Sentence_Identification_in_Noise.docx

by Author Author

Submission date: 21-Jul-2020 09:14AM (UTC+0530)

Submission ID: 1360248592

File name: Sentence_Identification_in_Noise.docx (53.04K)

Word count: 2084

Character count: 10880

Sentence Identification in Noise: Effect of Signal to Noise Ratio and Audio-Visual Modality

Background

Persons with hearing loss show poor performance in speech perception in noise, even with a hearing aid or cochlear implant. This is attributed to suprathreshold auditory skills which are usually indicated as distortion factor (Apeksha & Kumar, 2017; Phatak, Yoon, Gooler, & Allen, 2009). Speech perception in noise tests are valid tools in the assessment of the distortion factor(Phatak et al., 2009). Such standardised tests use a variety of stimuli ranging from monosyllables to sentences, presented with different types of noiseslike wideband noise and multi-talker babble.

In real life, listening to speech in the presence of noise is a common scenario. Several studiesthrow light on the effect of noise on masking speech signals(Miller, 1947; Miller & Nicely, 1955). Studies using smaller speech segments like consonants or vowels to understand speech perception in noise provided crucial information on the role of such segments in speech perception in degraded conditions(Gelfand, 2004).Sentences were also widely used as stimuli in literature. Most of such studies focused on the development and validation of clinical tests, or the utilisation of such tests to understand the effectiveness of management options like hearing aid as they provide more approximation to real-life situation (Theunissen, Swanepoel, & Hanekom, 2009).

Integration of information from multiple sensory modalities by the brain is vital in speech perception. Of the different modalities, the importance of visual information in perceiving speech during degraded conditions is thoroughly discussed in the literature(Balan & Maruthy, 2018; Kristin J. Van Engen, Zilong Xie, 2017; Lalonde & Werner, 2019). The variables related to the influence of both sentences and noise in visual speech perception havealso been studied using sentence identification in noise tasks (Helfer & Freyman, 2005). However, there is a dearth of such studies using sentences in Indian languages. Hence the present study aimed to check the effect of SNRs on sentence identification presented across modalities (i.e. Auditory only, Audio-visual, and Visual only).

Material and Methods

Participants

A total of 41 young individuals (9 males and 32 females) in the age range of 17-35 years (mean age of 22.07 years) participated in the study.Based on a structured interview, it was verified that none of them reported any difficulty in understanding speech in noise in daily listening conditions. All the participants' pure-tone air conduction thresholds were within 15 dB HL for all octave frequencies ranging from 250 Hz to 8 kHz in both the ears. The participants were native speakers of Kannada (a language spoken in the south Indian state of Karnataka) with a minimum educational qualification of secondary school examination. The visual acuity of the participants was normal or corrected (6/6). Written informed consent was obtained from all the participants before the participation and the method adhered to the ethical guidelines prescribed by the ethical committee for research at the [Institute name removed to ensure double-blind review](Venkatesan, 2009).

Stimuli/Material

Seven standardised sentence listsfromthe Kannada sentence lists developed by Geetha, Shivaraju, Kumar, Manjula, and Pavan (2014)were adopted for the study.In each sentence list, there were ten sentences, each with four keywords, hence, making 40 keywords per list. All the sentenceswere both audio and video recorded in a sound-treated, well-lit room from a native Kannada speaker with clinically normal speech. The video recording of the stimuli was done using a Sony HD professional video camera fixed on a tripod at zero degree azimuth focussed to speaker's face. Each stimulus was initiated and ended with a mouth closed position as neutral face. The speaker was instructed to utter each stimulus for at least three times with the same duration, natural intonation, clear pronunciation, least eye blinks and head movements. As to ensure good clarity for auditory portions of the stimuli while video recording, a simultaneous audio recording of stimuli was done using Adobe Audition (version 3) software with Motu Microbook II sound card interface with a sampling frequency of 44100 Hz. The good clarity auditory stimuli were digitised and normalised. The auditory stimuli were mixed with speech noise to generate 0 dB SNR, -5 dB SNR, and -10 dB SNR conditions using MATLAB 2014 (Mathworks Inc., Natick, MA, USA). The auditory stimuli with speech noise were synced with the video counterparts of the stimuli to create audio-visual (AV) condition. The audio signal was then extracted from the synched videotaped stimuli using Adobe Premiere Pro CC to create auditory only (AO), visual only (VO) conditions.

