (suburban train, crowd of people, car, exhibition hall, restaurant, street, airport and train station noises). Study reported in [6] shows performance of different speech enhancement algorithms on noisy speech corpus database (NOIZEUS). The same database has been used in [18], where the authors have shown that by modifying the phase spectrum in the enhancement process the quality of the resulting speech can be improved.

In the paper [22], the relation between uncertainty in phase and word error rate (WER) in human speech recognition through the subjective tests has been investigated. It has been shown that at an SNR of -10 dB, having random phases at all frequencies results in a WER of 63% compared to 24% if the phase was unaltered. With the SNR of 0 dB, random phase results in a 25% WER in respect to 11% for unaltered phase. It has been concluded that at high SNRs (i.e. 20 dB) the effect of phase on WER is small in comparison with low SNRs (such as 0, -5 and -10 dB), where the effect of phase on the recognition error rate can be significant.

Results from human listening tests [1] indicate that even for small window durations (20-40 ms), the phase spectrum can contribute to speech intelligibility as much as the magnitude spectrum.

Phase spectrum is used for quantifying speech quality in [4, 10]. In [10] authors proposed phase distortion deviation measure (PDD), which was evaluated in a database of dysphonic speakers with spasmodic dysphonia. They have shown that PDD is highly correlated with subjective ranking from medical doctors. Study [4] demonstrated that phase deviation objective metric is a reliable speech quality estimator, with performance in line with PESQ metric [8].

Good literature surveys about magnitude and phase importance in signal processing applications can be found in [13, 23]. While paper [23] presents a review on techniques for signal reconstruction without phase, paper [13] demonstrates the importance of phase in different applications including: speech enhancement, automatic speech, speaker recognition and speech synthesis.

This paper is focused on studying the phase importance in digital processing of speech signals through a large-scale subjective study. A human study was conducted using limited transmission bandwidth and phase degradations as speech sentence's distortions. Subjective trial was run at the University of Defence in Belgrade, Serbia, during the latter half of July 2015 and involved 18 listeners, evaluating 144 speech sentences. The listener responses gathered in the response directory at the end of trial were analysed and their basic statistics evaluated for each degraded sentence. Subjective trial provided an opportunity of evaluating listener performance within the context of speech observation.

The quality of speech as perceived by listeners is becoming increasingly important, due to the large number of listening and speech communication devices that humans utilize as the end users of speech.

The rest of the paper is organized as follows: Section 2 illustrates the magnitude- and phase-only signal reconstruction; in Section 3 the description of the used filters (signal degradations) is done; Section 4 presents the characteristics of the equipment and software tools used for signal acquisition; Section 5 describes the original and distorted sentences used, while Section 6 describes the subjective results and statistical analysis of the obtained results; finally, the conclusion is given in Section 7.

2 Reconstruction of Speech Signal Using Magnitude or Phase Spectrum

Information preservation about source (speech) signal in magnitude and phase spectrum is shown through example in Figure 1. The figure illustrates the waveforms of source signal (Figure 1(a)), signal reconstructed from the magnitude spectrum of source signal (Figure 1(b)) and the waveform of a signal produced using phase spectrum of source signal (Figure 1(c)). Reconstruction is performed by calculating inverse discrete Fourier transform (IDFT) over magnitude and phase spectrum of the source signal. When reconstructing from magnitude spectrum, it is assumed that the phase spectrum equals zero, and when reconstructing from the phase spectrum, it is assumed that the magnitude spectrum equals one. From Figure 1 it is obvious that the signal reconstructed from magnitude spectrum significantly differs from the source signal. Waveform of the source signal is preserved in the phase spectrum reconstructed signal.

