

EFFECT OF WITHIN- AND BETWEEN-TALKER VARIABILITY ON WORD IDENTIFICATION IN NOISE BY YOUNGER AND OLDER ADULTS

Huiwen Goy¹, Kathy Pichora-Fuller^{1,3}, Pascal van Lieshout^{1,2,3,4}, Gurjit Singh^{1,3} and Bruce Schneider¹

¹Dept. of Psychology, University of Toronto, Mississauga, 3359 Mississauga Rd North, Ontario, Canada L5L 1C6

²Dept. of Speech-Language Pathology, University of Toronto, 160-500 University Ave, Ontario, Canada M5G 1V7

³Toronto Rehabilitation Institute, 550 University Ave, Ontario, Canada M5G 2A2

⁴Institute of Biomaterials and Biomedical Engineering, University of Toronto, 164 College St, Ontario, Canada M5S 3G9

1. INTRODUCTION

Speech produced in noise differs from speech produced in quiet. The Lombard Effect is the reflex by which talkers increase their vocal effort as environmental noise increases [1]. Other speech differences include increased fundamental frequency (F_0) [2], word duration [3], and high-frequency energy [4]. Although the Lombard Effect is well-documented, standardized tests of speech intelligibility in noise usually use speech materials recorded in quiet, with background noise added later. However, word recognition in noise is significantly better for speech materials produced in noise than for materials produced in quiet, even when vocal level differences are removed [4].

Older adults, even if they have audiometric thresholds within clinically normal limits, often have more trouble than younger adults understanding speech in noise [5]. The intelligibility advantage seen when listeners are tested with speech produced in conditions matched to the listening condition may be even greater for older than for younger adults. The main purpose of the present study was to investigate age-related differences in the extent to which results of speech intelligibility testing might be improved if more ecologically appropriate spoken versions of materials were used. Another purpose was to examine the possible acoustical explanations for any differences in intelligibility.

2. METHOD

2.1 Participants

Participants were 16 younger adults (mean age = 19.9 years, $SD = 1.7$) and 16 older adults (mean age = 69.0 years, $SD = 4.1$). All participants spoke English as a first language, and had pure-tone audiometric thresholds of ≤ 25 dB HL at frequencies from .25 to 3 kHz in the test ear. The vocabulary scores on the Mill Hill test were 13/20 ($SD = 2$) for the younger and 16/20 ($SD = 3$) for the older group.

2.2 Stimuli

The stimuli used were the eight sentence lists of the SPIN-R test. Each list contains 25 low-context and 25 high-context sentences. List 1 of the original SPIN-R recordings was used [6]. A new male talker (DF) with similar F_0 and speaking rate as the original SPIN-R talker recorded the eight lists in different speaking environments. Based on measures for sentences in List 1, the average F_0 was 121.7 Hz for DF and 120.5 Hz for the original talker. The average speaking rate was 4.3 syllables/sec for both talkers. List 1

was spoken by DF in quiet without headphones, and Lists 2, 7 and 8 were spoken in quiet while DF wore headphones (Sennheiser HD265 Linear). DF was instructed to “speak normally” for Lists 1 and 2, to “speak clearly” for List 7, and to “speak loudly” for List 8. Lists 3 to 6 were spoken while DF heard noise presented over headphones: Lists 3 and 5 were spoken in speech spectrum noise at 62 and 66 dB SPL, respectively; Lists 4 and 6 were spoken in multi-talker babble at 62 and 66 dB SPL, respectively. All sentences were digitized at 24 kHz, low-pass filtered at 6 kHz, and matched on RMS level.

2.3 Experiment procedure

Each participant was tested on the eight lists (eight talking conditions), with list order counterbalanced. Sentences were presented monaurally under TDH-50P earphones at 50 dB SL (relative to participant’s average pure-tone threshold for 0.5, 1, and 2 kHz). Each sentence was presented with corresponding files of multi-talker babble at 0 dB S:N. Participants were instructed to report the last word of every sentence. There were no time constraints on responding, and guessing was encouraged.

