Interspeech 2018
2-6 September 2018, Hyderabad

Effects of User Controlled Speech Rate on Intelligibility in Noisy Environments

John S. Novak, III [1], *Robert V. Kenyon* [1]

[1] University of Illinois at Chicago, United States of America

Jnovak5@uic.edu, kenyon@uic.edu

Abstract

Talkers intentionally producing high-intelligibility speech for listeners in challenging situations often reduce their speech rate. This study affords listeners fine-grained control over the playback rate of a desired speech signal in varying levels of background noise, and tests listener intelligibility with their preferred and unmodified rates of speech. We find clear listener preference for decreased rates of speech as background noise increased. However, we also found degraded performance on a speech-in-noise intelligibility test relative to unmodified speech in these same conditions.

Index Terms: perception of prosody; adverse listening conditions; speech intelligibility

1. Introduction

Improving the intelligibility of speech is usually the purview of the talker, with the listener as a passive agent. A great deal of work has examined how listeners process different characteristics of a talker's speech behavior and how such changes impact intelligibility and other factors. For example, in noisy or adverse listening environments, talkers adapt their speech (ostensibly to improve intelligibility) in a number of ways: reducing speech rate, changing pitch, changing formant patterns, and increasing consonant-vowel energy ratios [1], [2]. Such speech adaptations are also found when speaking to non-native speakers [3], those with a hearing-impairment [1], and when talking to infants [4, 5]. However, little has been done to give control to *listeners* to improve intelligibility other than amplification and, more recently, noise cancellation. Hearing aids, e.g., allow listeners to change filter and amplifier settings. However, more may be possible given advances in electronic processing power [6]. Giving listeners control of the characteristics of incoming audio signals may open a new avenue to improve intelligibility in noisy and other adverse conditions. To provide such affordances we must know what changes in the auditory signal can assist a listener.

Due to the prevalence of talker *speech rate* adaptation in these challenging environments, prior work examined the effects of artificially modified speech rates on intelligibility [7, 8], often using a small number of experimenter-chosen expansion factors applied uniformly across test subjects. Both experiments showed decreases in intelligibility with time expansion. However, related experiments featuring user control or pacing in noiseless environments have shown improvement in comprehension and listener recall [9, 10]. These studies found positive effects.

Given these improvements when listeners control the auditory speech rate [9] or pacing [10], we hypothesize that a similar paradigm applied to speech rate manipulation might improve intelligibility in a noisy environment, and have designed a study, which is the subject of this communication.

2. Materials

We adapted QuickSIN [11, 12] speech tests each consisting of a single list of six recorded Harvard [13] sentences, spoken in a conversational manner by a single female talker ("signal"), in the presence of four-person babble ("noise"). Each sentence contains five keywords, whose successful recitation forms the basis of the QuickSIN test score. We altered the standard test by altering the noise levels to include 0, 2.5, 5, 6.7, 8.3, 10, 12.5, 15, and 20 dB. (New values underlined, 25 dB removed.)

The QuickSIN tracks were separated into twelve signal and twelve noise tracks, and converted to 16 bit, 44.1 KHz sampled WAV format sound files, and are referred to as Lists 1 through 12. One non-separable track was used as a "practice" list (List 13). Finally, one audio segment of an actor reciting the Gettysburg Address [14] was converted to the same format. All WAV files were normalized with a sound level meter for sound intensity.

Software was written to allow subjects to manually control the rate of audio playback in real time, using a frame-based, magnitude-interpolating phase vocoder [15] as described in [6]. Subjects controlled audio playback rate with an on-screen slider bar, which could be dragged quickly with a computer mouse. The subjects controlled the dilation ratio of the audio signal (defined as the ratio of the unmodified to modified length of an audio track) which was linearly related to the movement of the on-screen slider bar. Subsequently these data were transformed into expansion ratio (1/dilation ratio) to maintain consistency with conventional analyses in audio research. Therefore, subjects adjusted the expansion factor from 1.0 to 2.5 in non-uniform increments of not larger than 0.06. Audio playback rate responded smoothly, in real time. The time stretching technique did not change the pitch or tonal qualities of the audio. In all cases, the slider controlled the playback speed of the *signal tracks*, while in all but one case, the speed of the *noise tracks* was unchanged. The software set Signal to Noise Ratio (SNR) values by keeping the signal constant and changing the intensity of the noise tracks.