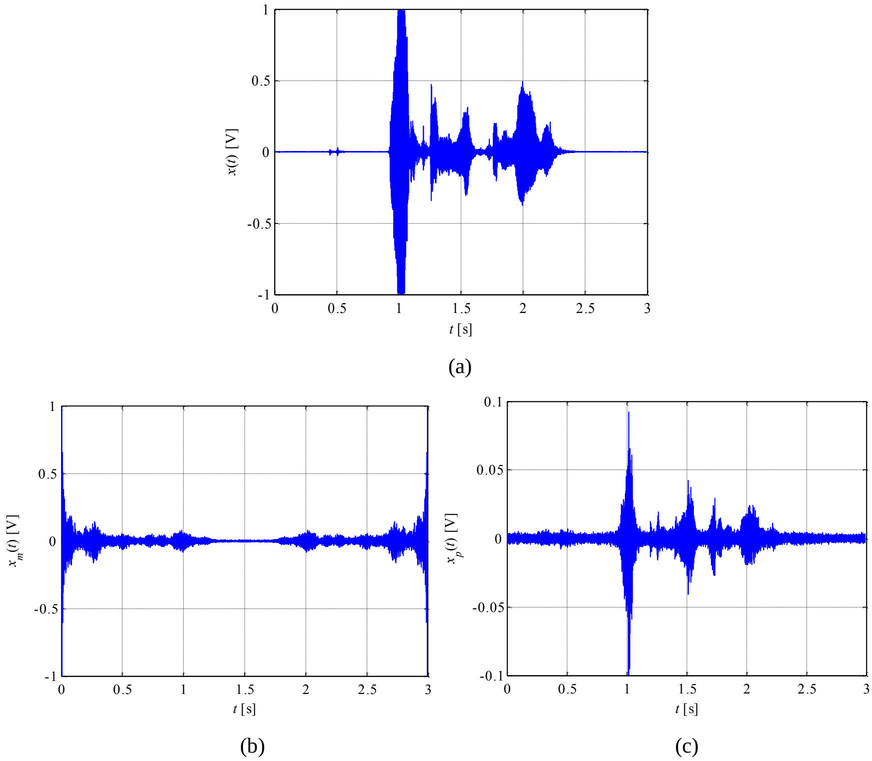


Figure 1
Waveforms of the: (a) source signal, (b) signal reconstructed from the magnitude spectrum of source signal and (c) signal reconstructed using the phase spectrum of source signal

3 Description of the Used Filters

For the purpose of phase spectrum importance research in digital signal processing of speech signals, verbal sentences are filtered with four types of low pass filters, which are marked as Type 1, Type 2, Type 3 and Type 4 (Figure 2). Type 1 represents the ideal low pass filter, with zero phase response. Type 4 is the real elliptic low pass filter. The remaining two types are produced by combining the magnitude ($\mu(f)$) and phase responses ($\theta(f)$) of the above-mentioned ideal and elliptic filters (Figure 2). Type 2 filter has ideal magnitude response and phase response taken from elliptic filter, and Type 3 filter has zero phase response and magnitude response taken from the elliptic filter.

Ideal low pass filter is designed for the cutoff frequencies of 1 kHz and 2 kHz, while the specifications of elliptic filter are given in Table 1.

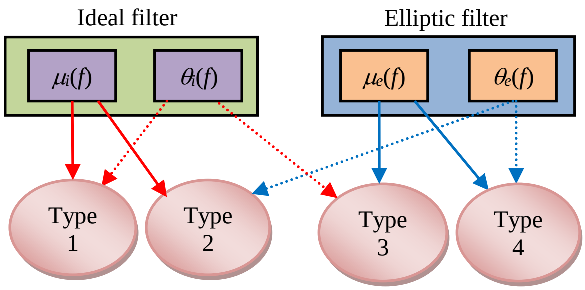


Figure 2
Filter design scheme

Table 1
Specifications for elliptic digital filter design

Cutoff frequency of the passband, f_g	1 kHz	2 kHz
Cutoff frequency of the stopband, f_a	2 kHz	3 kHz
Peak passband ripple	1 dB	
Minimum stopband attenuation	30 dB	
Sampling frequency	16 kHz	

The elliptic filter is an infinite impulse response (IIR) filter, and has a steep roll-off with equiripple in both passband and stopband. Using an elliptic filter design it is possible to achieve the lowest order for a given set of specifications [12]. The cutoff frequencies of projected filters are within a frequency range that contains the majority of the speech signal energy content (from 300 Hz to 2.8 kHz).

Elliptic function computations carried out by the following MATLAB® functions: (1) `ellipord` – for elliptic filter order selection and (2) `ellip` – which returns the filter coefficients.

