Implication of speech level control in noise to sound quality judgement

Sara Akbarzadeh[1], Sungmin Lee[2], Satnam Singh[3], and Chin Tuan-Tan[4]
University of Texas at Dallas, Richardson, US
E-mail: ([1]sara.akbarzadeh, [2]sung.lee, [3]satnam.singh, [4]Chin-Tuan.Tan)@utdallas.edu

***Abstract*- Relative levels of speech and noise, which is signal-to-noise ratio (SNR), alone as a metric may not fully account how human perceives speech in noise or making judgement on the sound quality of the speech component. To date, the most common rationale in front-end processing of noisy speech in assistive hearing devices is to reduce "noise" (estimated) with a sole objective to improve the overall SNR. Absolute sound pressure level of speech in the remaining noise, which is necessary for listeners to anchor their perceptual judgement, is assumed to be restored by the subsequent dynamic range compression stage intended to compensate for the loudness recruitment in hearing impaired (HI). However, un-coordinated setting of thresholds that trigger the nonlinear processing in these two separate stages, amplify the remaining "noise" and/or distortion instead. This will confuse listener's judgement of sound quality and deviate from the usual perceptual trend as one would expect when more noise was present. In this study, both normal hearing (NH) and HI listeners were asked to rate the sound quality of noisy speech and noise reduced speech as they perceived. The result found that speech processed by noise reduction algorithms were lower in quality compared to original unprocessed speech in noise conditions. The outcomes also showed that sound quality judgement was dependent on both input SNR and absolute level of speech, with a greater weightage on the latter, across both NH and HI listeners. The outcome of this study potentially suggests that integrating the two separate processing stages into one will better match with the underlying mechanism in auditory reception of sound. Further work will attempt to identify settings of these two processing stages for a better speech reception in assistive hearing device users.**

I. Introduction

Auditory perception of noisy speech also relies on the level of speech above audition, than its relative level to the masking noise, which is signal-to-noise ratio (SNR) alone. The active mechanism in cochlear actually broadens the auditory filter bandwidths when the overall intensity of sounds increases, which also reduces the frequency selectivity. One would hypothesize that the noise at higher sound pressure level will have a greater masking effect, even when the speech is kept the same sensational level above noise. This nonlinear masking effect with increasing sound intensity may yield a perception of sound quality that cannot be fully accountable with SNR alone.

Most noise reduction technique in the front-end processing for audio instruments or hearing devices enhance speech signal in advantage of signal-to-noise ratio (SNR) regardless of the overall sound intensity. In addition, the nonlinear distortion introduced by the noise reduction technique may result in a higher 'perceivable noise' than the anticipated residual noise after reduction, which further add to the nonlinear growth of masking. The sound quality of the output speech may be perceived differently by listener at different overall sound intensities.

Kates [1] showed that the nonlinear distortions introduced by noise reduction have an adverse effect on speech quality perception, which may not directly relate to the nonlinearity in the ear. However, our previous works [2, 3] showed that perceived sound quality degrades faster with increasing amount of nonlinear distortion than linear distortion (for instance additive noise and linear spectral shaping) with NH listeners. Likewise, Gabrielsson et al [4] also reported that linear distortions were perceived as change in timbre and tone quality, but not necessary a drastic change in perceived sound quality. In [3], we also found that reduced frequency selectivity in the listener's ear could be reflected by their consistency in rating their perceived sound quality.

In this study, we would like to relate the effect of overall sound intensity and the nonlinear distortion to sound quality judgement, in attempt to establish a metric for noise reduction with an optimal auditory reception [5]. We will explain the consequence of nonlinear growth of masking in the ear, by examining sound quality perception of two different sets of noisy speech sentences at different SNRs. The sets of stimuli were separately generated when the speech sentences are at two different sound pressure levels (SPL), with and without noise reduction. Listeners were asked to perform the evaluation task twice to serve as a check for their consistency in rating sound quality as they perceived. We will present the details of the experiments and explain the results in the subsequent sections. Finally, we will attempt to draw the implications from the outcome of the study.

II. Methods

A. *Experimental setup*

Noise reduction algorithms primarily reduces the noise in noisy speech but it also introduces nonlinear distortion. The remaining residual noise and the nonlinear distortion will add variations to the stimuli for sound quality evaluation by the listeners. Speech sentences were set to two SPL levels (75 and 65 dB SPL) and noises were set to three SPL levels (75, 65 and 55 dB SPL); noisy speech was generated by adding them in all possible combinations. Speech sentence at 75 dB SPL and noise 65 dB SPL was considered a different listening condition than a speech sentence at 65 dB SPL and noise at 55 dB SPL, even though both these cases have the same SNR (i.e. 10 dB SNR). Likewise, speech and noise were both at 75 dB SPL was considered a different condition from speech and noise were both at 65 dB SPL.