3. Methods

3.1. Participants

Twenty-eight young (age 18 - 30), healthy (self-reporting no hearing problems), native English speaking adults took part in this study. Participants were entered into a raffle for a $50 Amazon Gift Card, which has since been disbursed. This study was approved by the UIC Office for the Protection of Research Subjects. Participants were recruited by announcements to UIC student mailing lists and in lectures, and provided written statements of informed consent prior to participation. Participants were informed of their right to halt participation at any time, without removal from the raffle.

10.21437/Interspeech.2018-63

3.2. Procedures

Custom software was installed on a Windows laptop, connected to Sennheiser HD 598SE over-ear headphones, with audio track sound levels calibrated to present signal audio at 65 dB SPL. Participants then engaged in the four-phase experiment described below, followed by subject interviews. The experiment took approximately 45 minutes per subject.

Training Phase: Subjects trained with the interface, controlling the expansion of a 183 second (unmodified) clip of a recitation of the Gettysburg Address with a horizontal slider.

Practice Phase: Sentences in this phase were presented with audible four-person babble noise, with a fixed signal to noise ratio of 10 dB. Subjects listened to as many sentences as necessary to be able to remember and distinguish the talker's voice, and to understand the sentence lengths (3 to 5 seconds.)

Personalization Phase: Subjects listened to QuickSIN Lists 1 through 9, each containing six sentences. Each list was presented at a different SNR level, as above. List, noise condition, and sentence orders within each list were randomized. At the beginning of each list, the initial position of the slider was randomized to prevent historicity. Note that each noise condition is confined to a single list, previously shown to be of equivalent difficulty. [16]

Subjects set the expansion to the speed deemed most useful for understanding the target speech, with no guidance as to what settings might be "useful." Subjects were asked to leave the slider in its most useful location before proceeding to the next group of sentences. These final settings were recorded as the subjects' personalized Preferred Speech Rates (PSR).

Evaluation Phase: Eighteen sentences in noise were drawn from QuickSIN lists as follows: Five each from Lists 10 and 11 were used for five standard QuickSIN SNR values, in modified and unmodified conditions respectively. The remaining sentences from those lists were combined with those of List 12 to extend the QuickSIN test to four non-standard SNR values. This is summarized in Table 1, below.

Table 1: *Allocation of List Sentences*

SNR, dB	Modified	Unmodified	Notes
0, 5, 10, 15, 20	List 10	List 11	Standard
2.5, 6.7, 8.3, 12.5	List 10, 12	List 11, 12	Non-standard

These sentences were presented in random order, and after each sentence the subject immediately repeated the sentence back to the researcher. All keyword responses were transcribed as they were spoken or noted if not spoken.

Subject Interviews: After each experiment, subjects were interviewed and asked whether they believed the overall technique of time expansion helpful, harmful, both or neither; whether they had adopted any strategies or patterns of use; and whether they had specific improvements to suggest.

3.3. Experimental Records

Electronic Records: In all phases, the software logged subjects' activity, including wherever applicable the identities of signal and noise tracks, the SNR of combined audio, and all expansion factors.

Scoring Records: QuickSIN test scores are calculated based on the number of keywords correctly recited back to the test administrator. Subjects' recitations were transcribed, except where subjects omitted keywords entirely.

