

The role of spatial separation of two talkers' auditory stimuli in the listener's memory of running speech: listening effort in a non-noisy conversational setting

Edina Fintor, Lukas Aspöck, Janina Fels & Sabine J. Schlittmeier

To cite this article: Edina Fintor, Lukas Aspöck, Janina Fels & Sabine J. Schlittmeier (2021): The role of spatial separation of two talkers' auditory stimuli in the listener's memory of running speech: listening effort in a non-noisy conversational setting, *International Journal of Audiology*, DOI: [10.1080/14992027.2021.1922765](https://doi.org/10.1080/14992027.2021.1922765)

To link to this article: <https://doi.org/10.1080/14992027.2021.1922765>



© 2021 The Authors. Published by Informa UK Limited, trading as Taylor & Francis Group on behalf of British Society of Audiology, International Society of Audiology, and Nordic Audiological Society.



Published online: 15 Jun 2021.



[Submit your article to this journal](#)



Article views: 229




[View related articles](#)



[View Crossmark data](#)

The role of spatial separation of two talkers' auditory stimuli in the listener's memory of running speech: listening effort in a non-noisy conversational setting

Edina Fintor^a , Lukas Aspöck^b, Janina Fels^b and Sabine J. Schlittmeier^a

^aTeaching and Research Area Work and Engineering Psychology, RWTH Aachen University, Aachen, Germany; ^bInstitute for Hearing Technology and Acoustics, RWTH Aachen University, Aachen, Germany

ABSTRACT

Objective: To investigate the role of the spatial position of conversing talkers, that is, spatially separated or co-located, in the listener's short-term memory of running speech and listening effort.

Design: In two experiments (between-subject), participants underwent a dual-task paradigm, including a listening (primary) task wherein male and female talkers spoke coherent texts. Talkers were either spatially separated or co-located (within-subject). As a secondary task, visually presented tasks were used. Experiment I involved a number-judgement task, and Experiment II entailed switching between number and letter-judgement task.

Study sample: Twenty-four young adults who reported normal hearing and normal or corrected to normal vision participated in each experiment. They were all students from the RWTH Aachen University.

Results: In both experiments, similar short-term memory performance of running speech was found independently of talkers being spatially separated or co-located. Performance in the secondary tasks, however, differed between these two talkers' auditory stimuli conditions, indicating that spatially separated talkers imposed reduced listening effort compared to their co-location.

Conclusion: The findings indicated that auditory-perceptive information, such as the spatial position of talkers, plays a role in higher-level auditory cognition, that is, short-term memory of running speech, even when listening in quiet.

ARTICLE HISTORY

Received 15 April 2020

Revised 9 March 2021

Accepted 20 April 2021

KEYWORDS

Short-term memory;
running speech; listening
effort; spatial location;
spatial hearing; listening



Introduction

Face-to-face conversations between two talkers are probably the most common verbal form of communication in everyday life. Often, one or more persons are listening to these conversations, such as in the office, classroom, or at a party. In these situations, the listener is challenged to follow the talkers' speech and remember the content to join the conversation, if needed or wanted, or to remember information later.

Listening to such a conversation and remembering what has been said requires auditory processes, cognitive functions, and the interplay between them (see Edwards 2016, for auditory-cognitive models). The auditory perception of the talkers' spatial location or the pitch of their voices relies predominantly on the physical stimulus aspects of the incoming speech signals (see Bregman 1990, for auditory scene analysis). Extracting and processing its semantic and conversational content, however, necessitates the interplay of several cognitive functions and processes (e.g. short-term memory, verbal-logical reasoning, and focussed attention). Hence, listening to running speech, comprehending, and remembering what has been said has to be considered a highly demanding task since all auditory-perceptive and cognitive processes need to be accomplished for speech signals, which are inherently strictly sequential, not-repeatable and presented at a given speed (Imhof 2010).

Since the physical speech signal conveys auditory-perceptual information, as well as semantic meaning, auditory-perceptual and cognitive speech processes, are not independent of each other. If listening conditions are adverse because of a noisy environment, the relevant speech signal's auditory-perceptive cues are reduced or distorted. Yet, even if the intelligibility of relevant speech remains near to perfect in such conditions, short-term memory for its semantic content can be diminished as for example, demonstrated by Surprenant (1999) for immediate recall of unrelated items in a verbal serial recall task. This phenomenon is viewed as a strain on cognitive resources when carefully listening to speech in noise (i.e. listening effort, which we address later in more details) that results in a reduction of available resources for further cognitive processing of the overheard speech, such as maintenance of its semantic content for recall. Conversely, one could assume that when listening to a two-talker conversation, the availability of auditory-perceptual cues, like talkers' differing spatial positions or voice timbres, might facilitate memory for heard conversational content. This might even apply in the ideal – namely quiet – listening settings.

Some cognitive processes, like maintaining information when listening to a conversation, might be extended after the auditory information has been encoded in the listener's cognitive system and attention has been focussed on the target source (Wingfield 2016).

