PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Semantic coherence and speech production in adverse listening conditions

Carine SIGNORET[1]; Emil HOLMER[2]; Mary RUDNER [3]

Linnaeus Centre HEAD, Swedish Institute for Disability Research,

Department of Behavioural Sciences and Learning, Linköping University, Sweden

ABSTRACT

According to the Framework for Understanding Effortful Listening (FUEL), multiple factors determine listening effort, including bottom-up factors such as the quality of the speech signal and top-down factors such as knowledge. Listening effort can be understood as the cognitive resources deliberately applied to a listening task. Poor sound quality may make a listening task more difficult to perform, consuming more cognitive resources and making it more effortful. Knowledge, on the other hand, may make the task easier to perform and reduce effort. Both of these factors may affect the ability to hear an interlocutor correctly and also the ability to make an appropriate response. We investigated this as follows: Twenty young adults listened to prerecorded spoken sentences with semantic coherence that was high e.g. “Her daughter was too young for the disco” or low. “Her hockey was too tight for the cotton”. Then, they either repeated the sentences or recast them as questions, e.g. “Was her daughter too young for the disco?”. The repeat and recast tasks were presented in two separate blocks each comprising equal numbers of sentences with high and low coherence. Sound quality was either clear or degraded (6-band noise-vocoding). Results showed that high semantic coherence was more beneficial during the more cognitively demanding recast task, and when speech was degraded. Better cognitive abilities were associated with the ability to make use of semantic coherence when speech was degraded, as well the ability to cope with speech degradation when semantic coherence was low. This pattern of results extends the FUEL by suggesting that semantic coherence alleviates listening effort at both perceptual and cognitive levels and is in line with the ELU model by showing that cognitive skill is important under challenging listening conditions.

Keywords: Framework for Understanding Effortful Listening (FUEL), Ease of Language Understanding Model (ELU), Semantic coherence, Noise-vocoding, Repeat, Recast

1. INTRODUCTION

We hear with our ears but we listen with our brains. This means that the harder it is to hear speech, the harder our brains must work to understand it. Degraded speech is harder to hear than clear speech, and thus more effortful to listen to. Listening effort can be understood in terms of the allocation of cognitive resources to the task of listening. According to the Framework for Understanding Effortful Listening (FUEL, 1), listening effort is a function of task-related and individual-related factors, such that listening effort increases when task demands increase, but only when individual factors allow this. For example, listening effort increases when the quality of the speech signal is degraded but only when individual resources are sufficient to keep on processing the speech signal despite increased task difficulty. If task demands are too high, listening effort will drop as the listener disengages from the task. In other words, if task demands become too high the listener may decide to conserve cognitive energy rather than straining to hear a signal that is too hard to discern.

The Ease of Language Understanding model (ELU, 2) describes how cognitive resources are allocated during listening under challenging conditions. According to the ELU model, working memory, the capacity to maintain information for processing, is engaged in speech understanding under challenging conditions. One function of working memory is to maintain predictions about the

[1] carine.signoret@liu.se
[2] emil.holmer@liu.se
[3] mary.rudner@liu.se

content of speech during listening; these predictions constrain possible interpretations of the signal. This means that individuals with greater working memory capacity are likely to be able to make better use of semantic coherence when listening to degraded speech (3)

Audiological assessment often includes speech recognition in noise tests. Typically, the test participant is presented with speech against a background of noise and the task is to repeat the speech. Because the aim of audiological rehabilitation is to achieve ergonomic listening and most speech is heard against a background of noise, it is important to measure the ability to recognize speech in noise. In everyday life, however, we often listen in order to communicate rather than simply to repeat a message we have heard. Formulating a response to a spoken message requires allocation of cognitive resources in competition with those needed to understand the message. Thus, when a response is made to a degraded spoken message, there is likely to be competition between resources required for recognizing the speech and the process of formulating the response.