3.4. Analysis Techniques

In addition to PSR and keyword error counts, we used two specialized measures: A modified QuickSIN SNR-Loss, measuring overall intelligibility across five SNR settings; and glimpse increase, expressing the amount of hypothetical perceptual benefit provided by time stretching. [20]

SNR-Loss: QuickSIN intelligibility is reported as an SNR-Loss score, i.e., the difference between the subjects' SNR-50 and that of a theoretical young healthy individual. SNR-50 is the SNR at which a subject understands 50% of an utterance. In a standard test, an SNR-Loss is calculated from the number N of correctly repeated keywords across all six sentences (i.e., SNR-Loss ≡ 25.5 – N) using the Tillman-Olsen method [17].

During the Test Phase, subjects performed two modified and interleaved QuickSIN tests. We altered these tests, inserting additional SNR conditions to better probe the more challenging, lower signal to noise ratio region. The complete set of SNR values is as follows: 0, 2.5, 5, 6.7, 8.3, 10, 12.5, 15, and 20 dB. (New values underlined.) The standard value of 25 dB was removed since we expected very little change from the 20 to 25 dB conditions in a young, healthy population. [18]

We derive an SNR-Loss formula for our altered test following the procedures in [19]. These procedures require equally spaced steps of SNR, a condition met only by the non-underlined values. The SNR-Loss scores reported herein are calculated only as above, using the restricted, equally spaced five-point data set. Therefore, only keywords corresponding to sentences from this restricted set were used in this calculation (i.e., SNR-Loss ≡ 20.5 – N). As described in Table 1 above, the SNR-Loss calculation for the modified and unmodified conditions use sentences from Lists 10 and 11 respectively, thus confining calculations to within single lists shown previously to be of equivalent difficulty. [16]

Glimpsing Data: We analyzed audio expansion's effects from a psychoacoustic standpoint using glimpse patterns [20]. We generated cochleagrams [20], [22] from unmodified test sentence and noise tracks by dividing these sound files into 58 uneven bands corresponding to cochlear frequency sensitivity from 50 to 7500 Hz with an integration time of 8 ms and frame length of 10 ms. We aligned each pair of signal and noise pattern against each other before combining, to match the randomized stimuli heard by subjects. Glimpses were defined as time-frequency cells where signal exceeded noise by 3 dB or more. Glimpse areas (GAs) were the total number of such cells in each pair of signal and noise tracks.

Second, each unmodified test sentence was fed to SPPAS automated speech segmentation software [23] with a transcript, which returned segmentation data for the words of each sentence, identifying onset/offset silences. These were used to generate masks for the glimpse patterns, including whole-utterance masks (excluding offset and onset silences) and keyword-segmented masks (including only keywords.)

Third, audio files for the modified sentences heard by subjects during their Evaluation Phase were generated from the combination of signal and noise tracks, according to their discovered PSRs. Corresponding segmentations and masks for these expanded sentences were generated by directly expanding the SPPAS results for the unmodified sentences. (SPPAS is not designed for use on stretched speech.)

Finally, as above, glimpse patterns and masks were generated by aligning the modified sentences with noise tracks as heard by the subjects, and noting time-frequency cells where the signal exceeds noise by 3 dB.

4. Results

4.1. Preferred speech rates

PSRs were determined for each SNR. A Lilliefors test failed to show normality. Therefore, median values are presented (Figure 1) with 95% confidence interval estimates. Medians ranged from 1.05 at 20 dB SNR to 1.50 at 0 dB SNR. We note a trend of increasingly large confidence intervals with increasing noise, with median values increasing nearly linearly ($R^2 = 0.91$) as SNR declines.

A Kruskal-Wallis non-parametric test was performed ($\chi^2 = 36.97$, df = 8, p = 1.17e-05), followed by Dunn's test with Benjamini-Hochberg post hoc correction, controlling False Discovery Rate to 0.05. The results of the Dunn's test, shown Table 2, analysis show that the expansion factors are statistically different at opposite ends of the SNR range, with results for all of the SNRs 0 dB, 2.5 dB and 5 dB differing ($p < 0.05$) from all of the SNRs 12.5 dB, 15 dB and 20 dB.