2.4 Acoustical analysis procedure

Several acoustical measures were taken of the stimuli, using the PRAAT speech analysis program [7]. For every sentence, the time boundaries of target words and their phonemes were marked. A script was then used to generate measures of duration, F_0 , and intensity for each target word, as well as F_1 to F_3 for the vowel in the word.

3. RESULTS

3.1 Word identification

Figure 1 shows the performance of younger and older adults. In general, words were identified better in the high-context than low-context conditions. Younger adults correctly identified more words in noise than older adults, especially in the low-context condition. Both age groups found the new talker to be more intelligible than the original one. Word identification was best for talking conditions with louder noise environments (conditions 4 and 6), when the type of noise in the talking environment matched that of the listening environment (conditions 3 and 4), and when the talker was instructed to “speak loudly” (condition 8).

This description was confirmed by an Analysis of Variance (ANOVA) with age as a between-subjects factor, and talking condition and context as within-subjects factors.

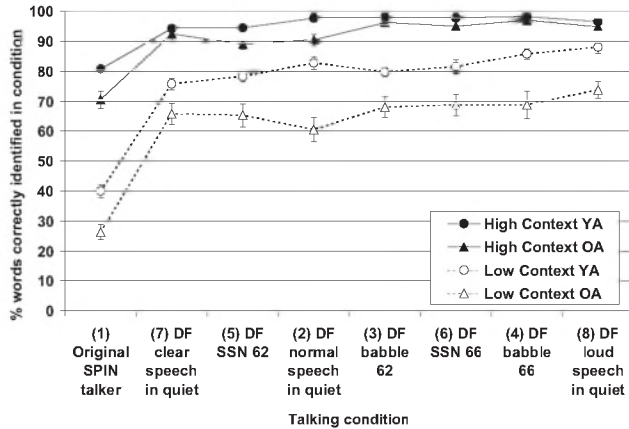


Fig. 1. Mean percent of correctly identified target words in 8 talking conditions, for 16 younger adults (YA) and 16 older adults (OA). Results are categorized into high- and low-context conditions. Standard error bars are shown.

There were main effects of age, $F(1,30) = 17.03, p < 0.001$, context, $F(1,30) = 531.02, p < 0.001$, and talking condition, $F(7,210) = 173.99, p < 0.001$. Performance in all conditions with the new talker was better than with the original talker, and performance in the set of conditions 3, 6, 4 and 8 was equivalent, but better than in all other conditions ($p = 0.05$). There were significant interactions of age and talking condition, $F(7, 210) = 3.260, p < 0.01$, and age and context, $F(1,30) = 24.83, p < 0.001$, with age-related differences being greater in the low-context condition ($p = 0.01$).

3.2 Talker differences

Compared to the original talker, DF's speech had more energy in the higher frequency range (Fig. 2). For DF, his speech in noise had more energy in the 2-5 kHz range than his speech in quiet. Clear speech had less energy in the 3-4 kHz, but more energy in the 4-5 kHz range, whereas loud speech was the opposite, with more energy in the 2-4 kHz and less in the 4-6 kHz range.

There were significant differences between lists for target word properties (Table 1), as confirmed by one-way ANOVAs for duration, $F(4, 120) = 3.76, p < 0.01$, intensity, $F(4, 120) = 11.58, p < 0.001$, and F_0 , $F(4, 120) = 14.23, p < 0.001$, but not for F_1, F_2 or F_3 . Tukey's *HSD* test showed that duration, F_0 , and intensity were not significantly different between the original talker and DF. Within-talker, there were no significant differences between Lists 2, 4, and 7, but word duration in Lists 4 and 7 was slightly longer than in List 2 ($p = 0.05$). List 8 words had significantly higher F_0 and intensity compared to other lists ($p = 0.01$).