The transfer functions $H(z)$ of the filters designed according to the specifications in Table 1 are given by:

$$H_{1\,\mathrm{kHz}}(z) = \frac{0.0274 - 0.0130z^{-1} - 0.0130z^{-2} + 0.0274z^{-3}}{1 - 2.5107z^{-1} + 2.2227z^{-2} - 0.6833z^{-3}} \tag{1}$$

$$H_{2\,\mathrm{kHz}}(z) = \frac{0.0529 - 0.0467z^{-1} + 0.0940z^{-2} - 0.0467z^{-3} + 0.0529z^{-4}}{1 - 2.6643z^{-1} + 3.2092z^{-2} - 1.9105z^{-3} + 0.4849z^{-4}} \tag{2}$$

Equations (1) and (2) show that the given specifications can be realized with elliptic filters of 3rd (f_g=1 kHz) and 4th order (f_g=2 kHz).

Locations of poles (x) and zeros (o) of the designed elliptic filters, Eqs. (1) and (2), are illustrated in Figure 3. The designed filters are stable because their poles are within the unit circle [3, 21, 24].

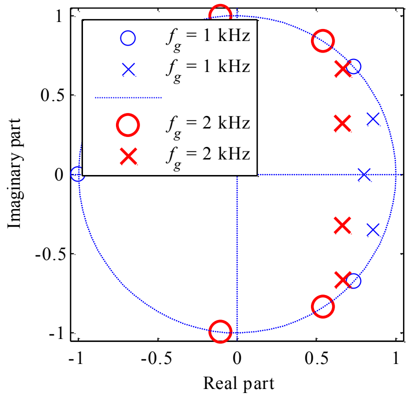
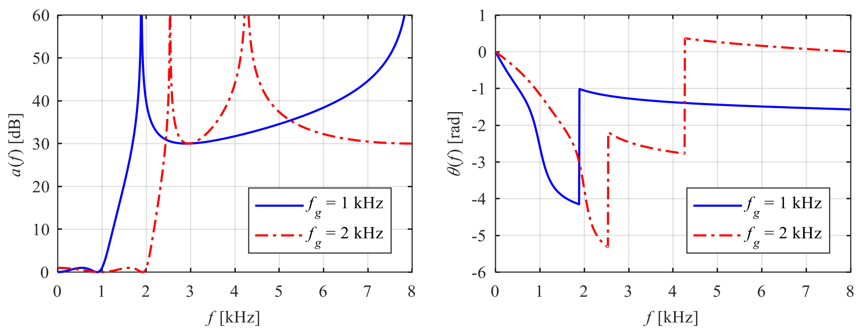


Figure 3
Locations of poles and zeros in the z-plane of the elliptic filters designed according to the specifications in Table 1

The magnitude and phase responses of the designed elliptic filters are given in Figure 4. Figure 4(a) shows that the attenuation requirements – peak passband ripple and minimum stopband attenuation, are satisfied. The phase responses of designed filters, Figure 4(b), are nonlinear. The nonlinear phase response resulting in group delays, *gd*, that are not constant in the passbands of the filters (Figure 5), which means that the components that enter the filter will not be delayed by the same value of time on its exit.



(a) (b)

Figure 4
(a) magnitude responses and (b) phase responses of the filters designed according to the specifications given in Table 1

An ideal filter (it is ideal in the sense that it is not realizable) is noncausal, and it was applied by calculating discrete FT (using MATLAB® function `fft`) of the input signal and by leaving all frequency components of a signal below a designated cutoff frequency (all components above are rejected).

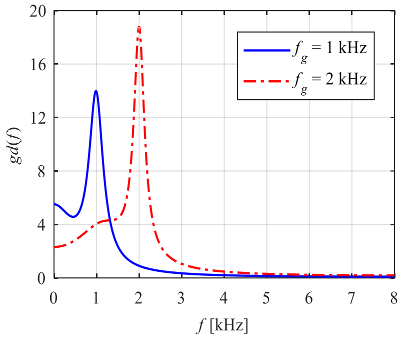


Figure 5

Group delays of the filters designed according to the specifications given in Table 1

4 Characteristics of the Equipment and Software Tools Used for Signal Acquisition

Test signals used in this research are audio – speech sentences. For speech acquisition Genius HS–400A Headband PC Headset with rotating microphone is used. Its features are shown in Table 2 and Table 3.

Table 2

Headphone features

Wearing style	on ear
Driver unit	40 mm
Frequency response	20 Hz – 20 kHz
Impedance	32 Ω
Sensitivity	102 dB
Cable length	1.8 m
Volume control	yes
PC in	2x3.5 mm stereo jack

Table 3

Microphone features

Sensitivity	-54+/-3 dB
Directivity	omni-direction
Frequency response	100 Hz – 10 kHz

PC used for recording and signal processing is a laptop computer Acer Aspire 5750G. Software used for signal processing was MATLAB® – version R2013a. A/D signal conversion was done with sampling frequency of 16 kHz, with 16-bit resolution.

