

# An analysis of the masking of speech by competing speech using self-report data (L)

Trevor R. Agus<sup>a)</sup> and Michael A. Akeroyd

*MRC Institute of Hearing Research (Scottish Section), Glasgow Royal Infirmary, Alexandra Parade, Glasgow G31 2ER, United Kingdom*

William Noble and Navjot Bhullar

*School of Behavioural, Cognitive and Social Sciences, University of New England, Armidale NSW 2351, Australia*

(Received 19 March 2008; revised 20 October 2008; accepted 21 October 2008)

Many of the items in the “Speech, Spatial, and Qualities of Hearing” scale questionnaire [S. Gatehouse and W. Noble, *Int. J. Audiol.* **43**, 85–99 (2004)] are concerned with speech understanding in a variety of backgrounds, both speech and nonspeech. To study if this self-report data reflected informational masking, previously collected data on 414 people were analyzed. The lowest scores (greatest difficulties) were found for the two items in which there were two speech targets, with successively higher scores for competing speech (six items), energetic masking (one item), and no masking (three items). The results suggest significant masking by competing speech in everyday listening situations. © 2009 Acoustical Society of America. [DOI: 10.1121/1.3025915]

PACS number(s): 43.71.Ky, 43.71.Lz, 43.66.Sr [RLF]

Pages: 23–26

## I. INTRODUCTION

In recent years there has been growing interest in the “informational masking” of speech (e.g., Freyman *et al.*, 1999; Brungart, 2001; Li *et al.*, 2004). This phenomenon is the extra masking observed when the identifiability of speech is measured in a background of competing speech versus that predicted from control conditions with comparable acoustics, such as a background of noise (cf. Carhart *et al.*, 1969). Informational masking is associated with uncertainty, target-masker similarity, and other “higher-level” aspects of the target and masker stimuli (cf. Durlach *et al.*, 2003). However, most of these measurements reflect situations where there are few cues available with which to distinguish the target and competing talkers. Laboratory measurements of the amount of informational masking of target speech by competing speech can be as much as 22 dB (Arbogast *et al.*, 2002), but is reduced considerably when supplementary cues are provided, including differences in talker gender (Brungart, 2001), pitch (Drullman and Bronkhorst, 2004), azimuth (Freyman *et al.*, 1999), and speech reading (Helfer and Freyman, 2005). Since many of these cues are typically available to most listeners in everyday listening situations, it seems unlikely that large informational-masking effects would be observed outside of the laboratory.

One approach to investigating the difficulties in real life due to competing speech is to actually ask participants about them. To this end, the items included in the “Speech, Spatial and Qualities of Hearing” scale (SSQ) (Gatehouse and Noble, 2004) are particularly useful because they ask about listening situations involving competing speech. The SSQ was designed to measure “auditory disability,” the restriction of auditory ability to perform an activity considered to be

normal for a human being, as distinct from “auditory impairment,” the underlying loss of auditory function (cf. World Health Organisation, 1980). The items cover a wide range of auditory abilities, including speech perception (e.g., “Can you easily have a conversation on the telephone?”), spatial hearing (e.g., “Can you tell how far a bus or truck is, by the sound?”), and “qualities” of hearing (e.g., “Do other peoples’ voices sound clear and natural?”). The questionnaire is designed to be completed by interview, with participants being read each vignette and then asked to respond on a visual scale from 0 (“not at all”) to 10 (“perfectly”). Thus, lower scores correspond to greater auditory disability.

Here, we reanalyze some previously collected SSQ data and focus on the 14 items that relate to speech perception: many of them refer to situations involving the comprehension of speech in competing speech, and so responses might be expected to indicate the total masking of competing speech and to help distinguish informational masking from energetic masking.

## II. METHOD

The data were collected as part of an earlier survey.<sup>1</sup> A short questionnaire was mailed to a random sample of 9000 individuals in the Glasgow area. The questionnaire used a set of screening questions including: (1) whether the individuals self-reported any hearing loss, the difficulties any hearing loss caused them, and whether they wore hearing aids, (2) a set of six of the SSQ items, with two from each of the speech, spatial, and qualities of hearing sections (Gatehouse and Noble, 2004), (3) what their opinions were about the provision of hearing aids by the UK National Health Service, (4) their age and their sex, and (5) their overall quality of life. The study was designed to stratify 50 people in each of the eight groups of a  $2 \times 2 \times 2$  factorial design: 50–64 years versus 65–80 years; no hearing difficulty versus hearing dif-

<sup>a)</sup>Electronic mail: trevor@ihr.gla.ac.uk

TABLE I. Mean and standard deviations of scored responses to speech items from the SSQ for those with and without self-reported hearing difficulties. Items are grouped by categorization then ordered by mean responses.