In the present study, we investigated the effects of signal degradation and semantic coherence on response generation under controlled conditions. Participants listened to clear and degraded spoken sentences with high and low semantic coherence and performed two different tasks in balanced order. One task was a speech repetition task as used in a classic speech recognition in noise test. The other task involved recasting the sentence as a question by changing the word order. The dependent variable in both tasks was the proportion of words reported correctly. We predicted poorer performance in the recast than repeat task due to the cognitive load of changing the word order. For similar reasons, we also predicted that the effects of signal degradation and semantic coherence would be greater for recast than for repeat. In line with the ELU model (2), we predicted that greater cognitive skill would be associated with better ability to perform both tasks when speech was degraded but semantic coherence high. In line with the FUEL (1), we predicted that in the most difficult condition (recasting degraded low-coherence sentences) performance would be low indicating that participants had opted out.

2. METHODS

2.1 Participants

Twenty native Swedish speakers (10 females) took part in the present study. Mean age was 25 (SD = 3.1). All participants reported that they had no hearing impairment and gave informed written consent.

2.2 Materials

The materials used in the present study consisted of 200 Swedish propositional sentences (3, 4) based on an English model (5). Half of the sentences had high coherence (HC, e.g. “Her daughter was too young for the disco” and the other half had low coherence (LC, e.g. “Her hockey was too tight for the cotton”]). HC and LC sentences were matched on words, syllables and letters.

The sentences were recorded by a female native Swedish speaker in a soundproof room. The resulting audio file was edited manually to create one audio file per sentence and 10 ms linear onset and offset ramps were added. All sentences were sampled at 16 bits and amplitudes were root mean square equalized. The sound quality of the 200 sentences was degraded with noise-vocoding (NV) at 6 bands using a custom Matlab vocoder (6). This resulted in speech with an intelligibility level of around 50% (7).

2.3 Experimental tasks

The participants listened to the sentences. In the repeat task, they were instructed to repeat the sentences and in the recast task, they were instructed to rephrase the propositional sentences as questions (e.g. “Was her daughter too young for the disco?”). The tasks were blocked and there were 40 trials per block. Task order was balanced over participants and semantic coherence and sound quality were pseudorandomized within task blocks. Responses were audio-recorded and scored by 2 independent raters. The dependent variable was the proportion of words correctly reproduced in task-appropriate order.

Perception of NV speech improves with practice (8), and thus to avoid a learning effect during the task, all participants listened to 20 NV sentences while reading the corresponding text before performing the tasks. These sentences were not included in the subsequent task. All two hundred sentences were presented in a balanced fashion across participants. No sentence was repeated within participant, and all sentences were presented in each condition across participants.

2.4 Cognitive test battery

A cognitive test battery including tests of working memory, inference making ability, lexical access speed, and cognitive speed was administered to the participants.

Size-Comparison Span. Working memory was tested using the Size-Comparison Span test (SiCSpan; 9). In the SiCSpan test, series of questions are presented on a computer screen about the relative size of two semantically related objects (e.g., is a cow larger than a cat?). A yes/no button press response is given after each question and then a target item belonging to the same semantic category is presented for retention. At the end of each set of questions ranging in size from two to six, the participant is cued to recall all the target words from the set. The dependent variable is the number of target words recalled. The maximum score is 40.

Logical Inference-making Test. Inference-making ability was tested using the Logical Inference-making Test (10). Each trial is based on inferring the answer to a question on the basis of two statements. For example, "Is a JAL larger than a PONY?", "A JAL is larger than a TOC.", "A TOC is larger than a PONY". All information is provided simultaneously on a computer screen. The dependent variable was the number of correct responses (max = 16).

Lexical decision Task. In the lexical decision task (4, 10) participants judge whether three-letter strings are Swedish words. Forty items were used: 20 words, 10 pseudowords (i.e., pronounceable letter strings) and 10 nonwords (i.e., unpronounceable letter strings). The dependent variable is lexical access speed, i.e. response time in msec.

Physical Matching Task. The Physical matching task (10) is a measure of cognitive processing speed. The participant judges whether two simultaneously presented letters have the same physical shape (e.g., A-A), or no, (e.g., A-a). The dependent variable is response time in msec.

3. RESULTS

3.1 Experiment

Mean performance in all four conditions of the repeat and recast tasks is shown in Figure 1.