5 Forming the Speech Database and Running the Subjective Tests

The phase importance research was realized through four phases: (1) gathering source speech sentences, (2) forming the speech database, (3) subjective tests running, (4) results arrangement and analysis.

First phase included source speech sentences recording. Three native speakers (two male and one female) were reading six Serbian sentences each (translation in English is in brackets):

"Aba je najbolji bend." (Abba is the best band.)

"Ada je recno ostrvo." (Ada is a river island.)

"Igi Pop je muzicar." (Iggy Pop is a musician.)

"Op'o mi pritisak." (My blood pressure went down.)

"Oko mi je crveno." (My eye is red.)

"OTO je predmet u osnovnoj skoli." (OTO is a subject in elementary school.)

Every sentence was recorded. Hardware and software tools used for signal recording and processing were described above.

In this way 18 source signals were recorded (sources-originals-references) lasting 3 seconds each. These sentences are the signals to be filtered in the next research step.

These sentences meet the ITU conditions [8]. Namely, sentences should be formed in that manner so that the six most frequent consonants (b, d, g, p, k, t), should be placed between two same vocals.

In the second research phase test speech sentences were formed (degraded sentences). Eighteen signals mentioned before were filtered with four types of filters described earlier. Filters were projected for two different cutoff frequencies, so every source sentence was modified in 8 different ways (4 types of filters x 2 cutoff frequencies). In this way, 144 test sentences were produced. So, the complete database contains 162 speech recordings (18 originals + 144 degraded signals), and they represent the ground basis for subjective testing. Using different types of filters and different cutoff frequencies, different signal quality is achieved. Therefore, different grades are expected.

Procedures and standards for subjective evaluation of speech signals have existed for many years. Because of the variety of digital audio contents in general, running subjective evaluation tests, for estimating audio quality, became very simple.

Quality estimation of test speech sentences was performed through subjective tests (third research phase). Subjective testing was performed on a representative sample of several listeners. The testing followed ITU recommendation [8], which defined a phonetic approach in subjective testing of speech signals. Test signals must be produced by male as well as female speakers. Recommended duration of sentences is 1 to 3 seconds, where pure speech must be in a range of 40 to 80%. Minimum sampling frequency is 8 kHz and resolution should be 16 bits. Optimal number of listeners is 12.

Subjective tests included 18 listeners. The process went like this: the original sentence is played first, and the listener knew that, so he/she didn't evaluate the first sentence. After that, four respective degraded sentences are played (with cutoff frequency of 1 kHz of the passband) and at the end the original is played again, and the listener didn't know that he listens to the original (so he/she evaluated the original). All five sentences are graded with scores from 1 to 5 (1 – bad quality, 5 – excellent quality). During evaluation, speech comprehension sound purity, sound loudness and the sentence tone are taken into account. When the first round ends, sentences at the cutoff frequency of 2 kHz of the passband are listened to again and evaluated.

Each listener had to evaluate 27 different sentences (24 degraded and 3 originals). The listener, firstly, listened to the three originals from one speaker (and its respective degraded sentences), and then listened to subsequent three originals, and so on. Degraded sentences are randomly played, as suggested in [9].

6 Results and Analysis

The outcomes of subjective tests are quality-grades given by listeners, which are presented through mean opinion scores (MOS). MOS is the most commonly used method of generalising a subjective score given by a number of independent observers with respect to perceived quality of a signal. Extensively used in both subjective audio (in particular voice over IP (VoIP) communications) and video quality evaluation, MOS is simply the arithmetic mean of all individual scores assigned by the listeners (test participants) to a signal:

$$MOS_i = \frac{1}{N_S}\sum_{n=1}^{N_S} SQ(n,i) \tag{3}$$

where are:

i – degraded sentence index,

$SQ(n,i)$ – subjective grade given by n-th listener to i-th sentence and

N_s – the number of listeners who graded i-th sentence.

MOS is within the same range as the quality scale adopted during the subjective trials so there is no need for additional normalisation of the range. This is most commonly 0 or 1 to 5, as is the case with subjective trials ran here where 1 means the lowest and 5 the highest quality. MOS is a democratic measure in that it treats each subjective vote equally and the only true mean opinion.