Table 1. Acoustical measures of target words in five selected lists

Background	Quiet				Babble at 66 dB
	No	Yes			Yes
Headphones	Normal				Yes
Instructions	Normal	Clear	Loud	Normal	Normal
Talker & list	SPIN 1	DF 2	DF 7	DF 8	DF 4
Duration (ms)	488	521	559	522	547
Intensity (dB)	66.1	66.8	66.1	68.1	66.8
F_0 (Hz)	112	121	102	145	117
F_1 (Hz)	578	637	576	648	610
F_2 (Hz)	1437	1429	1529	1528	1516
F_3 (Hz)	2434	2493	2447	2524	2462

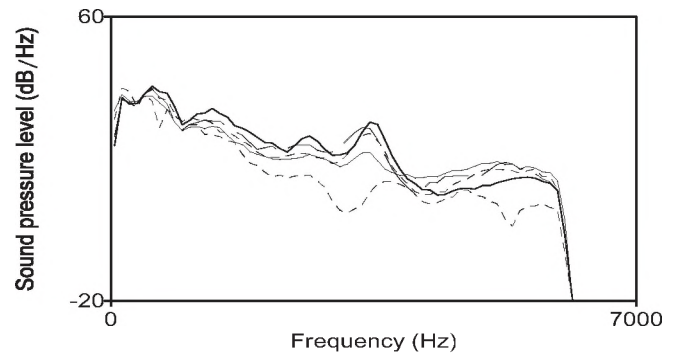


Fig. 2. Long-term average spectrum of target words spoken by the original talker (lower dotted line); DF in quiet (upper dotted line); DF in 66 dB SPL babble (dashed line); DF "speaking clearly" (non-bolded line) and DF "speaking loudly" (bolded line).

4. DISCUSSION

Our results support previous findings that speech produced in noise is more intelligible than speech produced in quiet, when tested in noise. For both age groups, accuracy was influenced by within- and between-talker differences, but older adults seemed to benefit more from speech produced in relevant conditions. Since all sentences were equated for overall intensity, the differences in intelligibility between lists may be due to properties of the target word itself. Specifically, there were changes in the intensity, duration, and F_0 of the target words. Although target words spoken by DF in quiet were more intelligible than those of the original talker, they did not differ significantly in duration, F_0 or intensity. However, as shown in Fig. 2, DF produced more high-frequency energy, which may account in part for the between-talker differences. For DF, words spoken in noise were more intelligible than words spoken in quiet, but they do not differ significantly in duration, F_0 or intensity. However, words spoken "loudly" had greater F_0 and intensity compared to words spoken "normally" in quiet, which may explain their greater intelligibility in noise. A combination of acoustic cues may be needed to explain the pattern of results.

REFERENCES

- [1] L. E. Lombard. (1911). Le signe de l'elevation de la voix. *Ann. Mal. Oreil. Larynx*, 37, 101-119.
- [2] T. Letowski, T. Frank, & J. Caravella. (1993). Acoustical properties of speech produced in noise presented through supra-aural earphones. *Ear & Hearing*, 14, 332-338.
- [3] J. Junqua. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *JASA*, 93, 510-524.
- [4] W. V. Summers, D. B. Pisoni, R. H. Bernacki, R. I. Pedlow & M. A. Stokes. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *JASA*, 84, 917-928.
- [5] M. K. Pichora-Fuller & P. Souza. (2003). Effects of aging on auditory processing of speech. *Int. J. Audiol*, 42, S11-S16.
- [6] R. Bilger, J. Nuetzel, W. Rabinowitz & C. Rzezckowski. (1984). Standardization of a test of speech perception in noise. *JSHR*, 27, 32-48.
- [7] P. Boersma and D. Weenik, "Praat: Doing phonetics by computer (version 4.5.08)" [Computer program], 2006 Apr 4 [cited 2007 Jul 22], Available HTTP: <http://www.praat.org/>

ACKNOWLEDGMENTS

This project was funded by CIHR and NSERC.