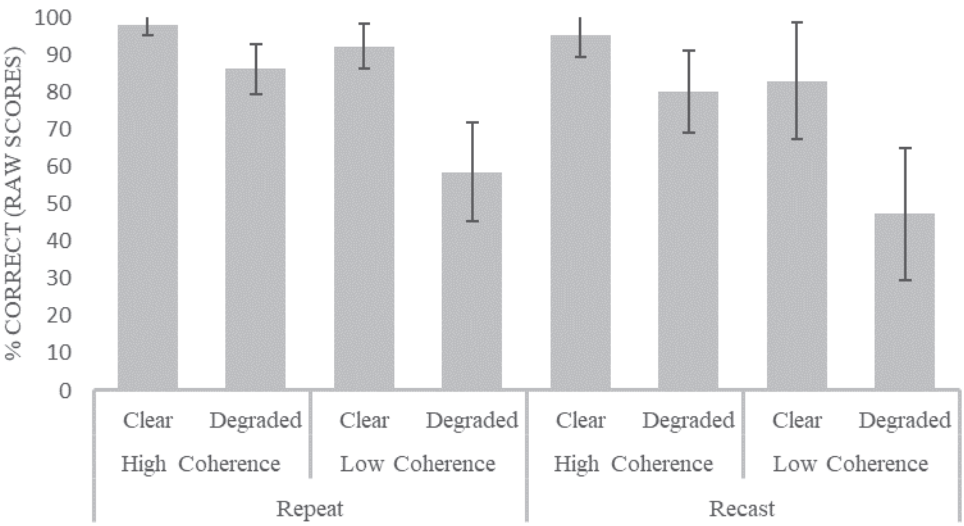

Figure 1. Mean performance (raw scores) in the repeat and recast tasks with high and low coherence sentences presented as clear or degraded.

It appears from Figure 1 that performance is better in the repeat than the recast task, better with high than low coherence sentences and when presentation was clear rather than degraded. We tested whether this pattern was statistically significant with a repeated measures within subjects analysis of variance (ANOVA). Because of ceiling performance for clear high coherence sentences, all scores were log-transformed to reduce skew. However, it should be noted that performance was still around 50% in the most difficult condition (recast task with degraded speech and low coherence). Two participants scored under 25% (14% and 19%) and both had higher than average scores on SicSpan, the test of working memory. Thus, there is no clear evidence that participants gave up with accumulating task difficulty.

Results of the ANOVA showed statistically significant main effect of all three factors: Task, $F_{(1,19)}$ = 12.99, p = .002, η^2 = .41; Coherence, $F_{(1,19)}$ = 97.42, p < .001, $\eta 2$ = .84, and Sound quality, $F_{(1,19)}$ = 135.09, p < .001, η^2 = .88. Further, there were two statistically significant 2-way interactions: Task x Coherence $F_{(1,19)}$ = 5.41, p = .031, η^2 = .22, and Coherence x Sound quality, $F_{(1,19)}$ = 51.30, p < .001, η^2 = .73. However, the Task x Sound quality interaction did not reach significance: $F_{(1,19)}$ = 3.33, p = .084, η^2 = .15, nor did the 3-way, p = .39. Thus, the ANOVA partially confirmed the pattern apparent from Figure 1 that performance is better in the repeat than the recast task, better with high than low coherence sentences and when presentation is clear rather than degraded. Further, the ANOVA confirmed that the benefit of coherence was greater when speech was degraded, p < .001. However, it was not confirmed that the effect of sound quality was greater in the recast task.

3.2 Cognitive test battery

The results of the cognitive test battery are shown in Table 1.

Table 1. Minimum and maximum scores as well as means and standard deviations (SD) for the individual tests in the cognitive test battery.