Since subjective MOS scores are constructed from several individual quality scores, their estimate of absolute video quality has a statistical uncertainty associated with it. In cases where individual subject quality scores vary widely this uncertainty is large. Subjective score uncertainty can be measured using a number of methods, but the benchmark metrics of uncertainty are standard deviation, or its square the variance, and standard error. Standard deviation is evaluated directly from the individual subjective scores for sentence i, SQ_i (index n of the listener used previously is ignored here for the sake of brevity) and the MOS for that speech signal:

$$\sigma_i = \sqrt{\mathrm{E}\left[SQ_i^2\right] - \left(\mathrm{E}\left[SQ_i\right]\right)^2} = \sqrt{\mathrm{E}\left[SQ_i^2\right] - MOS_i^2} \tag{4}$$

where $E[X]$ is expected value or an average score of the random variable. Standard deviation is mostly shown together with an average score as its positive and negative variation.

The alignment of quality scores vary from listener to listener. Since listeners are free to choose which grade to assign, it is natural to expect that grades given to sentences differ from one listener to another. More generous listeners provide higher scores, while those who are stricter choose lower ones. Although the absolute range of the subjective scores doesn't influence the speech signal ranks, contradictory scores may influence MOS.

Equations (3) and (4) show the way of determining average grades and standard deviations of subjective scores of a single sentence (i-th sentence), which listeners evaluated. In this paper, the analysis of gathered scores was done through calculating average scores and standard deviations of subjective grades given to test sentences which are produced from the same source sentence, subjective scores of the sentences which were listened by every listener alone, subjective scores of test sentences which are produced by every speaker alone and subjective scores associated to different modifications (filter types) of a source speech sentences.

Figure 6 shows MOS scores histogram with 10 equally spaced bins – vertical axes corresponding to the number of sentences in each bin. We can see that the subjective trial contained a good spread of speech quality as mean opinion scores show variation of around 75% of the entire score range (from 2 to 5). Histogram maximum is in the medium quality area, wherein the dissipation of subjective scores around the maximum can be approximated with Gaussian (normal) distribution.

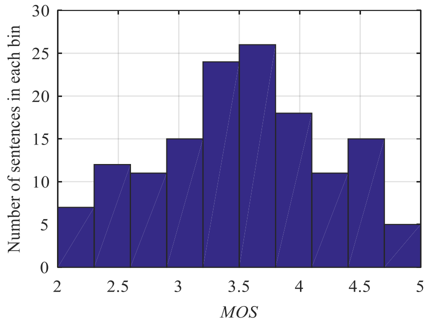


Figure 6
MOS scores histogram with 10 equally spaced bins

Figure 7 shows MOS values with respective reliability intervals (MOS±σ) of subjective scores of test sentences which are produced from the same source sentence (3 speakers x 6 sentences = 18 source sentences). A group of source sentences with the ordinal numbers 1 to 6 originating from the first speaker, group of sentences with the ordinal numbers from 7 to 12 originating from the second speaker, and the last group of six sentences originating from the third speaker. MOS values are within the range from 3.25 (sentence number 3) to 3.79 (sentence number 13), while standard deviations are within the range of 0.78 (sentence number 10) to 1.06 (sentence number 12).

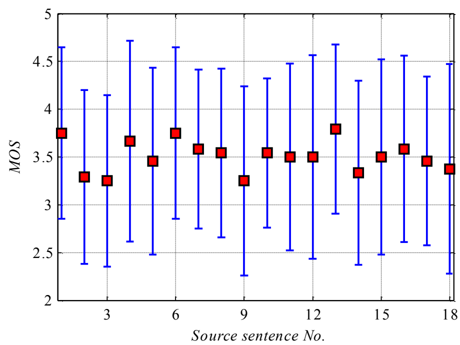


Figure 7
MOS values with respective reliability intervals of subjective scores of test sentences which are produced from the same source sentence

Listener performance is another important indicator of subjective trial success and quality/usefulness of resulting results. With respect to individual listeners' opinions we are primarily interested in analysing the consistency of their scoring. Obviously trials with high levels of agreement between individual subjects are more useful as the certainty of the resulting subjective quality scores is high. Conversely, in cases where subjects disagree and we have a large variance in subjective quality and ranking we cannot be certain which speech sentence really has a quality advantage or it if the advantage really exists. In reality agreement between subjects is never ideal and a certain level of uncertainty remains for both absolute quality and ranking based on quality. The simplest approach to measure general subject performance within a trial is to evaluate some basic statistics of their quality responses.