CONTACT Edina Fintor  Edina.Fintor@psych.rwth-aachen.de  Teaching and Research Area Work and Engineering Psychology, RWTH Aachen University, Jägerstraße 17-19, Aachen 52066, Germany

© 2021 The Authors. Published by Informa UK Limited, trading as Taylor & Francis Group on behalf of British Society of Audiology, International Society of Audiology, and Nordic Audiological Society.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Consequently, the availability of auditory-perceptive cues when listening to speech arguably affects the listener's memory and comprehension of speech content. Studies that investigated the role of auditory-perceptive information in listening emphasised the outstanding importance of spatial acoustic cues in, for example, selecting the target speech signal from interferers (see, for example, Bronkhorst 2000; Carlyon et al. 2001; Cherry 1953; Ihlefeld and Shinn-Cunningham 2008; Oberem et al. 2014). In contrast, several studies claiming that other acoustic cues, such as voice differences between talkers, are more relevant for selecting and focussing on the target speech in such a multi-talker environment (Gallun et al. 2013; Humes, Lee, and Coughlin 2006).

Although these studies provided valuable insights, for example, into how speech recognition depends on auditory-perceptual signal characteristics reflecting differing spatial locations of talkers (see, for example, Bronkhorst 2000 for a review), they used mostly basic stimuli and relatively simple audio presentation methods. Thus, whether these findings could be transferred to situations in which listening to running speech by two talkers taking turns in a conversation is an open question. For example, memory for verbal material has been predominantly investigated using spoken digits or unrelated sentences with single-talker recordings (e.g. Conway, Cowan, and Bunting 2001). Undoubtedly, such stimuli require a series of perceptual and cognitive processes, like organising auditory information into meaningful units, such as words, and activating the mental lexicon. However, the cognitive processes necessary to answer questions about coherent running speech go well beyond this. Here, words must be related to each other, content must be linked across sentences, and information must be stored at least over a short period of time to draw any conclusions.

Memory and comprehension of running speech can be measured directly by posing questions to the listener (see Hafter et al. 2013, for measuring listening comprehension). Yet, such direct performance measures are often not sensitive to experimental variations of, for example, talkers' spatial location. For example, a multi-talker environment study by Xia et al. (2015) investigated the effect of spatial and gender cues in a speech-recognition task. It measured the speech performance when the target voice was masked by two other voices (i.e. so-called cocktail-party like situation). Although they found comparable speech recognition rates when the target voice was separated from interferers either by spatial location or by gender cue, they found a performance benefit in a visual tracking task, conducted in parallel with the speech-recognition task when talkers were spatially separated. The observed performance differences in that visual tracking task were interpreted as a release of cognitive resources by spatial separation of talkers. Accordingly, measuring cognitive resource allocation or the so-called listening effort has been proven to be a suitable indicator of direct performance measurements in listening research (see also Zekveld et al. 2014).

Listening effort refers to the number of processing resources (e.g. perceptual, attentional, cognitive) a listener allocates to a specific auditory task when the listener aims to reach a high-level performance on the task and/or the environment is suboptimal for speech understanding (see, for example, Gagné, Besser, and Lemke 2017; McGarrigle et al. 2014). While listening effort provides a valuable insight into listening research, it is not easy to measure. As already shown in the study by Xia et al. (2015), it is possible to obtain similar performance measures in the task related to the heard speech signal while the listening effort necessary to achieve that performance could differ. In this case, the dual-task paradigm, where two tasks must be performed in

parallel, offers an advantageous approach as a bridge between performance directly concerned with the presented speech signal and listening effort measures. The basic idea behind the dual-task paradigm is that the cognitive resources that a person has available are limited (Kahneman 1973). Thus, the more resource-demanding task, the fewer resources are available for the second, parallel task. In listening research, generally, the task that is supposed to measure speech performance directly, as speech recognition, or in our case remembering conversational contents, is defined as the primary task and the task that is used as an indicator for listening effort is defined as the secondary task. Accordingly, although if primary-task performance remains similar between two experimental conditions A and B, reduced secondary-task performance in condition A would indicate higher listening effort in this experimental condition (see Gagné, Besser, and Lemke 2017; McGarrigle et al. 2014).

Although an increasing body of research has used the dual-task paradigm to understand the processes underlying listening, these studies focussed on speech recognition of single syllables, words or isolated sentences as primary tasks, as mentioned above (see Gagné, Besser, and Lemke 2017, for a review), and did not consider using running speech, for example, consisting of a row of interrelated sentences, or measuring more complex cognitive processes, like memory for heard content and text comprehension. Moreover, the existing studies investigated listening in adverse conditions in which, for example, speech is presented in noise or accompanied by task-irrelevant talkers (see Hafter et al. 2013; Xia et al. 2014, for listening comprehension in the so-called cocktail party situation) while less is known about the role of auditory-perceptive information in the ideal – in terms of quiet – listening situations.

The mutual dependence of auditory-perceptive and cognitive processing on mental resources suggests that memory for running speech contents should be affected by listening effort. The ease of language understanding (ELU) model, which links auditory-perceptive information, language comprehension, and listening effort, presents a theoretical framework for these effects (Rönnerberg et al. 2013). The ELU model distinguishes between an implicit and explicit cognitive processing route of speech understanding. If the implicit route fails, which could happen in case of a mismatch between the mental representation of a reduced or distorted auditory signal and long-term memory representations, the explicit route is triggered. The explicit route is more focussed, slow, and resource-demanding compared to the more automated implicit route. Thus, the more explicit processing, the more listening effort a listener needs for speech understanding.

Although the ELU model was designed to understand adverse listening situations where the auditory signal is distorted, the model could be applied to understand the role of auditory-perceptive information in memory functioning in a quiet listening environment. We argue that spatial separation of talkers supports speech understanding by driving the implicit route, that is, a more automated route. This is because the auditory system uses spatial location – derived from a combination of different auditory-perceptive information – to unambiguously identify perceptual objects as separate entities, that is, two talkers of a conversation. In real life, it is quite rare, rather impossible, that two different talkers' audio signals come from the same spatial location. Thus, co-located talkers might call for the explicit route in at least two ways. Firstly, co-location might necessitate explicit route processing to compensate for a lack of information on auditory-perceptive distinctiveness which is “normally” available to unambiguously separate the two entities, respectively.