	Units	Minimum	Maximum	Mean	SD
SicSpan (working memory)	Correct	22	38	31	4
Inference-making ability	%	75	100	90	9
Lexical access speed	ms	555	1395	832	190
Physical matching (Cognitive speed)	ms	472	1146	680	160

3.3 Correlations

We predicted that participants with better cognitive skills would make better use of semantic coherence when the signal was degraded in both tasks. There was no significant correlation between SicSpan (working memory) and experimental performance in any of the conditions. However, performance in the repeat task with signal degradation and high coherence correlated significantly with inference-making ability, r = .55, p = .01, and performance on the recast task under same conditions correlated significantly with physical matching (cognitive speed), r = -.51, p = .02. When the speech signal was clear and semantic coherence was low, performance on both tasks correlated significantly with cognitive speed: repeat, r = -.48, p = .03; recast, r = -.50, p = .03.

4. DISCUSSION

4.1 Experiment

The results of the experiment showed poorer performance on the recast than repeat task supporting our prediction that recasting a sentence as a question is a harder task than simply repeating it. This supports the notion that preparing to make a response during communication is more cognitively demanding than simply listening. Signal degradation reduced performance across tasks and semantic coherence increased it, supporting our predictions that signal degradation increases task demands while semantic coherence reduces them. In line with our prediction, the effect of semantic coherence was greater in the harder, recast task, suggesting that the top down predictions allowed by semantic coherence become more important when the task is harder. This in turn suggests that semantic coherence not only reduces the task demands involved in speech recognition but also the task demands involved in recasting the sentence. This means that in a communication situation, context is more important when it comes to responding rather than simply listening to what someone is saying. However, there was no evidence that the same was true of signal degradation; instead, it seems that speech degradation simply reduces the task demands involved in speech recognition, a process involved in both tasks, but does not significantly affect the task of recasting.

The significant interaction between signal quality and coherence suggests that semantic coherence reduced the task demands generated by speech degradation. Thus, semantic coherence not only reduces the task demands of recasting but also the task demands of signal degradation. The lack of three-way interaction between task, degradation and coherence suggests again although semantic coherence influences the additional task demands of sentence recasting, signal degradation does not. Thus, bottom-up speech degradation affects perceptual but not cognitive process while top-down knowledge affects both perceptual and cognitive processes.

4.2 Correlations

There were only a few statistically significant correlations between performance on the experimental task and the cognitive test battery. In particular, we were surprised by the absence of correlations with working memory. This may have been due to the relatively small number of participants in the present study but it may also be explained by working memory not being a crucial factor for young adults with normal hearing during a listening based task, even when the task is so challenging that performance is low (c.f. 11).

We predicted that participants with better cognitive skills would make better use of semantic coherence in both tasks. We found some support for this notion in the significant correlation between inference-making ability and performance on the repeat task when the speech signal was degraded but semantic coherence was high, and the significant correlation between cognitive speed and performance on the recast task under the same conditions. The repeat task may be compared to an immediate serial recall task in that all items are to be recalled in serial order. The recast task, on the other hand may be compared to a reverse serial recall task in that item order has to be changed on recall in a systematic manner. Thus, one tentative interpretation of this pattern of results is that inference-making ability is important for serial recall of semantically coherent but degraded items. This may be because good inference-making ability facilitates use of context when speech is degraded. This should be investigated in future research. Cognitive processing speed on the other hand seems to be of the essence for order processing of semantically coherent but degraded items. This may be due to speed of processing facilitating the working memory processing that underpins reordering of items. Cognitive processing speed was important in both the repeat and recast tasks when speech was clear but semantic coherence low. When sentence coherence is low, it may be difficult to maintain serial order of items and thus, a similar interpretation may be offered here: cognitive processing speed is of the essence for order processing during conversation. However, the correlation pattern should be interpreted with caution considering the small number of correlations and large number of potential comparisons among variables.

5. CONCLUSIONS

Audiological testing often includes speech-in-noise testing to determine speech processing ability. However, in everyday situations it is more appropriate to generate a response, such as requesting clarification, rather than simply repeating a statement you have heard. Therefore, we studied the effect of speech degradation and semantic coherence on the ability to provide a particular kind of response to a spoken statement, which involved recasting it as a question. In line with predictions, we found that semantic coherence was more important when a sentence was recast as a question rather than simply repeated and that this effect was greater when speech was degraded. This finding extends the FUEL (1) by showing that semantic coherence eases listening effort at both the perceptual and cognitive levels. We also found some evidence that cognitive ability was associated with the ability to make use of semantic coherence when the speech signal was degraded, as well as with the ability to cope with low semantic coherence when speech was clear, in line with the ELU model (2). There was little evidence of participants giving up, as predicted by FUEL (1), even in the hardest conditions. Future work should further investigate the perceptual and cognitive effects of semantic coherence during speech processing in noise, including response generation, and under a range of challenging conditions.