SSQ item	Vignette	Category	Mean (SD) hearing difficulties	
			Yes	No
10.	You are listening to someone talking to you, while at the same time trying to follow the news on TV. Can you follow what both people are saying?	Two targets	4.4 (2.4)	6.1 (2.5)
14.	You are listening to someone on the telephone and someone next to you starts talking. Can you follow what's being said by both speakers?	Two targets	4.8 (2.6)	6.2 (2.4)
6.	You are in a group of about five people in a busy restaurant. You cannot see everyone else in the group. Can you follow the conversation?	Multitalker masker	4.9 (2.4)	6.7 (2.3)
4.	You are in a group of about five people in a busy restaurant. You can see everyone else in the group. Can you follow the conversation?	Multitalker masker	5.4 (2.5)	7.3 (2.2)
11.	You are in conversation with one person in a room where there are many other people talking. Can you follow what the person you are talking to is saying?	Multitalker masker	5.7 (2.4)	7.3 (2.3)
8.	Can you have a conversation with someone when another person is speaking whose voice is the same pitch as the person you're talking to?	Single-talker masker	5.9 (2.4)	7.3 (2.2)
9.	Can you have a conversation with someone when another person is speaking whose voice is different in pitch from the person you're talking to?	Single-talker masker	6.0 (2.2)	7.4 (2.1)
1.	You are talking with one other person and there is a TV on in the same room. Without turning the TV down, can you follow what the person you're talking to says?	Single-talker masker	6.0 (2.4)	7.4 (2.2)
5.	You are talking with one other person. There is continuous background noise, such as a fan or running water. Can you follow what the person says?	Noise masker	6.4 (2.3)	7.8 (2.1)
3.	You are in a group of about five people, sitting round a table. It is an otherwise quiet place. You can see everyone else in the group. Can you follow the conversation?	No masker	6.8 (2.5)	8.2 (2.0)
13.	Can you easily have a conversation on the telephone?	No masker	7.9 (2.2)	8.8 (1.9)
2.	You are talking with one other person in a quiet, carpeted lounge-room. Can you follow what the other person says?	No masker	8.1 (1.9)	8.9 (1.8)
12.	You are with a group and the conversation switches from one person to another. Can you easily follow the conversation without missing the start of what each new speaker is saying?		5.4 (2.5)	6.9 (2.5)
7.	You are talking to someone in a place where there are a lot of echoes, such as a church or railway terminus building. Can you follow what the other person says?		5.6 (2.5)	7.5 (2.2)

difficulty; low score on the six-question form of the SSQ versus high score. There were 3824 respondents, of whom 1979 were 50–80 years old. Of these, 576 were selected for a face-to-face interview, to try to get 50 completed interviews per group. There, an interviewer conducted a hearing handicap questionnaire and the SSQ questionnaire (both are fully listed in Gatehouse and Noble, 2004), together with a small number of questions about the individual. The hearing difficulty was categorized by their yes/no response to the question “Do you have difficulty with your hearing?” In practice, 414 successful interviews were obtained, giving group sizes from 49 to 54. We used part of the stratification here, comparing the “younger” versus “older” groups ( $N=208, 206$ , respectively) and “no hearing difficulty” versus “hearing difficulty” groups ( $N=209, 205$ , respectively).<sup>2</sup>

For the present analysis, the speech items of the SSQ were split into six categories (see Table I) according to the configuration of target speech and competing speech described by the vignette. In four categories, the items asked about following one person in a variety of types of masker: no maskers (item Nos. 2, 3, and 13), a noise masker, interpreted as purely energetic (No. 5), single-talker maskers (Nos. 1, 8, and 9), and multitalker maskers (Nos. 4, 6, and 11). In a fifth category, the items asked about following two targets at the same time (Nos. 10 and 14). The sixth category contained two items that were omitted from the analysis because they were felt to be ambiguous about whether or not

the masker should be considered to be competing speech (Nos. 7 and 12).