Secondly, co-located talkers might be perceived as a deviancy in the cognitive system resulting in explicit route processing to be resolved. Vice versa, spatial separation of talkers should drive the implicit route and thus more automated processing, which, in the end, results in more cognitive resources available for other cognitive performances, like memory recall of a heard conversational contents. Accordingly, we expected and tested improved short-term recall performance and/or reduced listening effort in conditions in which talkers were spatially separated.

Research intent

The present study investigated the role of two talkers' spatial position, that is, either co-located (i.e. same position) or spatially separated, in-text comprehension and memory of coherent speech in a quiet setting where two talkers took turns in a conversation. For this purpose, we developed a listening task containing 16 spoken texts and corresponding questions assessing memory and text comprehension. These texts were comparable in terms of necessary previous knowledge of the chosen topic (family stories considering three generations: grandparents, parents, children), length (words and number of sentences), and difficulty of the questions to be answered. Consequently, within-subject testing of different experimental manipulation was possible and was pre-tested.

We used the dual-task paradigm to measure listening effort, with the primary task being listening to and remembering the contents of a two-talker-conversation. These two talkers were either presented binaurally via headphones at two different positions in space ($\pm 60^\circ$ relative to the listener) or at the same position (0°), that is, co-located. For the secondary task, we used a number-judgment task in Experiment I performed concurrently with the primary task. This number-judgment task has already been used successfully as a secondary task in the dual-task paradigm (see, for example, Sarampalis et al. 2009, see also Seeman and Sims 2015 for dual-task measures of listening effort). In Experiment II, we used the same primary task as in experiment I but with a more demanding secondary task wherein participants had to switch between a number and a letter task while listening to coherent texts. In both experiments, we tested for effects of conversing talkers being either co-located or spatially separated on memory for what has been said (performance in the primary task) and listening effort (indicated by performance in the secondary task).

Experiment I

Method

Participants

Twenty-four participants (19 women, between 18 and 29 years, $Md = 21$ years) who reported normal hearing and normal or corrected to normal vision took part in the experiment. They were all students at the RWTH Aachen University and received credit points or 8€ for participating in the experiment. Only persons signing informed consent took part in the experiment.

Stimuli, task, and apparatus

The experiment was programmed in Psychopy 3.1.5 and run entirely on Dell Latitude 3590 laptop. All visual material was presented on the notebook's non-glare 15" screen, and all auditory stimuli were played back via a Focusrite Scarlett 2i2 2nd Gen external soundcard and Sennheiser HD 650 headphones.

In the listening (primary) task, the auditory stimuli were spoken coherent texts about family stories considering three

generations (grandparents, parents, and children). Each text consisted of 10 sentences, and it was presented as a conversation between one talker with a female voice and another talker with a male voice. All speech material was spoken by a professional female speaker and recorded in a soundproof booth with very low reverberation time. The mean pitch of the female voice was 177 Hz according to analysis using Praat (Boersma and Weenink 2020). The voice signal of this female talker was altered in timbre utilising pitch shifting (Adobe Audition) so that it sounded like a male voice. The mean pitch of the male voice was 138 Hz after pitch tuning. This procedure allowed to ensure that female and male voices had the same talking style, prosody, and pace of speech. In a pre-test with $n = 8$ participants, the participants correctly assigned the two voices to the male and female gender and judged them as natural voices. The turn-taking between the female and the male voice aimed to simulate a natural conversation, so sentences linked closely together were spoken by the same voice (the number of sentences spoken by either the female or the male voice was counterbalanced). Most importantly for the present study, the talkers never spoke simultaneously so that no partial masking of speech signals took place.

For each text, nine questions had to be answered, which asked for names of family members, relations between family members, and further information (e.g. profession, locations, hobbies, age). Questions, related to one specific text were arranged in a fixed order but did not follow the order of the information from the heard text. For example, the name of one's father was told in the second sentence but asked back in the fifth question. These patterns – that is, the order of the information in the text and the order of the questions recalling that information – was random across the texts. Questions were presented visually one after the other, and responses were given by typing. Each question could be answered with one word, and error coding was done after the experiment.

For the two experimental conditions of co-located vs. spatially separated talkers, the male and the female voice were binaurally presented at a distance of 2.5 m from the listener either at $\pm 60^\circ$ (spatially separated condition) or at the same position in space (0° ; co-located condition). All texts were presented at $L_{Aeq} = 54$ dB(A). The binaural signals were created by spectral convolution of the corresponding head-related transfer functions (HRTFs) with the recorded voice signals. The HRTFs were taken from a database of a measured head and torso simulator (Schmitz 1995), the corresponding HRTF dataset can be found in (Aspöck, Viveros Munoz, and Fels 2020). In order to improve the binaural sensation of the signals, the frequency response of the headphones used in the experiment (Sennheiser HD 650) was additionally equalised according to (Masiero and Fels 2011).

In the number of judgement (secondary) tasks, visual stimuli were presented. Here, white numbers from 1 to 9, except for 5, were presented in the centre of the notebook screen on a black background with a width and height of 1.5 cm. The viewing distance was approximately 60 cm.