The basic statistics of listeners' subjective scores (average scores and standard deviations) are shown on Figure 8. Average values are within a range from 3.3 to 3.83, with relatively equal reliability intervals. Although the difference between the MOS values exist, it doesn't point out one listener from the majority.

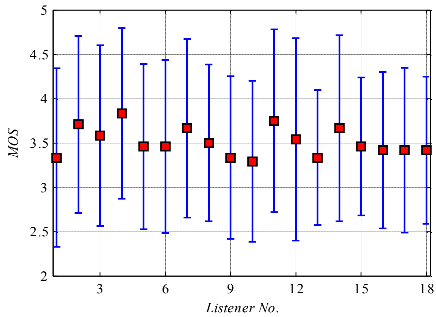


Figure 8

MOS values with respective reliability intervals of subjective scores given to sentences by one listener

Figure 9 shows MOS values with reliability intervals of subjective scores of test sentences which came from the same speaker (three speakers). It is obvious that the mean scores are relatively equal and that they are about 3.5. Standard deviations are also very close, and they are between 0.916 (speaker 2) and 0.968 (speaker 3).

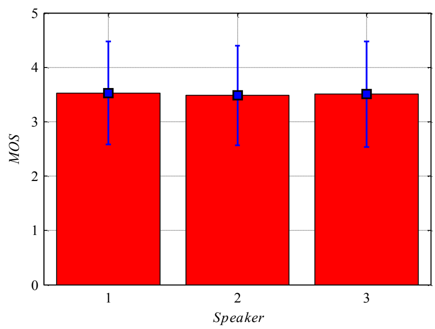


Figure 9

MOS values with respective reliability intervals of subjective scores of test sentences which came from the same speaker

To analyse the ranking of various degraded sentences the results are best viewed on a filter level. The scores are aggregated according to filter and shown on Figure 10, which shows MOS values of subjective scores given to the sentences with the same type of degradation – filter type analysis. Filter types are described in Section 3. Beside the average scores of the test sentences, Figure 10 shows the average value of scores given to the source sentences (without degradation).

Figure 10 shows that the biggest MOS value belongs to the originals – source sentences (4.85), which was expected. Additionally, listeners evaluated that better quality is maintained when sentences passed through filters Type 3 and 4 then through filters Type 1 and 2. This was also expected because filters Type 1 and 2 have ideal magnitude responses (both with cutoff frequencies of 1 kHz or 2 kHz), while filters Type 3 and 4 have magnitude responses with transition zones 1 kHz wide (both types include transition zones from 1 kHz to 2 kHz or from 2 kHz to 3 kHz). Transition zone enables for the spectral components which are higher than cutoff frequency of the passband, not to be completely attenuated, i.e. to be perceptually noticeable.

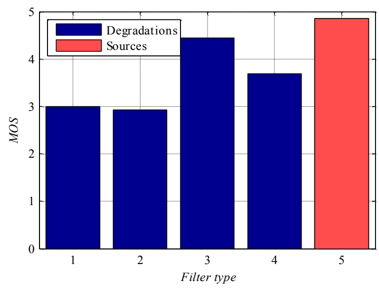


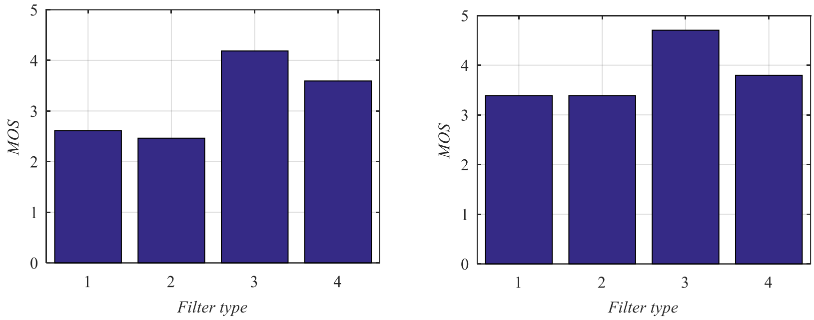
Figure 10

MOS values of subjective scores given to the sentences with the same type of degradation

