

PAPER • OPEN ACCESS

Investigation of Influence of Speech Specter Differences on Intelligibility Estimation Results

To cite this article: A V Ivanov *et al* 2018 *J. Phys.: Conf. Ser.* **944** 012044

View the [article online](#) for updates and enhancements.

You may also like

- [EEG-based auditory attention detection: boundary conditions for background noise and speaker positions](#)

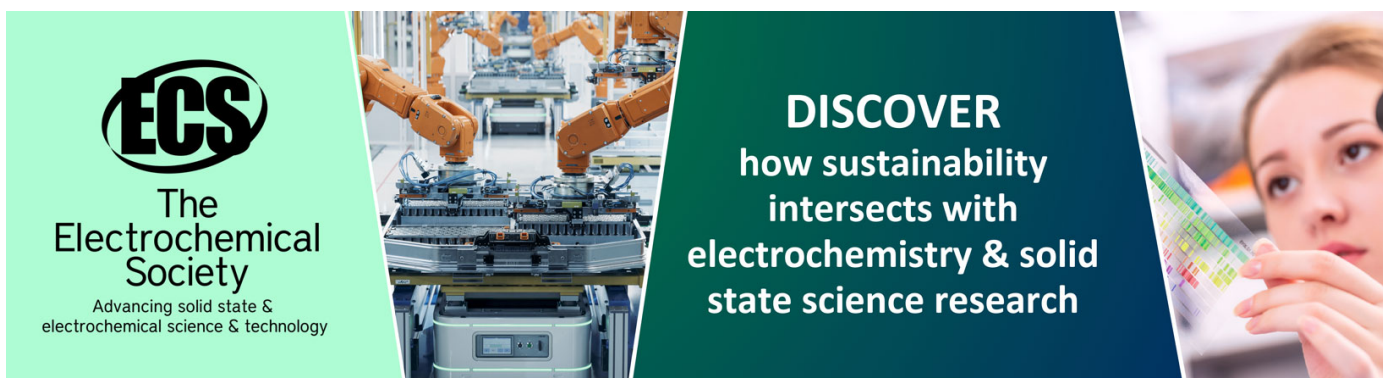
Neetha Das, Alexander Bertrand and Tom Francart

- [Binaural Speech Intelligibility in a Real Elementary Classroom](#)

G Minelli, G E Puglisi, A Astolfi et al.

- [Direct speech reconstruction from sensorimotor brain activity with optimized deep learning models](#)

Julia Berezutskaya, Zachary V Freudenburg, Mariska J Vansteensel et al.



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Investigation of Influence of Speech Specter Differences on Intelligibility Estimation Results

A V Ivanov, I L Reva, and E E Shemshetdinova

Novosibirsk State Technical University, 20, Karla Marksa ave., Novosibirsk, 630073, Russia

E-mail: andrej.ivanov@corp.nstu.ru

Abstract. The purpose of this study is to investigate influence of difference of speakers' speech specters on word intelligibility. Experiment on estimation of speakers' speech specter difference was conducted with taking into account an effect of a text which a speaker uses. Theoretical calculation of speaker's speech specter influence on intelligibility was done with use of voice information protection approach. The influence of speaker's speech specter on intelligibility was confirmed by immediate appreciation tests. Recommendations on possible consideration of the effect are provided.

1. Introduction

Standard approach of estimation of voice information protection [1] is the approach of word intelligibility estimation, based on the Pokrovsky's calculating-experimental formant approach [2]. The approach was developed for conditions, which are essentially different from information protection. As the result, it led to appearance of essential limitations, many of which are recently discussed by a lot of authors who try to adjust the chosen approach for information security conditions [3-8].

One of the important aspects of the approach is use of average speech specter. However, it is known that different people have different speech specters because every human's articulation organs are unique, beside essential differences of men's and women's speech specters. Therefore, there are questions: is it correct to use the average speech specter in word intelligibility estimation; what is a range of values of word intelligibility with use of different speakers' speech specters?

Thus, the purpose of the study is to investigate how differences of speech specters influence on the result of word intelligibility.

2. Word intelligibility estimation approach

Word intelligibility is the index of voice information protection, which is stated as ratio of correctly received speech elements to the number of transmitted speech elements.

Essence of the approach [1] is estimation of sound pressure levels (vibrational acceleration) of the following signals octave bands:

- test signal, formed with use of white noise generator or tone signals with frequencies corresponding to average compound frequencies of octave bands;
- background noise or noise created by active protection means, measured behind a protective structure with a turned off test signal generator and a turned on active protection mean (if any);



- signal-noise mixture, measured behind a protective structure with a turned on test signal generator;

Calculation of word intelligibility is made with results of the measurements. The approach is described in details in [1, 2].

It should be mentioned that test signal is a signal with integrity levels in octave bands corresponding to the average speech level. Distribution of energy in spectrum for integrity level 70 dB is characterized by octave bands' integrity levels, shown in table 1.