Procedure

The experiment was conducted in single sessions in a soundproof booth (Studiobox, premium edition) at the Teaching and Research Area Psychology with a Focus on Auditory Cognition of the RWTH Aachen University. Written instructions appeared on the screen, and the experimenter explained the task orally before the experiment. Participants were asked to respond quickly and accurately to the secondary task while listening carefully. Participants

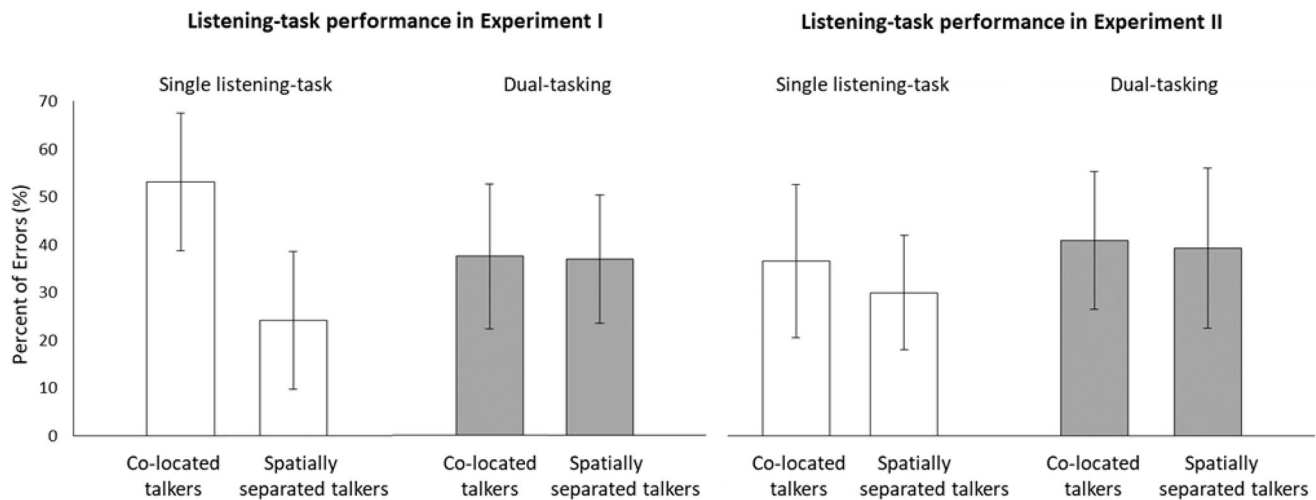


Figure 1. Listening-task performance in Experiment I and in Experiment II. *Note.* Listening-task performance in Experiment I (left panel) and in Experiment II (right panel) as a function of number of tasks (single listening-task vs. dual-tasking) and talkers' auditory stimuli (co-located talkers vs. spatially separated talkers). Error bars represents the 95% confidence interval of the mean.

started with two single-task blocks of the listening (primary) task, consisting of one text spoken by two talkers, either co-located or spatially separated (order was counterbalanced across participants). It was followed by a single-task block of the number-judgement task (secondary task), consisting of 32 trials (baseline condition for secondary-task performance). After these two single-task conditions, each participant took part in two dual-task conditions in which listening and number-judgement tasks had to be performed in parallel. In one dual-task condition, the running speech was spoken by two co-located talkers.

In contrast, in the other dual-task condition, talkers were spatially separated (condition order was counterbalanced across participants). Both dual-task conditions consisted of one practice block of one text followed by six experimental blocks each consisting of one text (text order was randomised within conditions). In the dual-task blocks, the number of trials of the number-judgement task was defined by the time of the listening task, and participants could perform as many trials as possible.

In the number-judgement task, each trial started with the onset of the stimulus and lasted until either response was made or until 1500 ms had elapsed. The participants indicated whether the presented number was lower or higher than five. Responses were given by pressing the “P” or “j” button with the left and right index finger, respectively. Stimulus-response mapping was counterbalanced across participants. The response-stimulus interval was 500 ms. After each text of the listening task, the corresponding questions were presented on the screen one after the other, and participants entered their responses via the notebook's keyboard. Each question lasted until a response was made. The participants could omit a question by pressing the return button, but going back to a missed question or correcting a response was not possible. The experiment lasted about one hour.

Results

Memory of conversational contents (performance in the primary-task)

Memory performance was analysed using the percentage of errors (PE). Each incorrect answer on a text question was counted as an error, and individual error rates were pooled over

participants for each of the two talkers' auditory stimuli conditions, namely the co-located and the spatially separated talkers.

First, we investigated memory performance in the single-listening task in which participants did not work on a secondary task in parallel with the listening task. A paired-sample *t*-test verified a significant difference between the two talkers' auditory stimuli conditions, $t(23) = 8.936$, $p < 0.001$, Cohen's $d = 2.55$. Here, participants made more errors in the co-located talkers condition compared to the spatially separated condition ($M = 53.2\%$, $SD = 17.9$ vs. $M = 24.1\%$, $SD = 17.8$).