Procedure

The participants were made to sit in a well-lit, quiet room. The prepared stimuli stored in the Lenovo laptop (running on Windows 10 OS, Intel(R) i3-2370M CPU) were presented through calibratedheadphone (Sennheiser HD 569) at 75 dB SPL in all conditions except VO. The participants underwent sentence identification task at three SNRs (0dB, -5 dB, and 10 dB) across three modalities (AO, VO and AV conditions) resulting in seven conditions. The stimuli for each modality and the SNR conditions wererandomisedto avoid order effect using paradigm software (version 2.5.0.68). In each condition, separate lists of 10 sentences were used to avoid familiarisation effect. Hence, each participant had to repeat the sentence heard for 70 stimuli presentation (10 sentences * 7 conditions).After each sentence presentation, the participant had to repeat verbatim.The responses were voice recorded for offline scoring.

Data Analysis

Each keyword identified correctly in the sentence was given a score of 1 and each keyword wrongly identified was given a score of 0. The maximum score achievable in each stimulus condition was 40. The raw scores of each of the seven conditions were used to calculate the visual gain (VG) and auditory gain (AG) across the SNR conditions(Sumby & Pollack, 1954). The AG was calculated as the absolute difference between the raw scores of AV condition and VO condition[AG=AV-VO]. Similarly, VG was obtained from the absolute difference of raw scores for AV and AO conditions [VG=AV-AO].

The mean, median, and standard deviation of raw scores were estimated. The Shapiro-Wilk test was done on raw scores to check the normalcy of scores across conditions.The data were not normally distributed across each condition, and hence, non-parametric statistics were done. Friedman test was used to check the effect of modality and SNR on sentence identification. The data were further analysed using the Wilcoxon signed-rank test to check the pairwise significance.

Results

Descriptive statistics of sentence identification score, AG, and VG at three SNRs and across three modalities are given in table 1.The scores improved as the SNR became better in AO and AV modalities.

Table 1: Mean, median and standard deviation (SD) of sentence identification scores, auditory gain (AG), and visual gain (VG) at each SNR across modalities

Modality/ Gain	SNR	Mean	Median	SD
AO	-10 dB	3.88	2.00	4.595
	-5 dB	27.20	27.00	6.165
	0 dB	38.24	38.00	1.625
AV	-10 dB	17.22	16.00	8.572
	-5 dB	35.68	37.00	3.189
	0 dB	39.15	39.00	1.216
VO	Visual	4.41	4.00	3.486
AG	-10 dB	12.80	12.00	7.557
	-5 dB	31.27	32.00	3.735
	0 dB	34.73	35.00	3.529
VG	-10 dB	13.34	13.00	8.092
	-5 dB	8.49	8.00	5.192
	0 dB	.90	1.00	1.497

The sentence identification scores were compared between modality and SNR conditions using the Friedman test. There was a significant difference found between modality and SNR conditions($\Psi^2(6) = 234.98$, $p<0.01$);. The data were further analysed using the Wilcoxon signed-rank test to check the pairwise significance. The results are shown in table2. The sentence identification scores are significantly higher for 0dB SNR condition compared to -5dB and -10 dB SNR conditions in both AO and AV modalities. Similarly, a significantly higher score was obtained for -5dB SNR condition compared to -10dB SNR condition.

Table 2: Comparison of sentence identification score between different SNRs within each modality

Modality	**SNR**	**Z**	**p-value**
AO	-5 dB vs -10 dB	-5.58	0.00
	0 dB vs -10 dB	-5.59	0.00
	0 dB vs -5 dB	-5.58	0.00
AV	-5 dB vs -10 dB	-5.58	0.00
	0 dB vs -10 dB	-5.58	0.00

	0 dB vs -5 dB	-5.40	0.00

Table 3represents the comparison of the sentence identification score at each SNR across modalities. Sentence identification scores were significantly higher for AV conditions compared to AO at all SNRs. The scores were also significantly higher for AV compared to VO condition. The comparisons between AO and VO conditions showed that sentence identification scores were higher for AO compared to VO at all SNRs except at -10 dB SNR.

Table 3: Comparison of speech identification scores at each SNRs across modalities

SNR	**Modality**	**Z**	**p-value**
-10 dB	AO vs AV	-5.44	0.00
	AO vs VO	-1.43	0.15
	AV vs VO	-5.58	0.00
-5 dB	AO vs AV	-5.48	0.00
	AO vs VO	-5.58	0.00
	AV vs VO	-5.58	0.00
0 dB	AO vs AV	-3.28	0.00
	AO vs VO	-5.59	0.00
	AV vs VO	-5.59	0.00

Friedman test and Wilcoxon signed-rank test were done for AG and VG comparison also. Friedman test results revealed a significant difference across conditions ($\Psi^2(5) = 178.26$, $p<0.01$) and the results of the Wilcoxon signed-rank test is given in table 4. The AG showed a higher mean value in better SNRs compared to poorer SNRs. On the other hand, VG showed a reduction in mean value as the SNR improves. When a comparison (depicted in table 5) was made between AG and VG at each SNR, AG showed a significantly higher value compared to VG except at -10 dB SNR.