Table 2: *Dunn's Test of Preferred Speech Rates*

SNR	2.5	5	6.7	8.3	10	12.5	15	20
0	0.683	0.777	0.212	0.261	0.111	**0.009**	**0.004**	**0.001**
2.5	-	0.863	0.399	0.497	0.239	**0.023**	**0.015**	**0.002**
5		-	0.334	0.399	0.197	**0.016**	**0.010**	**0.001**
6.7			-	0.863	0.749	0.197	0.134	**0.021**
8.3				-	0.648	0.150	0.100	**0.015**
10					-	0.334	0.239	0.061
12.5						-	0.842	0.391
15							-	0.497

4.2. Intelligibility

We found intelligibility degraded significantly from 8.3 dB to 2.5 dB SNR ($p < 0.05$). We present the median change in score from the unmodified to modified conditions, for each SNR, shown Figure 1.

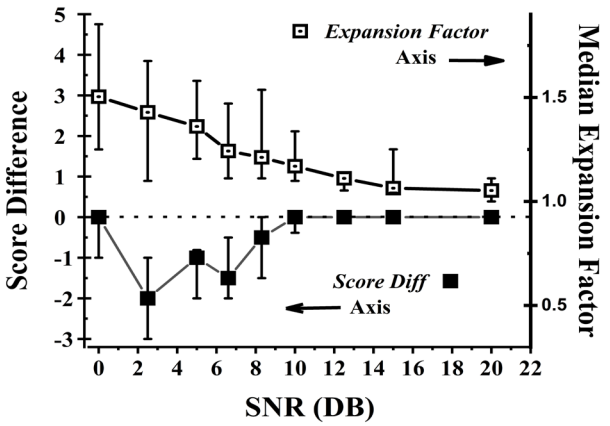

Figure 1: *Score Difference, Expansion vs SNR*

As before, the Lilliefors test failed to show normality. A sequence of one-sided Wilcoxon signed rank tests comparing words scored correct in modified vs unmodified conditions shows significant degradation in intelligibility ($p < 0.05$) at all SNRs from 0 dB through 8.3 dB.

We also examined overall SNR-Loss scores, as described above. SNR-Loss scores were calculated from raw data. The Lilliefors test indicated normality for the SNR-Loss scores. SNR-Loss worsened from a loss of 1.7 dB in the unstretched condition to 3.9 dB in the stretched condition.

4.3. Glimpse increase

Our analysis shows that an expanded sentence yields greater GAs on average, as measured by the glimpse increase ratio (GIR), i.e., ratio of the GA of a time-expanded sentence to a non-expanded sentence.

Expanding an utterance requires the creation of audio frames resulting in a longer track. Therefore, there are more time-frequency signal cells to compare to noise audio in the glimpse analysis, with the same statistical distribution as the original audio. On average this leads to a greater GA, and a greater GIR as temporal expansion increases.

However, our experiment makes direct comparisons between individual sentences difficult. First, to prevent subject learning effects, different sentences of differing lengths were used for test and control conditions. (Evaluation sentences range from 2.05 to 3.39 seconds.) Second, noise track order was randomized. Third, signal tracks contain leading and trailing silences, which are stretched during the Evaluation Phase, but discarded in the glimpse analysis. This changes the position of the voiced portion of a signal track relative to the noise track. Simulations indicate that GA and GIR are sensitive to both the second and third factors.

Since GIRs are highly variable, we grouped all GIRs according to their SNR conditions, and computed the average GIR by condition. This mean GIR is plotted against the mean expansion factor for each SNR along with a reference line of unity slope, shown Figure 2. This process was performed using both the sentence-level mask and the keyword-level mask. We constructed linear regression models, estimating the slope of the lines as 0.51 ($R^2 = 0.773$) and 0.65 ($R^2 = 0.666$), respectively. Both regressions reject the null hypothesis of zero slope ($p < 0.05$). The sentence-level regression rejects the hypothesis of unity slope ($p < 0.05$) while the keyword-only regression does not ($p > 0.05$). These analyses show that GAs increase with increasing temporal expansion; however, at the level of a whole utterance this increase is less than unity.