Second, we tested the effects of dual-tasking on memory performance in the listening task. Therefore, we conducted a 2×2 ANOVA on PE with the independent within-subject factors *talkers' auditory stimuli* (co-located vs. spatially separated) and the *number of tasks* (single listening-task vs. dual-tasking). The main effect of *talkers' auditory stimuli*, $F(1, 23) = 15.988$, $p = 0.001$, $\eta^2_p = 0.410$, and interaction between *talkers' auditory stimuli* and *number of tasks*, $F(1, 23) = 15.576$, $p = 0.001$, $\eta^2_p = 0.403$, was significant. The main effect on the *number of tasks* was not significant, $F < 1$. Due to the significant interaction, the main effects cannot be interpreted directly. Thus, post hoc comparisons using Tukey's test were conducted to clarify the exact effect pattern. These tests indicated that significantly more errors were made in answering the questions when talkers co-located rather than when they were spatially separated (PE: $M = 45.5\%$, $SD = 16.7\%$ vs. $M = 30.8\%$, $SD = 13.3\%$). The significant interaction of *talkers' auditory stimuli* with the *number of tasks* was due to the difference between dual-tasking and single listening-task being larger in the co-located than in the spatially separated condition, (PE: $M = -15.6\%$, $SD = 16.7$ vs. $M = 13.51\%$, $SD = 22.4$). As Figure 1 (left panel) indicates, the performance improved in the co-located condition but declined in the spatially separated condition when the listening task was performed in parallel with the number-judgement task (i.e. when dual-tasking) and not as a single-task.

Listening effort: performance in the number-judgement task (secondary task)

Participants performed the number-judgement task in three conditions: (a) as a secondary task in parallel with the listening task with the talkers being either co-located or (b) spatially separated

Number-judgment task performance in Experiment I

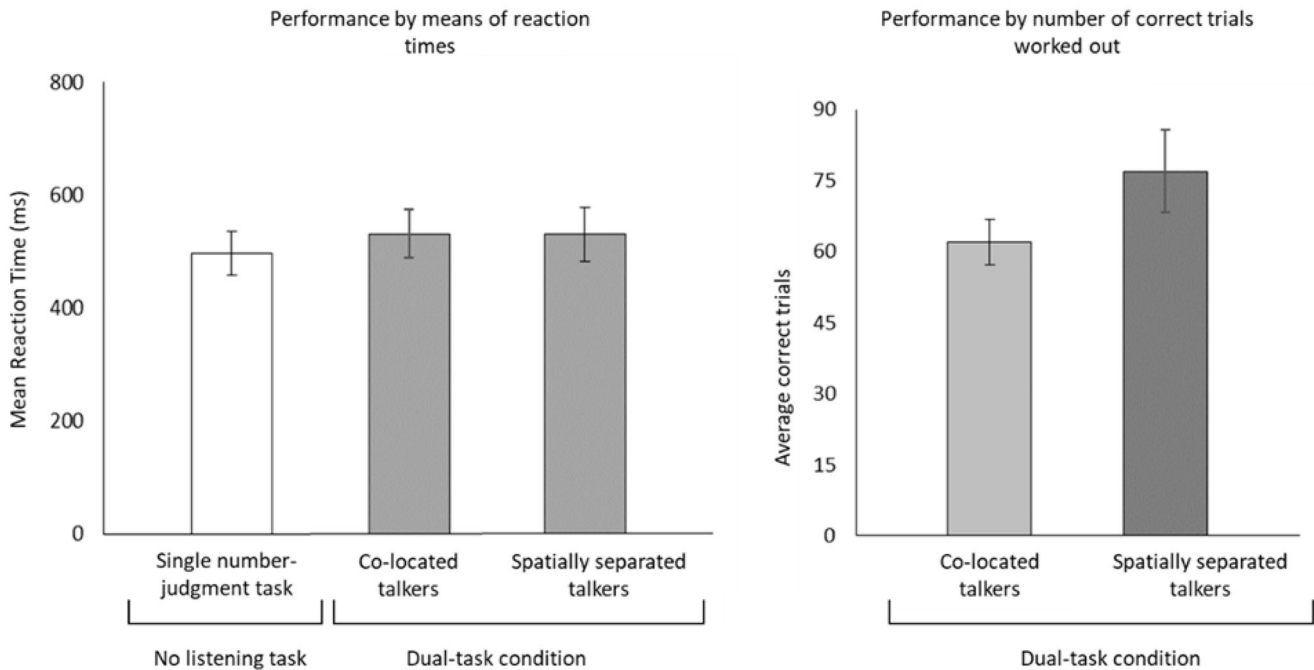


Figure 2. Number-judgement task performance in Experiment I. Note. Performance by means of reaction times (left panel) as a function of task conditions (single number-judgement task, co-located dual-tasking, spatially separated dual-tasking). Performance by number of correct trials worked out (right panel) a function of talkers' auditory stimuli (co-located dual-tasking, spatially separated dual-tasking). Error bars represents the 95% confidence interval of the mean.

and (c) without the listening task (baseline performance in the number-judgement task). The performance was analysed via three measures: reaction times (RTs), assessing the time from the appearance of a number to the participant's response, percentage of errors (PE), and the number of trials. For RT analyses, error trials (overall, 9.7% of the judgments were wrong), and outliers were excluded. To identify the latter, the median absolute deviation was calculated for each participant (see Leys et al. 2013), and values exceeding ± 3 were discarded.

First, we conducted two ANOVAs, one for RTs and one for PE, to compare performance in the three number-judgement task conditions (a)–(c). These ANOVAs revealed a significant effect of task conditions on RTs, $F(2, 46) = 3.563$, $p < 0.05$, $\eta^2_p = 0.134$, but not on PE, $F < 1$. To clarify the effect pattern on RTs, post hoc comparisons were conducted using Tukey's test. These verified that RTs in the number-judgement task were significantly longer when the listening task had to be performed in parallel with the number-judgement task (condition a: $M = 530$ ms, $SD = 59.1$ ms, and condition b: $M = 531$ ms, $SD = 53.4$ ms) as opposed to when only the number-judgement task had to be performed (condition c: $M = 496$ ms, $SD = 45.0$ ms). These results are depicted in the left panel of Figure 2.