### III. RESULTS

Table I reports the scores for each of the 14 speech items in the questionnaire separately for the normal-hearing and hearing-disabled participants. The items are ordered by mean score in the normal group (note that the ordering is the *same* for the hearing-disabled group). Figure 1 shows the data averaged across both groups, with items ordered in the same way: the highest mean responses (i.e., the least disability) were for the no-masker items and the lowest for the two-

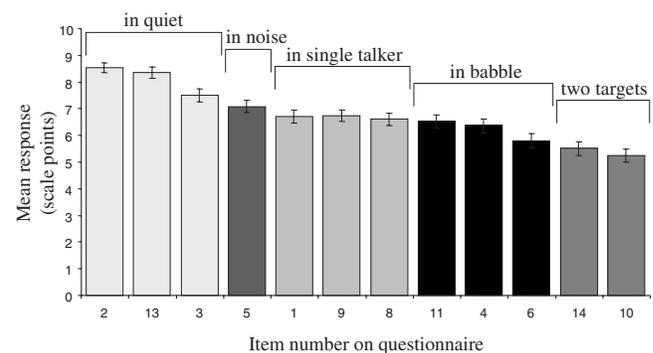


FIG. 1. Mean scores (out of 10) for each of the SSQ items included in the analysis, reverse-ordered by mean score, and grouped by categorization. Error bars are 95% confidence intervals.

target items. The mean responses for the other items fell between those two in an ordered manner, from noise-masker items, single-talker items, to multitalker items. Note that there was no overlap between the five categories of items.

The score for each item was lower for the hearing-disabled listeners than for the normal-hearing listeners by, on average, 1.4 scale points.<sup>3</sup> Accordingly, to compare the effects for these groups we calculated a within-listener “single-talker effect,” “multitalker effect,” and “double-target effect,” as the difference between the score for the respective category and the baseline of the noise-masker category. Figure 2 shows the results for the normal-hearing listeners (open bars) and hearing-disabled listeners (filled bars). The ordering of the magnitudes of the three effects followed that expected from Fig. 1—the single-talker effect showed the least, the double-target effect showed the greatest—but only the multitalker effect showed a difference across normal-hearing versus hearing-disabled (a subsequent MANOVA demonstrated that this was a significant effect: [ $F(1,406)=4.31, p=0.04$ ]). None of these three effects were correlated with the participants’ ages, handicap, or overall quality of life.

The mean responses of the SSQ and the handicap questionnaire were negatively correlated [ $r(410)=-0.60, p<0.001$ ], showing that the participants who reported the greatest disability also reported the greatest handicap, replicating Gatehouse and Noble’s (2004) result.

#### IV. DISCUSSION

This analysis has shown that the scores on those items of the SSQ questionnaire that involved competing-speech situations (i.e., listening in multitalker situations, or to two simultaneous targets) were lower than the scores of those items that involved other, arguably simpler situations. Given that the dataset came from face-to-face interviews with the general public, it clearly demonstrates that competing speech is a problem for many people and that it contributes to auditory difficulties in the real world. It also shows that the respondents are at least aware of the difficulties they have in the everyday listening situations that could involve informational masking.

Nevertheless, while it is possible that a source of these competing-speech effects is informational masking, we cannot necessarily be certain that it actually is. In none of the auditory circumstances described by these vignettes can we distinguish with confidence the informational-masking component of the total masking from the energetic-masking component; indeed, if a pair of vignettes adequately described two situations with acoustically equivalent maskers that differed only in their informational content, then it seems likely that the descriptions would have to be so carefully written and prescriptive that they would not be meaningful to most participants. Such is one of the costs of a questionnaire: it is not possible to describe a situation with the same control and precision as would be expected in a laboratory experiment. But such also is the power of the questionnaire: one can describe quickly and effectively a situation that many people would recognize and understand, and so obtain data on larger, more representative samples of people than would normally be possible experimentally. To distinguish informa-

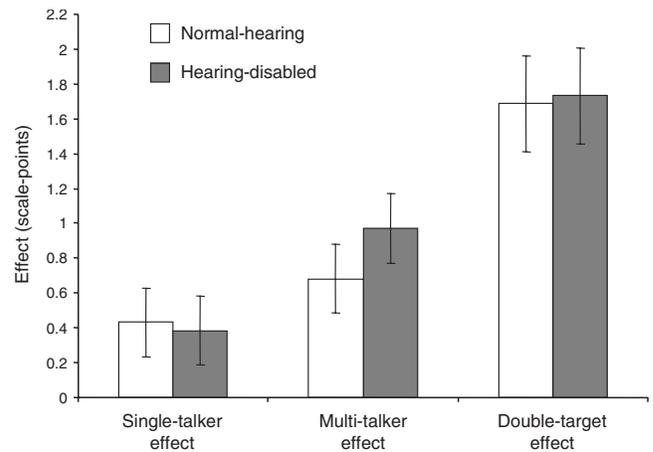


FIG. 2. Mean single-talker, multitalker, and double-target effects for normal-hearing and hearing-disabled listeners, relative to the score for the noise masker item (item 5). Error bars are 95% confidence intervals.

tional masking from energetic masking, it is perhaps necessary to return to laboratory situations, but it is not clear which simple laboratory experiments would best reflect the extent of the role of informational masking in everyday listening situations, especially as different laboratory measures lead to vastly different estimates of informational masking (compare Brungart, 2001; Arbogast *et al.*, 2002; Wu *et al.*, 2004).