Test sentences which are modified with Type 3 filter (magnitude response is the same as elliptic filter and phase response is zero) are graded with the highest average score of 4.44. Phase response of this filter is set to zero, so, input and output signals have the same phase spectrums, i.e. the original signal phase is completely preserved. The importance of phase preservation is obvious when MOS values of the filters of the same magnitude responses are compared (Type 1 versus Type 2, or Type 3 versus Type 4). In both cases, zero phase filters have the priority (Type 1 and Type 3). The benefit in phase preservation is much greater when analyzing the filter with non-ideal (real) magnitude response (MOS_3=4.44, MOS_4=3.69) and comparing it to the filter with ideal magnitude response (MOS_1=3, MOS_2=2.92). This result can be interpreted with the phase preservation influence – when analyzing filters with ideal magnitude response, phase response can be analyzed only in the passband, while when analyzing filters with real magnitude response the phase influences the transition zone and the stopband band as well (spectral components of the original signal are not completely attenuated).

Figure 11 shows MOS values of subjective scores given to the test sentences with the same type of degradation, when different cutoff frequencies of the passband of designed filters are taken into analysis separately.

Comments which refer to the source signal degradation analysis, when analyzing the complete database (Figure 10), are all in line when analyzing the source signal degradation with both cutoff frequencies separately (Figure 11). Additionally, MOS values of degraded sentences provided through filters with 2 kHz cutoff frequency of the passband (Figure 11(b)) are higher than MOS values of degraded sentences provided with filters with 1 kHz cutoff frequency of the passband (Figure 11(a)).



(a) (b)

Figure 11

MOS values of subjective scores given to the test sentences with the same type of degradation: (a) for the cutoff frequency f_g=1 kHz and (b) for the cutoff frequency f_g=2 kHz

The analysis results indicate that the test material was correctly chosen and prepared; also the subjective tests were well done. Because of that, the test sentence database with the entire subjective test results can be used for development of objective quality estimation algorithms for speech (audio) signals. The idea is to compare directly, MOS quality values with the values gained from the objective quality estimation algorithms.

We have decided to make the test material available to the research community free of charge [25]. Along with speech sentences and subjective scores, we provided MATLAB® files, too.

Conclusions

A subjective study to evaluate the phase importance on the perceptual quality of speech communication was presented. This study included 144 speech sentences derived from 18 original sentences using eight distortion types and were evaluated by 18 listeners. Subjective quality data collected through subjective trials was processed into the form of mean opinion scores expressing mean estimates of speech quality for each degraded sentence used in the study. The resulting database is unique in terms of content and distortion and is publicly available to the research community for further research on speech quality assessment.

For the purpose of validation, a specific set of software tools was constructed for conducting the validation and performing the comparison between subjective quality scores. Specifically, a set of statistical tools was designed and implemented in Matlab® development environment that allows reading, comparison and output of a set of both quantitative and qualitative validation scores.

The results of the performed study show that in the modeling and processing of the time-frequency signal representation, phase can't be ignored. Subjective results may with further analysis provide deeper insight into how people decide and what influences the decisions they make regarding perceived speech quality.

In future work the speaker database should be extended. Furthermore, we will develop objective speech quality assessment measure, with special attention to phase preservation measuring.

Acknowledgement

This research has been a part of the project No. VA-TT/1-17-19 supported by the Ministry of Defence, Republic of Serbia.

References

[1] Alsteris, L. D., Paliwal, K. K.: Further intelligibility results from human listening tests using the short-time phase spectrum, Speech Communication, 2006, Vol. 48, No. 6, pp. 727-736

[2] Behar, J., Porat, M., Zeevi, Y. Y.: Image reconstruction from localized phase, IEEE Transactions on Signal Processing, 1992, Vol. 40, No. 4, pp. 736-743

[3] Besancon-Voda, A.: Iterative auto-calibration of digital controllers: methodology and applications, Control Engineering Practice, 1998, Vol. 6, No. 3, pp. 345-358

[4] Gaich, A., Mowlaee, P.: On speech quality estimation on phase-aware single-channel speech enhancement, In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane (Australia), 2015, pp. 216-220

[5] Gerkmann, T., Krawczyk-Becker, M., Le Roux, J.: Phase processing for single-channel speech enhancement: History and recent advances, IEEE Signal Processing Magazine, 2015, Vol. 32, No. 2, pp. 55-66

[6] Hu, Y., Loizou, P. C.: Subjective comparison and evaluation of speech enhancement algorithms, Speech Communication, 2007, Vol. 49, No. 7, pp. 588-601