To investigate the difference between the two talkers' auditory stimuli conditions, we conducted a paired-sample t -test on the average number of correct trials in the number-judgement task per block. The significant t -test, $t(23) = -9.99$, $p < 0.001$, Cohen's $d = 3.10$, indicated that participants completed more trials when conversing talkers were spatially separated compared to when talkers were co-located ($M = 77$, trials, $SD = 7.9$ trials vs. $M = 62$ trials, $SD = 8.1$ trials; see Figure 2, right panel). The data of Experiment I are openly available at <https://osf.io/dqfng>.

Discussion experiment I

In Experiment I, we used a visual number-judgement task as a secondary task to measure listening effort when participants had to listen to a two-talker conversation wherein talkers were either spatially separated or co-located. In the listening (primary) task, we found similar performance in both talkers' auditory stimuli conditions in that participants could recall the same amount of information independently of whether the talkers were spatially separated or co-located. In the number-judgement (secondary) task, while we found no effect of the two talkers' auditory stimuli condition on RTs, there was an effect on the number of correctly worked out trials. Participants were able to work out more trials in the secondary task when listening to running speech spoken by two talkers auralized at different spatial locations instead of being co-located. Although variations in RTs are, by standard, interpreted as an indicator of differences in listening effort between auditory conditions (see Gagné, Besser, and Lemke 2017), this measure also has its limitations.

As only relatively few errors were made in the number-judgement task (9.7%), one could argue that this task was a very simple and easy secondary task to the participants. If a task is easy, it is possible that participants can adapt quickly to the task demands, and only a small amount of cognitive resources is necessary to complete the task (Hasher and Zacks 1979). Hence, we might have measured a ceiling effect. Picou and Ricketts (2014) used three different tasks to measure listening effort, and they found that the simple visual RT tasks (participants were asked to press a button when the presented rectangle was red but withhold any response when it was white) were not sensitive to listening effort changes. Similarly, Sarampalis et al. (2009) found that participants were quite fast in the RT task (similar

number-judgement task as we used in Experiment I) in the silent condition implemented as a baseline measure. Consequently, RTs in the number-judgement task might not be sensitive to changes in the listening effort necessary to accomplish the primary (listening) task. To test this assumption, we decided to conduct a second experiment in which we used the same listening task as in Experiment I but now with a more demanding secondary task.

Experiment II

In Experiment II, we combined the listening task of Experiment I with a more complex secondary task. A dual-task study by Picou and Ricketts (2014) already found that increasing the secondary task complexity increases the sensitivity to measure the effects of experimental manipulations on listening effort. Furthermore, we aimed to explore whether here, too, the number of correct trials varies across the present experimental conditions, which could indicate variations in listening effort. If the results of our first experiment were replicated, the number of correctly worked out trials could be considered as a valuable indicator of listening effort beyond RT data in listening-effort research.

To this end, a more demanding secondary task was introduced in parallel with the listening task. Participants had to switch between a number (N) and a letter (L) judgement task, so the so-called task-switching paradigm was implemented as the secondary task in the dual-task paradigm (see, for example, Koch et al. 2018, for a review of the task-switching paradigm). In this paradigm, participants are asked to work on at least two different tasks sequentially by switching between them. To this end, a letter-judgement task was introduced in Experiment II, additionally to the number-judgement task that was also used in Experiment I. If a digit was presented, the number task had to be performed. If a letter was presented, the letter task had to be done. In this case, a task sequence could be like N - N - N - L* - L - L - L - N* N...), where the asterisk indicated a “switch trial” when the required task differed from the previously solved task. The other trials, where the upcoming task was the same as the previous one, were called “repetition trial”. Generally, performance is better on repetition trials than on switch trials, and the difference between them is referred to as the so-called switch costs. The switch costs are assumed to reflect some cognitive processes taking place when activating and implementing a new task (see, for example, Rubinstein, Meyer, and Evans 2001, for further information on cognitive processes in task switching). According to the idea of listening effort, one could assume that the more resource-demanding the listening task, the fewer resources are available for task switching in the secondary task, resulting in larger switch costs.

Method

Participants

Twenty-four participants (1 male, between 18 and 29 years, $Md = 18$ years) who were students at the RWTH Aachen University participated in experiment II. No participant had taken part in Experiment I. All participants reported normal hearing and normal or corrected to normal vision. They received credit points or 8€ as compensation for their participation. Only persons who signed informed consent were included in the experiment.

Stimuli, task, and apparatus

Experiment II differed from Experiment I only in the secondary task. In addition to the number-judgement task, a letter task was presented as well. Here, instead of numbers, letters (A, E, O, U, B, F, G, H) were presented.

Procedure

The procedure was the same as in Experiment I, but in the secondary task, numbers and letters were presented randomly. If a letter was presented, the task was to decide whether the letter was a vowel or a consonant. Responses were given using the same button press as in the number task (counterbalanced across participants).

Additionally, in the condition in which the number-letter task was performed as a single-task, the number of trials was controlled by time, allowing the participants to produce as many correct trials as they could. Time was defined by the average length of all texts (84 s).