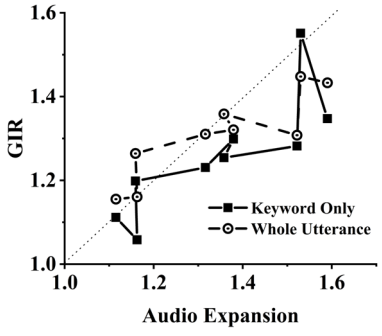

Figure 2: *Audio vs Glimpse Increase Ratio (GIR)*

4.4. Interview data

Post-experiment interviews revealed that 18 respondents (64%) expressed a full or qualified belief that expansion was helpful.

"Qualified belief" includes variations (paraphrased) such as "helpful at some noise levels," or "sometimes helpful, sometimes harmful." Six subjects (21%) expressed no opinion, or reported that expansion was neither helpful nor harmful. Four subjects (14%) reported that the technique was harmful. Three of the four subjects who reported the technique was harmful set the slider to no expansion for all noise conditions.

When asked about their usage, 15 subjects (54%) volunteered that they employed more stretching when they perceived more noise. Of these, 14 subjects were part of the subset who believed that stretching was helpful; the other thought stretching was neither helpful nor harmful. No subjects reported more expansion with decreasing noise levels.

Three subjects offered that stretched speech, especially highly stretched speech was perceptually odd or deficient ("unnatural," "not normal speech", words running together).

Finally, seven subjects referred to the ability to track a particular voice, and/or distinguished between that and the ability to understand the words as spoken by that voice. Three subjects thought slower voices were harder to track (one of whom thought expansion was generally harmful, two of whom thought expansion was neither helpful nor harmful) while three thought slower voices were easier to track (two of whom thought the expansion was helpful, one of whom thought expansion neither helped nor hindered.)

5. Discussion

Our experimental results are similar to previous investigations [8], namely, that expansion of a speech signal embedded in babble noise does not improve but *degrades* intelligibility. This degradation is concentrated at low SNR conditions where intelligibility is already degraded, but also where expansion was predicted to aid in intelligibility. However, in contrast to previous work where expansion values were imposed on subjects, this degradation occurs despite allowing subjects to choose expansions they feel benefit them most. Thus, given the ability to choose expansions, a majority chose to expand low SNR conditions despite degradation in their performance. It might be that subjects were using a different criteria or performance metric than intelligibility.

One possible explanation lies in our glimpse results, showing that time expansion is accompanied by increased GAs in whole utterances and keyword only portions. Although this glimpse increase did not increase intelligibility, it may enhance ability to track the target voice through the underlying noise as seven subjects (25%) noted. Of those seven, three believed expansion helped to isolate a voice, and was found to be non-harmful to intelligibility. Conversely, three believed expansion harmed the ability to isolate a voice believed that expansion was found to be non-helpful to intelligibility. Perhaps the advantage is that an expanded target voice in noise can be tracked more easily, explaining why most subjects used an expanded audio target signal as the noise level increased. If true, however, this ability to track a voice does not increase intelligibility.

However, the root cause of intelligibility degradation under time expansion is not clear. In [24] several potential explanations are summarized including algorithmically induced artifacts or distortions, which they discount on the strength of modern algorithms; that expansion factors larger than 1.4 may cause degradation by stretching syllables beyond a psychoacoustic "perceptual window"; and that the greatest benefit may be found in situations of cognitive load. Vocoder-based time-stretching algorithms such as the technique employed here are also introduce perceptual artifacts such as "phasiness" [26] and transient smearing which may reduce the psychoacoustic benefits of temporal expansion. Analyses of whole utterances and of the keywords show clear increases in GAs when signal tracks are stretched, indicating an overall increase in receivable signal to the listeners without signal amplification. However, we do note that the glimpse increase as measured across complete utterances increases more slowly than does the temporal expansion. This may be the result of the vocoding algorithm, and might be remedied with other more advanced algorithms if implemented in real-time.