ACKNOWLEDGEMENTS

The authors wish to thank Jonna Hammarsten and Hanna Svensson for help with data collection. This work was supported by the Swedish Research Council Grant 2017-06092 and through funding of the Linnaeus Centre HEAD Grant 349-2007-8654.

REFERENCES

1. Pichora-Fuller MK, Kramer SE, Eckert MA, Edwards B, Hornsby BW, Humes LE, Lemke U, Lunner T, Matthen M, Mackersie CL, Naylor G, Phillips NA, Richter M, Rudner M, Sommers MS, Tremblay KL, Wingfield A. Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). Ear Hear. 2016 Jul-Aug;37 Suppl 1:5S-27S. doi: 10.1097/AUD.0000000000000312.
2. Rönnberg J, Holmer E, Rudner M. Cognitive hearing science and ease of language understanding. Int J

Audiol. 2019 May;58(5):247-261. doi: 10.1080/14992027.2018.1551631.
3. Signoret C, Johnsrude I, Classon E, Rudner M. Combined effects of form- and meaning-based predictability on perceived clarity of speech. J Exp Psychol Hum Percept Perform. 2018 Feb;44(2):277-285. doi: 10.1037/xhp0000442.
4. Signoret C, Rudner M. Hearing Impairment and Perceived Clarity of Predictable Speech. Ear Hear. 2019 Jan 7. doi: 10.1097/AUD.0000000000000689. [Epub ahead of print]
5. Davis MH, Ford MA, Kherif F, Johnsrude IS. Does semantic context benefit speech understanding through "top-down" processes? Evidence from time-resolved sparse fMRI. J Cogn Neurosci. 2011 Dec;23(12):3914-32. doi: 10.1162/jocn_a_00084.
6. Wild CJ, Yusuf A, Wilson DE, Peelle JE, Davis MH, Johnsrude IS. Effortful listening: the processing of degraded speech depends critically on attention. J Neurosci. 2012 Oct 3;32(40):14010-21. doi: 10.1523/JNEUROSCI.1528-12.2012.
7. Zekveld AA, Rudner M, Johnsrude IS, Rönnberg J. The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. J Acoust Soc Am. 2013 Sep;134(3):2225-34. doi: 10.1121/1.4817926.
8. Hervais-Adelman AG, Davis MH, Johnsrude IS, Taylor KJ, Carlyon RP. Generalization of perceptual learning of vocoded speech. J Exp Psychol Hum Percept Perform. 2011 Feb;37(1):283-95. doi: 10.1037/a0020772.
9. Sörqvist P, Ljungberg JK, Ljung R. A sub-process view of working memory capacity: evidence from effects of speech on prose memory. Memory. 2010 Apr;18(3):310-26. doi: 10.1080/09658211003601530.
10. Rönnberg J, Lunner T, Ng EH, Lidestam B, Zekveld AA, Sörqvist P, Lyxell B, Träff U, Yumba W, Classon E, Hällgren M, Larsby B, Signoret C, Pichora-Fuller MK, Rudner M, Danielsson H, Stenfelt S. Hearing impairment, cognition and speech understanding: exploratory factor analyses of a comprehensive test battery for a group of hearing aid users, the n200 study. Int J Audiol. 2016 Nov;55(11):623-42. doi: 10.1080/14992027.2016.1219775.
11. Füllgrabe C, Rosen S. On The (Un)importance of Working Memory in Speech-in-Noise Processing for Listeners with Normal Hearing Thresholds. Front Psychol. 2016 Aug 30;7:1268. doi: 10.3389/fpsyg.2016.01268.