B. Participants

Eight participants (4 females and 4 males) identified by self-reported normal hearing, participated in the experiment. Their age ranges from 19 to 27 years with a mean of 23.63 and standard deviation of 3.33 years.

Two cochlear implant recipients (1 male and 1 female) with profound hearing loss were also recruited to participate in this study. They were 65 and 61 years in age, respectively. All participants are Native American English speakers. Listeners were compensated for their participation. The study was approved by institutional review board of the University of Texas at Dallas.

C. Stimuli: Speech Sentences

Four speech sentences, each concatenated two short phrases spoken by a male and female were extracted from AzBios database [6], and they were set to 75 dB and 65 dB SPL. Together with two types of noise, namely cafeteria and babble noise, set at 75 dB, 65 dB and 55 dB SPL, we created 8 different listening conditions. Each speech sentence was added to noises to form a set of 24 noisy speech sentences for each noise type. Two classical noise reduction (NR) algorithms, namely Wiener Filtering Method (NR 1) [7] and Binary Masking Method (NR 2) [8] were chosen to reduce noise and introduce nonlinear distortions in some listening conditions. Each NR algorithm generated 24 noise reduced sentences (4 sentences * 2 speech levels * 3 noise levels) with the two types of noises. Lastly 8 original sentences (4 sentences * 2 speech levels) were included as listening conditions. Together with these 8 original speech sentences, 48 noisy sentences and 112 noise reduced sentences, we have a total of 168 sentences with the two noise types was generated as different listening conditions for sound quality evaluation and stored in a PC.

All sentences were played back to the listeners using a 24-bit Lynx 1 sound card via Sennheiser HD600 earphones with a line amplifier for normal hearing participants, and via frontal-speaker for hearing impaired participants.

D. Procedure

Participants were asked to rate the sound quality of stimuli as perceived by them. Each participant was asked to perform the task twice with the same set of stimuli presented in two different randomized orders for each of two trials. After each stimulus was presented, the program waited infinitely for the listener to rate the perceived sound quality. Next stimulus was presented only after participant had made the rating. Repetition of the last stimulus was given as an option to the participant. Participants were required to rate the perceived sound quality of the stimuli on a scale of 1 to 10 with 1 being the most distorted and 10 being the most natural. The scale was displayed as a bar with 10 tabs numbered from 1 to 10 on the monitor. Participants were asked to provide their perceived rating by clicking on the numbered tab using a mouse provided.

III. RESULTS

A. Normal hearing participants

INTER-TRIALS CORRELATION

Each NH participant rated the perceived sound quality of the stimuli in two separate randomized trial. The inter-trial correlation on the ratings of each participant was computed and tabulated in the Table 1.

Table 1. Inter-trial correlation coefficients for NH participants

Participants	1	2	3	4	5	6	7	8
Correlation	0.981	0.957	0.956	0.979	0.948	0.971	0.970	0.981

The ratings were highly correlated between trials for all participants which indicates that they are consistent in their rating regardless of the randomized order in the presentation. Average of the ratings between two trials computed for each participant is also highly correlated to the grand average across all participants and trials (r>0.9). All NH participants had rated the perceived sound quality consistently among themselves.

SPEECH QUALITY RATING

Figure 1 shows grand average perceived quality ratings of 8 NH participants in the six listening conditions with original speech sentence (with no noise added). With or without NR, the ratings in these listening conditions were above 9.

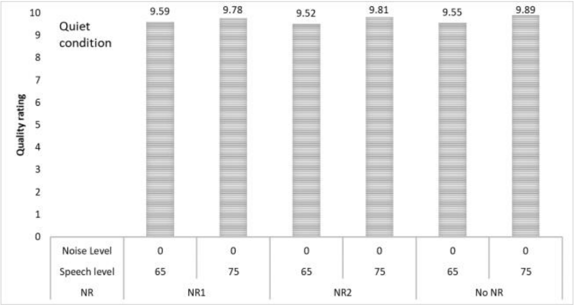

Figure 1. Grand average ratings for NH participants in quiet conditions.