Table 1. Octave levels of sound pressure of average speech specter

Frequency, kHz	0.125	0.250	0.5	1	2	4	8
Sound pressure level, dB	53	66	66	61	56	53	49

3. Experiment

1. Selection of text

An experiment on estimation of sound pressure levels in octave bands of different speakers was conducted.

The first problem of the experiment is selection of a text for speakers and its wordage. The problem was solved by another experiment. Four speakers (2 men and 2 women) were offered two texts:

- phrasal articulatory tables, presented in Russian national standard «GOST R 50840-95 Speech transmission by communication channels. Estimating methods of quality, intelligibility and awareness (Attachment D)» [9];
- a part of a technical text with the same wordage.

Recording duration was set as 1 minute. Speakers' speech was calm, clear, with average volume – characteristics of speech in real life. Noise level meter ZET 110 and microphone BC501 with 1kHz calibrated frequency using calibrator SVAN SV35 was used for the recording. The recording was made with sampling frequency 50 kHz. Another aim of the experiment was to investigate results repetition, that is why all the texts were recorded twice.

Specters of the recordings were plotted with use of the Adobe Audition functions. Samples of the specters help to calculate octave bands integrity levels. We should mention that audio editors' levels are estimated in dBFS – full scale (dB by voltage, relatively to the maximum A/D Converter value set on a PC). It means that 0 dBFS does not correspond to the limit value of signal level, and there will not be so-called «Clipping». Therefore, the minimum value of signal level fixed by the A/D Converter depends on its number of digits. For example, 16-digit A/D Converter's minimum signal level is -96 dBFS, 20-digit A/D Converter's minimum signal level is -120 dBFS, 24-sigit A/D Converter's minimum signal level is -144 dBFS. Traditional dB, which stand for sound pressure (relatively to 20 μ Pa) can be derived only if you know the A/D Converter's number of digits and microphone/acceleration meter sound pressure level (dB SPL). So, Adobe Audition does not allow to estimate absolute level of signal recording, but allows to compare spectral characteristics of signals.

Results of the experiments are octave bands signal levels for two recordings of one speaker with use of different texts (table 2) of the two recordings

Table 2. Octave levels of sound pressure with use of different texts

Sound pressure levels for different texts	Average compound octave bands' frequencies, Hz						
	125	250	500	1000	2000	4000	8000
Phrasal articulatory tables, dBFS	-36.69	-34.57	-34.72	-41.52	-45.33	-46.16	-57.01
Technical text, dBFS	-37.22	-33.65	-33.95	-40.75	-44.94	-46.43	-56.50

We can see that the difference of octave bands' sound pressure with use of different texts does not exceed 1 dD, so we can conclude that type of the text does not essentially influence on speech specter. Derived divergence (up to 1 dB) appears even when a speaker reads the same text. It means that the divergence appears because it is hard to speak on the same level.

We can try to reduce the effect by averaging levels of 2 texts.

Therefore, text type almost does not influence on speech specter. The technical text was chosen for further experiments because it is more similar with real life speech than special tables from the GOST.

2. Analysis of influence of speech specters differences on intelligibility

After selection of the text we can start statistics collection. 10 speakers (5 men and 5 women) without speech disorders and special characteristics were chosen for the experiment. The result of each speaker is calculated by averaging 2 recordings for increase of accuracy. The results are shown in table 3.

Reduce all levels to 70 dB for easy comparison of the results with average speech specter' level. On this step divergences are not critical, it is another work (estimation of amplitude speech constitution), described in details in [10]. We consider only specter forms with octave bands' levels.

Now we can substitute the results to [1] as test signal levels and analyze their influence on word intelligibility W . Now plot dependencies of W on signal-level ratio q with use of pink noise for all the speakers.

Table 3. Integrity levels of speakers' speech

F, kHz	Sound pressure levels in octave bands for different speakers, dB										
	№1	№2	№3	№4	№5	№6	№7	№8	№9	№10	Average
0.125	64.79	59.13	62.01	60.09	60.69	60.67	64.44	65.62	65.30	63.00	53
0.250	66.06	64.34	65.80	65.21	67.82	66.03	66.06	65.00	63.96	66.15	66
0.500	63.03	66.48	65.95	65.85	63.86	66.25	63.77	63.31	65.00	64.44	66
1	55.23	62.02	55.15	62.49	51.47	54.14	56.80	55.82	55.45	56.21	61
2	50.81	57.65	51.92	58.63	49.54	50.91	51.77	53.28	53.05	55.77	56
4	54.75	53.26	47.84	52.24	48.61	51.97	51.45	55.98	54.15	53.89	53
8	50.22	51.73	44.46	52.46	45.57	48.25	45.91	47.56	51.63	48.24	49

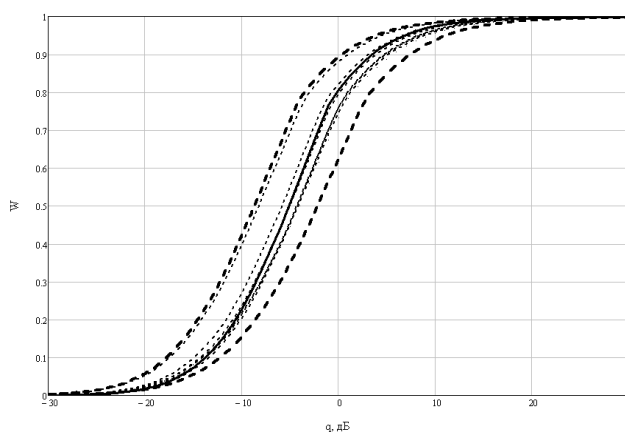


Figure 1. Dependences of word intelligibility W on signal-noise ratio q for pink noise (different speakers' speech specters). On the diagram: full line – average specter, bold dashed lines – curves with minimum and maximum values.