Another possibility is that uniform expansion of speech is not sufficiently faithful to natural slowed speech. Various sources note differences in expansion ratios by phoneme [1,2,24,27]. This is tentatively supported by experiments in [26] which test non-uniform time expansion of conversational speech intended to mimic clear speech. While this non-uniform expansion degraded intelligibility, this degradation was much less than that caused by uniform time expansion. In [8], speech signals were expanded based on a local power threshold, with the intent of stretching only vowels. Following this non-uniform stretching, additional distortion was added to simulate hearing loss, with mixed results: time stretching effects on intelligibility were not significant for simulated hearing loss, however, for simulated hearing loss with amplification the effect of time stretching was significant and harmful. However, as the authors note, while this non-uniform stretching tended to stretch vowels rather than consonants, the overall effect was "unnatural" and did not match the cadence of naturally produced slow speech.

The intelligibility degradation may be only one effect of expanded audio. The subjects' report of ability to track a stretched voice in noise may be of significant value: the three subjects for whom tracking a voice was easier with increased expansion performed below average; however, the three subjects who found tracking more difficult used less expansion performed better. While this small subject pool does not allow statistical analysis, it does highlight a possible trade-off between tracking and intelligibility that may occur with audio expansion. While the intelligibility of an anomalous string of words may be impeded by expansion, tracking a voice in a conversational environment may have additional benefit to the listener. If a natural voice is lost in environmental noise and a stretched voice is not, then even the distortions produced by expansion may be overcome by the subjects' ability to choose the correct word within the context of the sentence and the conversation. It would be interesting to evaluate this condition in future experiments.

6. Conclusions

This study gave listeners control of auditory signals, showing that individual preferences exist; increased noise results in more expansion; and listeners perceived intelligibility improvement. We found a statistically significant degradation in intelligibility even with listener chosen conditions.

We believe these results show the difficulties associated with listener control of personal acoustic experience. Especially, we believe that while some subjects may be conflating the ease of isolating or tracking a voice through noise with the intelligibility of such a tracked voice, this distinction may be a fruitful research direction for related applications such as cognitive load and user comfort.

7. References

[1] Picheny MA, Durlach NI, Braida LD. Speaking Clearly for the Hard of Hearing II: Acoustic Characteristics of Clear and Conversational Speech. *Journal of Speech, Language, and Hearing Research.* 1986 Dec 1;29(4):434-46.

[2] Garnier M, Henrich N. Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise? *Computer Speech & Language.* 2014 Mar 31;28(2):580-97.

[3] Papoušek M, Hwang SF. Tone and intonation in Mandarin babytalk to presyllabic infants: Comparison with registers of adult conversation and foreign language instruction. *Applied Psycholinguistics.* 1991 Dec 1;12(04):481-504.

[4] Van de Weijer J. Language input to a prelingual infant. In the GALA'97 *Conference on Language Acquisition.* 1997 (pp. 290-293). Edinburgh University Press.

[5] Fernald A, Simon T. Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology.* 1984 Jan;20(1):104.

[6] Novak III JS, Tandon A, Leigh J, Kenyon RV. Networked on-line audio dilation. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* 2014 Sep 13 (pp. 255-258). ACM.

[7] Picheny MA, Durlach NI, Braida LD. Speaking Clearly for the Hard of Hearing III: An Attempt to Determine the Contribution of Speaking Rate to Differences in Intelligibility between Clear and Conversational Speech. *Journal of Speech, Language, and Hearing Research.* 1989 Sep 1;32(3):600-3.