Grand average perceived quality ratings of 8 NH participants in listening conditions with noisy speech sentences were shown in Figure 2; the top panel (A) was added with babble noise and the bottom panel (B) was added with cafeteria noise. The overall ratings in Figure 2 were lower than those in Figure 1. Likewise, at same level of speech level (65 or 75 dB SPL), the perceived quality degrades as the level of noise increases. This trend is observed with both babble noise and cafeteria noise.

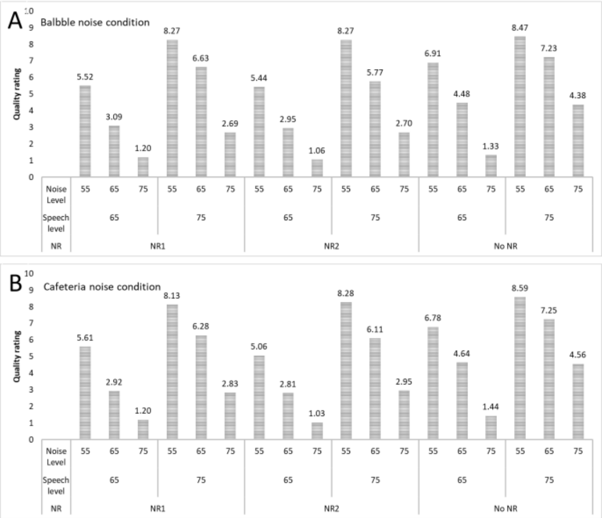

Figure 2. Grand average ratings for NH participants in babble and cafeteria noise.

Figure 3 compared only the grand average perceived quality ratings (extracted from Figure 2) at 10 dB SNRs with the speech level at 65 and 75 dB SPL. Two-way repeated measure ANOVA was performed separately babble and cafeteria noise, with different processing (No NR, NR1, and NR2) and speech levels (65 dB and 75 dB SPL) as two within subject factors.

For babble noise, the analysis showed a significant effect for the type of processing [$F(2,14)=8.217$, $p < 0.05$], but not the speech level [$F(1,7)=0.863$, $p=0.384$] with no interaction effect between two factors [$F(2,14)=0.375$, $p = 0.694$]. Pair-wise comparisons with Bonferroni adjustment showed that No NR and NR2 are significantly different ($p<0.05$). Similarly with cafeteria noise, the type of processing [$F(2,14)=6.447$, $p < 0.05$] has a significant effect, but not the speech level [$F(1,7)=3.069$, $p=0.123$] with no interaction effect between two factors [$F(2,14)=1.104$, $p = 0.359$]. Pair-wise comparisons with Bonferroni adjustment showed that No NR is significantly higher than NR 2 ($p<0.05$).

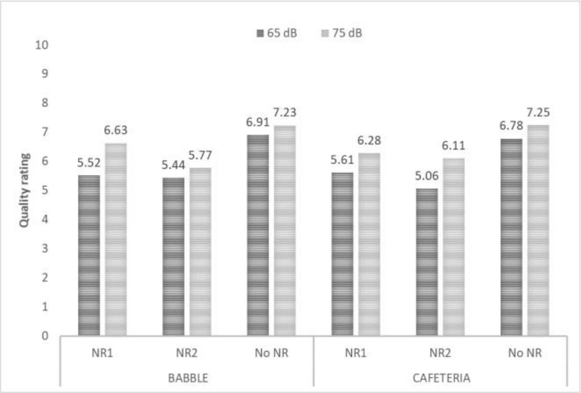

Figure 3. Grand average ratings at 10 dB SNR for NH participants, when speech level is at 65dB SPL (dark bar) and 75 dB SPL (grey bar)

B. *Cochlear Implant participant*

INTER-TRIAL CORRELATION

Likewise, each CI participant rated the perceived sound quality of the stimuli in two separate randomized trial. The inter-trial correlation on the ratings for each CI participant was computed; with 0.99 for CI participant 1 and 0.98 for CI participant 2, suggesting that they are consistent in their ratings.

SPEECH QUALITY RATING

Figure 4 shows grand average perceived quality ratings of 2 CI participants in listening conditions with original speech sentences. Ratings in all these condition were above 9.

Grand average perceived quality ratings of two CI participants in listening conditions with noisy speech sentences were shown in Figure 5; the top panel (A) was added with babble noise and the bottom panel (B) was added with cafeteria noise. The overall ratings in Figure 5 were lower than those in Figure 4. Likewise, at same level of speech level (65 or 75 dB SPL), the perceived quality degrades as the level of noise increases for all conditions except for few pairs of conditions (S65N65 dB < S65N75 dB with NR1 in babble noise, S65N65 dB < S65N75 dB with No NR in cafeteria noise (S=speech / N=noise). This trend is observed with both babble noise and cafeteria noise.