Results

Memory of conversational contents (performance in the primary-task)

In Experiment II, we measured memory performance in the condition in which no secondary task had to be performed in parallel. To test for the performance difference between the two experimental conditions, co-located talkers versus spatially separated talkers, we conducted a paired sample *t*-test, which was non-significant, $t(23) = 1.252$, $p = 0.22$, Cohen's $d = 0.40$. Thus, memory performance did not differ between the two talkers' auditory stimuli conditions.

However, the listening task was performed not only as a single task but also in parallel with a secondary task and thus under dual-task conditions. To test the effects of single- vs. dual-tasking on memory performance in the listening task, we conducted a 2×2 ANOVA with the independent within-subject factors *talkers' auditory stimuli* (co-located vs. spatially separated) and the *number of tasks* (single listening-task vs. dual-tasking) and the dependent variable PE. Neither the main effect of *talkers' auditory stimuli*, $F(1, 23) = 1.136$, $p = 0.28$, $\eta^2_p = 0.047$, nor the interaction, $F < 1$, was significant. However, there was a tendency on the main effect of *number of tasks*, $F(1, 23) = 3.120$, $p = 0.08$, $\eta^2_p = 0.119$. Here, post hoc comparisons using the Tukey's test indicated that the PE tended to be larger when dual-tasking rather than when performing the listening task alone ($M = 40.1\%$, $SD = 18.2\%$ vs. $M = 33.3\%$, $SD = 12.5\%$; see Figure 1, right panel).

Listening effort: performance in the switching task (secondary task)

Participants performed the switching task in three conditions: (a) as a secondary task in parallel with the listening task with the talkers being either co-located or (b) spatially separated and (c) without the listening task (baseline performance in the number-letter task). Additionally, in each condition, trials were defined as a switch trial (i.e. a number-judgement trial followed by a letter-judgement trial or vice versa) or a repetition trial. Similarly, as in Experiment I, the performance was analysed using three measures: reaction times (RTs), percentage of errors (PE), and the number of correct trials. For the RT analyses, error trials (33.2%) and outliers were excluded. For identifying the latter, the

Switching-task performance in Experiment II

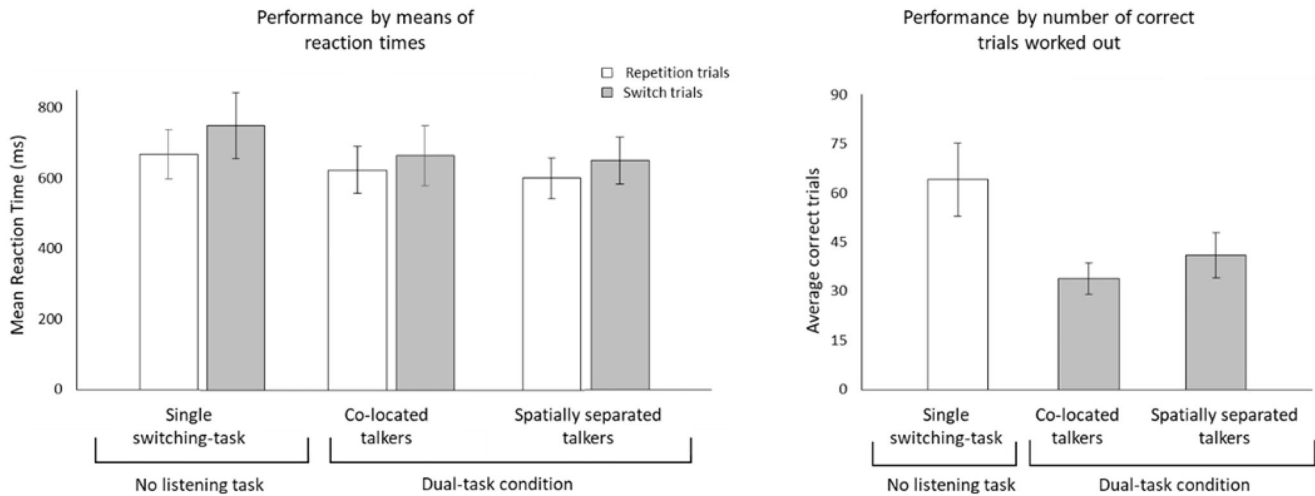


Figure 3. Switching-task performance in Experiment II. Note. Performance by means of reaction times (left panel) as a function of task conditions (single switching-task, co-located dual-tasking, spatially separated dual-tasking) and task switching (repetition trials, switch trials). Performance by number of correct trials worked out (right panel) as a function of task conditions (single switching-task, co-located dual-tasking, spatially separated dual-tasking). Error bars represents the 95% confidence interval of the mean.

median absolute deviation was calculated for each participant, and values exceeding ± 3 were excluded from the analyses.

Accordingly, we conducted two 3×2 ANOVAs – one for RTs and one for PE – on the independent within-subject factors *task conditions* ((a) co-located dual-tasking; (b) spatially separated dual-tasking; (c) single switching-task) and *task switching* (switch trial vs. repetition trial). The ANOVAs revealed a significant main effect of *task conditions* on RT, $F(2, 46) = 21.913$, $p < 0.001$, $\eta^2_p = 0.487$, but not on PE, $F < 1$. The main effect of *task switching* was also significant on RT, $F(1, 23) = 10.029$, $p < 0.01$, $\eta^2_p = 0.303$, but not on PE, $F(1, 23) = 1.332$, $p = 0.25$, $\eta^2_p = 0.054$. The interaction was significant neither on RT, $F(2, 46) = 1.343$, $p = 0.26$, $\eta^2_p = 0.55$, nor on PE, $F(2, 46) = 1.572$, $p = 0.21$, $\eta^2_p = 0.063$. To clarify the effect patterns on RTs, Tukey's tests were conducted as post-hoc comparisons. These verified that the RTs were longer when the switching task was performed alone than in parallel with the listening task (co-located dual-tasking: $M = 608$ ms, $SD = 90.2$ ms; spatially separated dual-tasking: $M = 601$ ms, $SD = 74.2$ ms; single switching task: $M = 710$ ms, $SD = 102.8$ ms). Additionally, the RTs were longer on switch trials than on repetition trials ($M = 664$ ms, $SD = 99.2$, vs. $M = 616$ ms, $SD = 78.9$ ms). These results are depicted in the left panel of Figure 3.