Some of the vignettes in the questionnaire parallel laboratory experiments on informational masking, such as those comparing single-talker maskers with multitalker maskers (Carhart *et al.*, 1975; Brungart *et al.*, 2001). In the current analysis, listeners reported more difficulty understanding one of many talkers than one of two talkers. This could be either a reflection of the expected extra difficulty when the number of maskers are increased or a reflection of the extra total power of the maskers when there are more talkers. It is not even clear whether the multitalker maskers would cause more informational masking than single-talker maskers: if the vignettes were interpreted as many voices each at similar levels, this might result in a semicontinuous speech babble with properties similar to a noise (see Carhart *et al.*, 1975; Brungart *et al.*, 2001), but in many real-life situations, some of the nearest talkers can be more clearly heard, and could cause informational masking (note also that one’s attention could be taken by competing talkers, so reducing, at least momentarily, what is heard from the target). The present results demonstrated that the multitalker effect was larger for the hearing-disabled people than for normal-hearing people; perhaps the hearing-disabled participants were most aware of their hearing difficulties in what were arguably the most acoustically challenging situations with the likely greatest amount of power in the maskers.

The tasks for which the greatest hearing difficulties were reported involved following two simultaneous talkers. It is likely that this task has greater cognitive demands than following just one of two simultaneous talkers. This cognitive effect is more apparent in a comparison of two of the items (1 and 10) whose vignettes describe a person talking at the same time as a television is on in the room. They are distinguished by whether a listener can follow just the talker (item

1) or both the talker and the TV (item 10). Thus, the two situations were acoustically similar yet with different tasks with seemingly different cognitive demands. The participants on average reported lower scores by 1.4 scale-points when following both talkers than when following just the talker [6.7 versus 5.2 scale-points;  $t(403)=13.49$ ,  $p<0.001$ ]. Since the two situations are acoustically similar, this effect is likely to be due to the extra cognitive challenges involved in trying to understand two simultaneous talkers. In the laboratory, it is possible to observe listeners' ability to correctly report one or both of two talkers. The results are varied: some show decreased performance with the increased cognitive load (e.g., Broadbent, 1954; Humes *et al.*, 2006), but others do not (Ihlefeld and Shinn-Cunningham, 2008a, 2008b). Since the self-report data reported here came from elderly participants, it may exaggerate any difficulty with divided attention in the general population (Humes *et al.*, 2006).

There are further informative contrasts available in the SSQ data. First, items 8 and 9 of the SSQ involved following speech in competing speech when the pitches of the two talkers were either the same or different. This is perhaps analogous to the psychophysical measurements of speech understanding that have shown that competing speech is a less effective masker when it is spoken by a talker of the opposite gender (e.g., Brungart, 2001) or a different pitch to the target speech (Darwin *et al.*, 2003). There was a small but significant difference of 0.2 scale-points between these two items [ $t(396)=2.34$ ,  $p=0.02$ ], such that the listeners' responses were lower for same-pitch competing speech than for different-pitch competing speech. Thus, this contrast in the SSQ data is in the same direction as would be predicted from the psychoacoustical data. Second, items 4 and 6 of the SSQ describe the same scene (of following a conversation in a busy restaurant) and only differ in that the listener can see everyone in the group (item 4) or cannot see everyone in the group (item 6). MacLeod and Summerfield (1990) showed that speech-reception thresholds were on average 11 dB lower when visual speech cues were provided. Thus, a higher response would be expected for item 4 than for item 6. The mean response to item 4 was indeed significantly higher than the mean response to item 6 [6.4 versus 5.8;  $t(398)=8.45$ ,  $p<0.001$ ]. Again, the direction of the effect supports the link between the SSQ and the experimental data.

In summary, the respondents generally reported the most hearing difficulties for speech items with more than one talker. The largest effects were for items that involved two targets, and the smallest effects were for items with a single competing talker. For multiple competing talkers, the effects were larger for hearing-disabled listeners than for normal-hearing listeners. The results show that masking by competing speech is tangible in everyday listening situations. The descriptions of these situations in the SSQ questionnaire suggest that they could involve informational masking, but there may also be some contribution from energetic masking.