Table 4: Comparison of auditory gain and visual gain between different SNRs within each modality

Modality	**SNR**	**Z**	**p-value**
AG	-5 dB vs -10 dB	-5.58	0.00

	0 dB vs -10 dB	-5.58	0.00
	0 dB vs -5 dB	-5.40	0.00
	-5 dB vs -10 dB	-2.68	0.01
VG	0 dB vs -10 dB	-5.42	0.00
	0 dB vs -5 dB	-5.41	0.00

Table 5: Comparison of auditory gain and visual gain at each SNRs across modalities

SNR	Modality	Z	p-value
-10 dB	AG vs VG	-1.43	0.15
-5 dB	AG vs VG	-5.58	0.00
0 dB	AG vs VG	-5.59	0.00

Discussion

The sentence identification scores improve as the SNR improves in both AO and AV modes. At favourable SNRs, the best scores were observed in AV modality followed by AO and least in VO. As the noise reduces, the consonant confusions caused by masking noise was reduced and this resulted in an improvement in sentence identification scores. In AO condition, the growth was drastic (i.e. 3% becomes 38%).However, in the AV condition, the reduction of sentence identification scores is comparatively lower as the SNR worsens. This indicates the effectiveness ofutilising visual cues for perceiving speech in noise. Earlier studies have also shown that the dependency on visual cues increases as the auditory cues become more degraded(Munhall, Kroos, Jozan, & Vatikiotis-Bateson, 2004; Pilling & Thomas, 2011; Tye-Murray, Sommers, & Spehar, 2007a, 2007b).Sentence identification score in VO condition is weakest and almost comparable to the score in AV at -10dB SNR. This may be due to the adverse SNR; the listener could utilise only visual cues present in the speech signal.

Deduction of AG and VG provides vital information on the relative role of auditory and visual modality contribution to speech perception in the presence of noise. Auditory gain reduces as the SNR reduces, whereas, visual gain increaseswith SNR reduction. This was expected because when SNR worsens listeners depend more on visual cues to perceive speech effectively(Stacey, Kitterick, Morris, & Sumner, 2016). At -10dB SNR,AG and VG were comparable,and it may be because of the ceiling effect in visual cues utilisation. At favourable SNRs the VG is lesser, because dependence on visual cues is minimal; the cues fromauditory modality alone are enough to perceive speech in such situations. These findings on VG agree with the study byBalan and Maruthy (2018).

Conclusion

The findings of the current study serves the following purposes. Thevisual cues has a critical role in the speech perception especially in degraded listening conditions even in young normal hearing individuals. The AG and VG measures also strengthens this assumption. As the effectiveness of visual cue is proved in normal hearing individuals, speech perception training using AV modality would be considered for individuals with hearing disorders like ANSD or APD. However, more research on the efficacy of AV speech training is required for such conclusions.

Acknowledgments

The authors would like to thank the Directors of [Institute names removed to ensure double-blind review]for the support. The authorswould also like to extend their gratitude to all subjects who participated in the study.

Conflict of Interest: NIL

Source of Funding: Department of Science and Technology project fund.

Sentence_Identification_in_Noise.docx

ORIGINALITY REPORT

6%	2%	3%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Southampton Solent University Student Paper	1%
2	Submitted to All India Institute of Speech & Hearing Student Paper	1%
3	scholarsarchive.byu.edu Internet Source	1%
4	extrahealthy.co.uk Internet Source	1%
5	Divya Seth, Santosh Maruthy. "Effect of phonological and morphological factors on speech disfluencies of Kannada speaking preschool children who stutter", Journal of Fluency Disorders, 2019 Publication	1%
6	ebcj.mums.ac.ir Internet Source	<1%
7	Connie Keintz. "Utilization of visual information and listener strategies in intelligibility impairment	<1%

related to bilateral facial paresis", International Journal of Speech-Language Pathology, 2011
Publication

8	www.mdpi.com Internet Source	<1%
9	Submitted to University of Southampton Student Paper	<1%
10	Earl Johnson, Todd Ricketts, Benjamin Hornsby. "The effect of extending high-frequency bandwidth on the Acceptable Noise Level (ANL) of hearing-impaired listeners", International Journal of Audiology, 2009 Publication	<1%
11	Submitted to University of Canterbury Student Paper	<1%
12	Sin-Young Park, Hack-Youn Kim, Juhui Choe. "Application of an Electric Field Refrigeration System on Pork Loin during Dry Aging", Food Science of Animal Resources, 2019 Publication	<1%
13	Vijaya Kumar Narne, C. S. Vanaja. "Perception of speech with envelope enhancement in individuals with auditory neuropathy and simulated loss of temporal modulation processing", International Journal of Audiology, 2009 Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		