[8] Nejime Y, Moore BC. Evaluation of the effect of speech-rate slowing on speech intelligibility in noise using a simulation of cochlear hearing loss. *The Journal of the Acoustical Society of America.* 1998 Jan;103(1):572-6.

[9] Zhao Y. The effects of listeners' control of speech rate on second language comprehension. *Applied Linguistics.* 1997 Mar 1;18(1):49-68.

[10] Piquado T, Benichov JI, Brownell H, Wingfield A. The hidden effect of hearing acuity on speech recall, and compensatory effects of self-paced listening. *International Journal of Audiology.* 2012 Aug 1;51(8):576-83.

[11] McArdle RA, Wilson RH. Homogeneity of the 18 QuickSIN™ lists. *Journal of the American Academy of Audiology.* 2006 Mar 1;17(3):157-67.

[12] Wilson RH, McArdle RA, Smith SL. An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. *Journal of Speech, Language, and Hearing Research.* 2007 Aug 1;50(4):844-56.

[13] Rothauser EH, Chapman WD, Guttman N, Nordby KS, Silbiger HR, Urbanek GE, Weinstock M. IEEE recommended practice for speech quality measurements. *IEEE Trans. Audio Electroacoust.* 1969 Sep;17(3):225-46.

[14] A Reading of the Gettysburg Address [Internet]. NPR.org. 2016 [cited 14 December 2016]. Available from: http://www.npr.org/templates/story/story.php?storyId=1512410

[15] Ellis D. A Phase Vocoder in Matlab [Internet]. Labrosa.ee.columbia.edu. 2016 [cited 4 March 2018]. Available from: http://labrosa.ee.columbia.edu/matlab/pvoc/

[16] Killion, Mead C., et al. "Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners." *The Journal of the Acoustical Society of America* 116.4 (2004): 2395-2405.

[17] Tillman TW, Olsen WO. Speech audiometry. *Modern developments in Audiology*. 1973;2:37-74.

[18] Killion MC. New thinking on hearing in noise: A generalized articulation index. In *Seminars in Hearing* 2002 (Vol. 23, No. 01, pp. 057-076). Copyright© 2002 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA.

[19] Niquette P, Gudmundsen G, Killion M. QuickSIN Speech-in-Noise Test Version 1.3. Elk Grove Village, IL: Etymotic Research. 2001.

[20] Cooke M. A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America.* 2006 Mar 1;119(3):1562-73.

[21] Ma N. Cochleagram Representation of Sound [Internet]. Staffdcs.shef.ac.uk. 2016 [cited 4 March 2018]. Available from: http://staffwww.dcs.shef.ac.uk/people/N.Ma/resources/ratemap/

[22] Brown GJ, Cooke M. Computational auditory scene analysis. *Computer Speech & Language.* 1994 Oct 31;8(4):297-336.

[23] Bigi B, Hirst D. SPeech Phonetization Alignment and Syllabification (SPPAS): a tool for the automatic analysis of speech prosody. In *Speech Prosody* 2012 (pp. 1-4).

[24] Gygi B, Shafiro V. Spatial and temporal modifications of multitalker speech can improve speech perception in older adults. *Hearing Research*. 2014 Apr 30;310:76-86.

[25] Laroche J, Dolson M. Phase-vocoder: About this phasiness business. In *Applications of Signal Processing to Audio and Acoustics*, 1997. 1997 IEEE ASSP Workshop on 1997 Oct (pp. 4-pp). IEEE.

[26] Puckette M. Phase-locked vocoder. In *Applications of Signal Processing to Audio and Acoustics*, 1995., IEEE ASSP Workshop on 1995 Oct 15 (pp. 222-225). IEEE.

[27] Uchanski RM, Choi SS, Braida LD, Reed CM, Durlach NI. Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech, Language, and Hearing Research*. 1996 Jun 1;39(3):494-509.