## ACKNOWLEDGMENTS

The first author was funded by a Ph.D. studentship from the Medical Research Council Institute of Hearing Research.

<sup>1</sup>The postal survey and face-to-face interviews were commissioned by Stuart Gatehouse (died, February 2007) and conducted by the National Centre for Social Research, London, in 2004. The full results are currently unpublished.

<sup>2</sup>A further two participants were excluded from the analysis because they omitted to answer the majority of the SSQ items relating to speech. There were 50 participants included in the analysis who omitted to answer one or more questions (a total of 103 missing responses). Where possible, means and other statistics were calculated from the data provided, without any corrections. In the rare cases where a participant answered none the items crucial to calculating a certain statistic, such a participant was excluded for the calculation of that statistic.

<sup>3</sup>This difference in overall SSQ score is somewhat surprising, as the original selection for interviews was designed to give approximately equal distributions of SSQ scores across the normal hearing versus hearing-disabled stratification. We believe that the effect was due to the initial selection of participants being based on a much-reduced, 6-item, version of the SSQ, rather than the full questionnaire. The difference partly motivated our analysis of the relative effects, rather than mean scores, in Fig. 2.

- Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (2002). "The effect of spatial separation on informational and energetic masking of speech," *J. Acoust. Soc. Am.* **112**, 2086–2098.
- Broadbent, D. E. (1954). "The role of auditory localization in attention and memory span," *J. Exp. Psychol.* **47**, 191–196.
- Brungart, D. S. (2001). "Informational and energetic masking effects in the perception of two simultaneous talkers," *J. Acoust. Soc. Am.* **109**, 1101–1109.
- Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (2001). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," *J. Acoust. Soc. Am.* **110**, 2527–2538.
- Carhart, R., Johnson, C., and Goodman, J. (1975). "Perceptual masking of spondees by combinations of talkers," *J. Acoust. Soc. Am.* **58**, 35.
- Carhart, R., Tillman, T. W., and Greetis, E. S. (1969). "Perceptual masking in multiple background sounds," *J. Acoust. Soc. Am.* **45**, 694–703.
- Darwin, C. J., Brungart, D. S., and Simpson, B. D. (2003). "Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers," *J. Acoust. Soc. Am.* **114**, 2913–2922.
- Drullman, R., and Bronkhorst, A. W. (2004). "Speech perception and talker segregation: Effects of level, pitch, and tactile support with multiple simultaneous talkers," *J. Acoust. Soc. Am.* **116**, 3090–3098.
- Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (2003). "Note on informational masking (L)," *J. Acoust. Soc. Am.* **113**, 2984–2987.
- Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (1999). "The role of perceived spatial separation in the unmasking of speech," *J. Acoust. Soc. Am.* **106**, 3578–3588.
- Gatehouse, S., and Noble, W. (2004). "The speech, spatial and qualities of hearing scale (SSQ)," *Int. J. Audiol.* **43**, 85–99.
- Helfer, K. S., and Freyman, R. L. (2005). "The role of visual speech cues in reducing energetic and informational masking," *J. Acoust. Soc. Am.* **117**, 842–849.
- Humes, L. E., Lee, J. H., and Coughlin, M. P. (2006). "Auditory measures of selective and divided attention in young and older adults using single-talker competition," *J. Acoust. Soc. Am.* **120**, 2926–2937.
- Ihlefeld, A., and Shinn-Cunningham, B. (2008a). "Spatial release from energetic and informational masking in a selective speech identification task," *J. Acoust. Soc. Am.* **123**, 4369–4379.
- Ihlefeld, A., and Shinn-Cunningham, B. (2008b). "Spatial release from energetic and informational masking in a divided speech identification task," *J. Acoust. Soc. Am.* **123**, 4380–4392.
- Li, L., Daneman, M., Qi, J. G., and Schneider, B. A. (2004). "Does the informational content of an irrelevant source differentially affect spoken word recognition in younger and older adults," *J. Exp. Psychol. Hum. Percept. Perform.* **30**, 1077–1099.
- MacLeod, A., and Summerfield, Q. (1990). "A procedure for measuring auditory and audiovisual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use," *Br. J. Audiol.* **24**, 29–43.
- World Health Organisation (1980). *International Classification of Impairments, Disabilities and Handicaps* (World Health Organisation, Geneva).
- Wu, X., Wang, C., Chen, J., Hongwei, Q., Li, W., Wu, Y., Schneider, B. A., and Li, L. (2004). "The effect of perceived spatial separation on informational masking of Chinese speech," *Hear. Res.* **199**, 1–10.