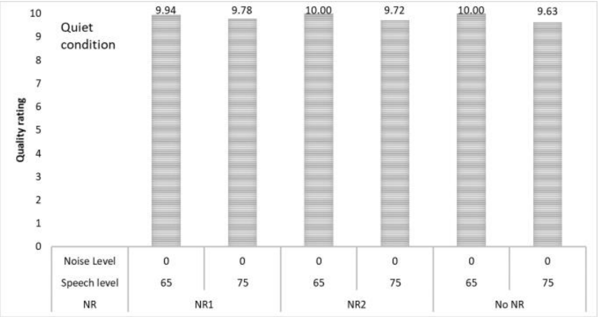

Figure 4. Grand average ratings in quiet conditions for CI participants.

Figure 6 compared only the grand average perceived quality ratings (extracted from Figure 5) at 10 dB SNRs with the speech level at 65 and 75 dB SPL. Contrary to the outcome with NH participants (Fig. 3), the perceived quality rating with speech level at 65 dB SPL was rated higher than with speech level at 75 dB SPL. Statistical analysis was not performed as there are only 2 CI listeners.

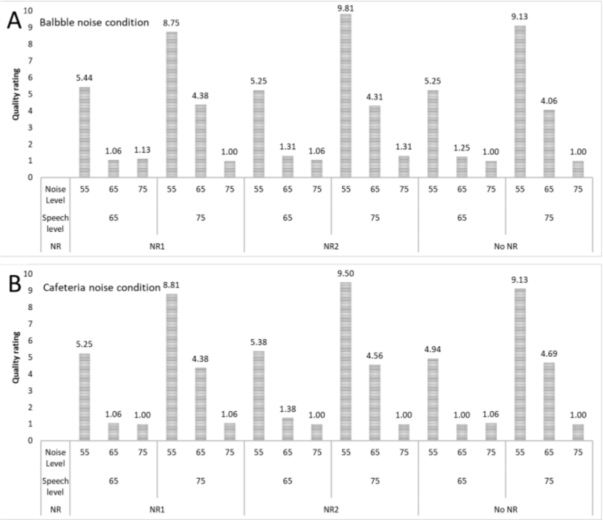

Figure 5. Grand average ratings in babble (A) and cafeteria (B) noise conditions for CI participants.

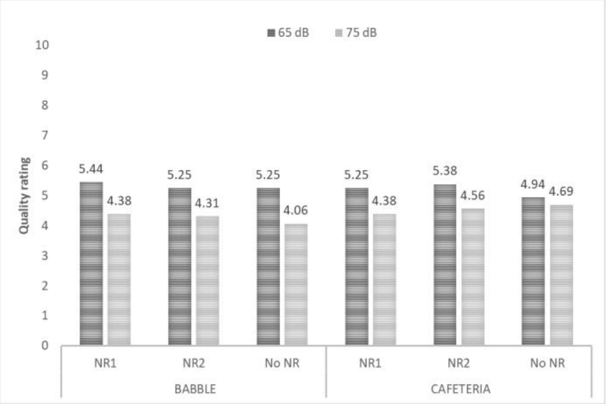

Figure 6. Grand average ratings obtained when presenting at 10 dB SNR for CI participants.

IV. Discussion

Our results showed that perceived speech quality of the original speech sentences (with no noise added) were rated highly in the range of 9.55 to 9.89 by NH participants and 9.63 to 10.0 by CI participants before or after NR processing, showing nearly equal ratings across conditions (Fig 1 & 4). In noise conditions for NH participants, however, the two NR algorithms chosen in this study influenced on the quality rating scores. No NR condition rated higher than either NR1 or NR2 condition, with the supporting statistic evidence showing the significant difference between No NR and NR 2. This implies the two NR algorithms introduced unnecessary distortion when they were processing in the noise environments.

In the outcome with NH participants, perceived quality ratings of noisy speech sentences were degraded systematically in a similar trend as noise is increasingly added. In the outcome with NH participants, this effect was observed in the perceived quality ratings of noisy speech sentences which degrades systematically in a similar trend as noise is increasingly added. Whether the speech level was kept at 65 dB SPL or 75 dB SPL, this trend remains with or without NRs. Notably, the overall perceived quality ratings were higher with speech level kept at 75dB SPL than those at 65dB SPL. However, the later observation was not statistically significant. During the sound quality evaluation, most NH participants have difficulty in rating the perceived quality of speech in the presence of noise. Participants might anchor on the whole sound level for making their ratings instead.