The diagram shows that, for example, if $q = -10$ dB for average speech specter, $W=0.22$. For speech specters of different speakers W falls in the limits from 0.15 to 0.43. Also we can notice that the lines with bigger W belong to men, with smaller W – to women.

3. Check of the results with use of immediate appreciation tests

We have done some appreciation tests to accept the results. Immediate appreciation test is a process of subjective estimation of word intelligibility. Its essence is addition of noises with particular signal-noise ratios to examined recordings, which after that are listened by auditors group.

Choose the speakers' recordings with the best and the worst results. These speakers are the speaker №5 (woman) and the speaker №4 (man), corresponding to table 2. By the way, theoretically (using the approach calculations), the difference between word intelligibility for different speakers is approximately 30%.

Add the pink noise and its generation in audio editor. Signal-noise ratio' value does not have big influence, but it should be the same for both speakers. Choose such a noise signal for the speech of speaker №5 (the minimum word intelligibility) to make only some words from it recognized, i.e. word intelligibility does not exceed 0.1-0.15.

After that we add the same noise with the same level to the speech of speaker №4 (maximal word intelligibility). The examined recordings were adjusted by level (nevertheless divergence between them was 0.3 dB).

As the result, even a small group of auditors admitted that word intelligibility of speaker №4 is essentially bigger (about 0.6-0.7). So, practical influence of speech specter form of speakers' speech have even bigger influence on word intelligibility than theoretical. Also, we should mention that the main differences were between man's and woman's voices.

4. Analysis of the results

The derived results show the following:

- speakers' speech specter essentially influence on word intelligibility, even from the approach point of view (difference is about 30%);
- immediate appreciation tests practically confirmed that the effect appears with bigger difference (50-60%);
- the main difference in word intelligibility is between man's and woman's speech.

The immediate appreciation tests' results can be explained that word intelligibility depends not only on speech specter, but more on distribution of formant in frequency range. The distribution is individual for each speaker and especially for men and women.

5. Conclusion

We can make several conclusions on the results. First of all, speech specter essentially influences on word intelligibility. Also, the work problem about text selection for tests done. The difference between phrasal tables from GOST and technical texts is less than 1 dB.

As the result, we can recommend to choose a speaker based on estimation of all speakers' speech specters.

If it is not possible, it is necessary to estimate protection level with use of the speech specter with the biggest word intelligibility instead of the average. Probably, for derivation of such a specter it is necessary to do additional immediate appreciation tests with bigger auditors' group.

Consideration of the effect in estimation of voice information protection would allow to increase information security level of private conversation rooms.

References

- [1] Zheleznyak V K, Makarov Yu K and Khorev A A 2000 *Spetsial'naya tekhnika* 4 39–45
- [2] Pokrovskiy N B 1962 *Raschet i izmerenie razborchivosti rechi* (Moscow: Svyaz'izdat) 390
- [3] Avdeev V B 2013 *Spetsial'naya tekhnika* 2 1–10
- [4] Didkovskiy V S, Didkovskiy M L and Prodeus A N 2008 *Akusticheskaya ekspertiza kanalov rechevoy kommunikatsii* (Kiev: Imeks-LTD) 420
- [5] Didkovskiy V S, Prodeus A N, Ladoshko O N and Samoylenko N A 2014 *Radioelektronika* vol 57 2 (620) 55-60
- [6] Gerasimenko V G, Lavrukhin Yu N and Tupota V I 2008 *Metody zashchity akusticheskoy rechevoy informatsii ot utechki po tekhnicheskim kanalam* (Moscow: Fakel) 258
- [7] Ivanov A V, Reva I L and Trushin V A 2010 (Novosibirsk: Novosibirsk State Technical University Press) 133–136
- [8] Ivanov A V, Trushin V A, Beresneva A V and Markelova G V 2014 The experimental research of security of speech in-formation of leakage from channels with technical account effect of forcing speech Actual problems of electronic instrument engineering 2014 12th International Conference on Actual Problems of Electronics Instrument Engineering (APEIE) 266-272
- [9] 1996 GOST R 50840-95 *Peredacha rechi po traktam svyazi. Metody otsenki kachestva, razborchivosti i uznavaemosti* (Moscow: Izdatel'stvo standartov) 198
- [10] Ivanov A V, Trushin V A and Khitsenko V E 2015 On the choice of the model of the test signal in the assessment of speech information security from leak on technical channels *Proc. of SPIIRAS* vol 40 122-133