[7] Huang, T. S., Burnett, J. W., Deczky, A. G.: The importance of phase in image processing filters, IEEE Transactions on Acoustics, Speech and Signal Processing, 1975, Vol. ASSP-23, No. 6, pp. 529-542

[8] ITU-T Recommendation P.862: Perceptual evaluation of speech quality (PESQ), An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, February 2001

[9] ITU-T Telecommunication standardization sector of ITU, Series P: Telephone transmission quality, telephone installations, local line networks, Supplement 23: ITU-T coded-speech database, February 1998

[10] Koutsogiannaki, M., Simantiraki, O., Degottex, G., Stylianou, Y.: The importance of phase on voice quality assessment, In Proceedings of the 15th Annual Conference of the International Speech Communication Association (INTERSPEECH), Singapore, 2014, pp. 1653-1657

[11] Langarani, M. S. E., Veisi, H., Sameti, H.: The effect of phase information in speech enhancement and speech recognition, In Proceedings of the 11th International Conference on Information Sciences, Signal Processing and their Applications (ISSPA), Montreal (Canada), 2012, pp. 1446-1447

[12] Mitra, S. K.: Digital Signal Processing – A Computer Based Approach, 3rd ed., New York (USA): McGraw-Hill, 2006, ISBN: 007-124467-0

[13] Mowlaee, P., Saeidi, R., Stylianou, Y.: INTERSPEECH 2014 Special Session: Phase importance in speech processing applications, In Proceedings of the 15th Annual Conference of the International Speech Communication Association (INTERSPEECH), Singapore, 2014, pp. 1623-1627

[14] Ni, X.S., Huo, X.: Statistical interpretation of the importance of phase information in signal and image reconstruction, Statistics & Probability Letters, 2007, Vol. 77, No. 4, pp. 447-454

[15] Oppenheim, A. V., Hayes, M. H., Lim, J. S.: Iterative procedures for signal reconstruction from Fourier transform phase, Optical Engineering, 1982, Vol. 21, No. 1, pp. 122-127

[16] Oppenheim, A. V., Lim, J. S.: The importance of phase in signals, Proceedings of the IEEE, 1981, Vol. 69, No. 5, pp. 529-541

[17] Oppenheim, A. V., Lim, J. S., Kopec, G., Pohlig, S. C.: Phase in speech and pictures, In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Washington, D.C. (USA), 1979, Vol. 4, pp. 632-637

[18] Paliwal, K., Wojcicki, K., Shannon, B.: The importance of phase in speech enhancement, Speech Communication, 2011, Vol. 53, No. 4, pp. 465-494

[19] Pearce, D., Hirsch, H.-G.: The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy

conditions, In Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP), Beijing (China), 2000, Vol. 4, pp. 29-32

[20] Pollak, P., Vopicka, J., Hanzl, V., Sovka, P.: CAR2 – Czech database of car speech, Radioengineering, 1999, Vol. 8, No. 4, pp. 1-6

[21] Precup, R.-E., Preitl, S.: PI and PID controllers tuning for integral-type servo systems to ensure robust stability and controller robustness, Electrical Engineering, 2006, Vol. 88, No. 2, pp. 149-156

[22] Shi, G., Shanechi, M. M., Aarabi, P.: On the importance of phase in human speech recognition, IEEE Transactions on Audio, Speech and Language Processing, 2006, Vol. 14, No. 5, pp. 1867-1874

[23] Sturmel, N., Daudet, L.: Signal reconstruction from STFT magnitude: A state of the art, In Proceedings of the 14th International Conference on Digital Audio Effects (DAFx), Paris (France), 2011, pp. 375-386

[24] Tar, J. K., Bitó, J. F., Rudas, I. J.: Contradiction Resolution in the Adaptive Control of Underactuated mechanical Systems Evading the Framework of Optimal Controllers, Acta Polytechnica Hungarica, 2016, Vol. 13, No. 1, pp. 97-121

[25] Tesic, L., Bondzulic, B., Andric, M., Pavlovic, B.: An experimental study on the phase importance in digital processing of speech signal – Phase importance (PhImp) database [Online]. Available: https://mega.nz/#!WsMEhTJQ!IHtEQ4XQ_Ir-tRPU-qqbmf0xcqNsl2TWmTJXt_PWCnQ, October 2017