Another interesting observation was found in the perceived quality rating of noisy speech sentences by NH participants in the listening conditions, where noise is added at 10dB SNR with speech level at 65 dB SPL and 75 dB SPL. The effect of the 3 types of processing (No NR, NR1 and NR2) on the perceived quality ratings in these listening conditions were found to be statistically significant. When we examined their average perceived quality ratings only in these listening conditions (Fig 3), we found that the ratings with speech level at 75 dB SPL are higher than that at 65 dB SPL. However, when we examined average perceived quality ratings by CI participants in the same listening conditions as in Fig 3 (Fig 6), we found that the ratings with speech level at 75 dB SPL are lower than that at 65 dB SPL.

Previous studies [9, 10] showed speech recognition score increases with the level of speech, which is typically known as performance intensity (PI) function in audiometric assessment. This PI function is used to measure systematically for the growth of intelligibility with speech intensity. Assuming that perceived speech quality ratings are associated with speech recognition scores [11], trend of the perceived quality ratings can be seen as PI function in our study, particularly with NH participants. However, the trend established with current 8 participants was not statistically significant. More participants will be recruited to further validate the significance of the effect of speech level on perceived quality rating (Fig 3). For CI participants, an opposite trend of perceived quality ratings was observed. In a previous study on measuring tuning curve with cochlear implant users [12], Nelson et al. found that CI users exhibited bandwidths that were approximately five times wider than NH listeners, but were in the same range as HI listeners with moderate hearing loss. The wider bandwidths which are associated with lower frequency selectivity may support the preference of CI users for lower speech level. Hence, higher perceived quality of noisy speech at lower speech level by CI participants.

References

[1] J. M. Kates, "Hearing-aid design criteria," *J. Speech Lang. Path. And Audiology*, Monogr. Suppl. vol. 1, pp 15-23, Jan. 1993.

[2] B. C. J. Moore, and C. T. Tan, "Perceived naturalness of spectrally distorted speech and music." *J. Acoust. Soc. Am*. vol. 114, no. 1, pp 408-419, July, 2003.

[3] B. C. J. Moore, C. T. Tan, Z. Nick, and M Ville-Veikko, "Measuring and predicting the perceived quality of music and speech subjected to combined linear and nonlinear distortion." *J. Audio. Eng. Soc*. vol. 52, no. 12, pp 1228-1244, Dec. 2004.

[4] A. Gabrielsson, L. Björn, and T. Ove, "Loudspeaker frequency response and perceived sound quality." *J. Acoust. Soc. Am*. vol. 90, no. 2, pp 707-719, Aug. 1991.

[5] J. E. Preminger, and D. J. Van Tasell, "Quantifying the relation between speech quality and speech intelligibility." *J. Speech Hear. Res.* vol. 38, no. 3, pp 714-725, June, 1995.

[6] A. J. Spahr, M. F. Dorman, L. M. Litvak, S. Van Wie, R. H. Gifford, P. C. Loizou, L. M. Loiselle, T. Oakes, and S. Cook, "Development and validation of the AzBio sentence lists," *Ear Hear,* vol. 33, no. 1, pp 112–117, Jan-Feb, 2012.

[7] Y Hu, P. C. Loizou, "Incorporating a psychoacoustical model in frequency domain speech enhancement." *IEEE sig. proc. Let.* vol. 11, no. 2, pp 270-3, Feb. 2004.

[8] J Lim, A Oppenheim, "All-pole modeling of degraded speech." *IEEE Trans. on Acoust Speech Sig Proc.* vol. 26, no. 3, pp 197-210, Jun, 1978.

[9] J Jerger, S Jerger, "Diagnostic significance of PB word functions." *Arch Otolaryngol*. vol. 93, pp 573–580, 1971.

[10] A Boothroyd "The performance/intensity function: An underused resource." *Ear Hear*. vol. 29, pp 479–491, 2008.

[11] J E Preminger, D J Van Tasell, "Quantifying the relation between speech quality and speech intelligibility." *J Speech Lang. Hear Res*. vol. 38, no. 3, pp 714-25, Jun, 1995.

[12] D A Nelson, H A Kreft, E S Anderson, G S Donaldson, "Spatial tuning curves from apical, middle, and basal electrodes in cochlear implant users," *J. Acoust. Soc. Am.* vol. 129, pp 3916-3933. Jun, 2011.