Similarly, as in Experiment I, we analysed the number of correct trials as a possible indicator of listening effort. We conducted an ANOVA on the average number of correct trials to test the performance difference between the three switching-task conditions (a)–(c). The analysis revealed a significant effect, $F(2, 46) = 51.72$, $p < 0.001$, $\eta^2_p = 0.692$ (see Figure 3, right panel). Post hoc comparison using Tukey's test indicated that more correct trials were worked out when the switching-task was performed alone (single-task) rather than in parallel with the listening task (dual-tasking). Importantly, when the switching-task was performed in parallel with the listening task, the number of correct trials was greater if talkers' auditory stimuli were spatially separated ($M = 41$ trials, $SD = 8.9$ trials) rather than co-located ($M = 34$ trials, $SD = 6.7$ trials). However, most trials were produced when the switching task was completed without the listening task ($M = 64$ trials, $SD = 14.6$ trials). The data of Experiment II are also available at <https://osf.io/dqfng>.

Discussion experiment II

In Experiment II, we implemented a more complex and possibly demanding secondary task, wherein participants had to switch between a number-judgement and a letter-judgement task while listening to a two-talker conversation. We found no memory performance difference between the two talkers' auditory stimuli conditions, that is, participants did not remember more facts listening to two-talker conversation when the talkers were spatially separated than co-located. In the switching (secondary) task, we found no effect of the two talkers' auditory stimuli on RTs, replicating the findings of Experiment I. One might argue that the lack of effect on RT suggests no variations of listening effort in the two talkers' auditory stimuli conditions. However, we assume that RTs are not sensitive enough to variations in listening effort in the quiet listening setting. In Experiment II, the switching task, that we used as a secondary task, seemed to be more difficult, since the error rates were generally larger than in Experiment I, and the switch costs, that is, RT difference when the number task switched to letter task or vice versa, was numerically smaller than in previous task-switching studies (see, for example, Rogers and Monsell 1995). The negligible switch costs in Experiment II could also indicate that RT data were insensitive to listening effort changes in our experimental manipulation of talkers' position in a quiet setting. We found the effects of the different talkers' auditory stimuli conditions on another secondary-task performance measure, namely, the number of correctly worked out trials. In Experiment II, the number of correct trials was greater when the two talkers in the listening task were spatially separated instead of co-located, replicating the findings of Experiment I. Thus, we argue for enhanced listening effort in the co-located condition.

Discussion

The present study aimed to examine the effect of spatial separation vs. co-location of two conversing talkers on a listener's memory and listening effort in a quiet setting. Thus, no noise or interferers was presented with the task-relevant speech signals. In two experiments, memory for running speech content was

measured directly by asking participants to answer content-related questions immediately after having heard a two-talker conversation. Additionally, the listening effort was measured indirectly through performance in a secondary task, which accompanied the primary listening task (dual-task paradigm). For this purpose, a relatively simple number-judgement task was used in Experiment I. In Experiment II, this number-judgement task was combined with a letter-judgement task, resulting in a switching task as a more demanding secondary task.

In both experiments, the patterns of effects, and thus results, were comparable. The memory performance for conversational contents was not different in the two talkers' auditory stimuli conditions (co-located vs. spatially separated talkers). Thus, the (non-)availability of binaural spatial cues did not affect the amount of information that listeners could recall. However, performance in the secondary task was higher when the conversing talkers were spatially separated, which we interpret as an indicator of reduced listening effort in that listening condition. That memory performance in the listening task did not vary with spatial cues while listening effort declined when spatial cues were available, which is in line with previous findings of Xia et al. (2015). These authors reported similar speech recognition rates when the target was separated from interferers either by spatial location or by gender cue. At the same time, the measured listening effort was reduced by the availability of spatial location cues. In most research on listening effort, like Xia et al. (2015), speech performance was tested in adverse listening conditions, specifically, in a listening setting in which the relevant speech signal was presented among other task-irrelevant talkers. However, the present study uniquely and experimentally verified the relevance of binaural spatial cues for listening effort even in quiet listening settings, that is, in the absence of broadband noise or task-irrelevant talkers. It thus extends the concept of listening effort to listening conditions, which are widely expected to be ideal for listening performance.

Notably, our results on the role of spatial cues for listening effort in quiet listening conditions are not based on differences between experimental conditions in reaction times (RTs) in the secondary task. Regarding this dependent variable, we found no differential effects of co-located and spatially separated talkers in both experiments. Instead, we found that the number of correct trials was significantly greater when spatial cues (due to talkers being binaurally auralised as being spatially separated) were available in the listening task. The same pattern of result emerged for both secondary tasks, the simpler number-judgement task in Experiment I as well as the more demanding switching task in Experiment II. The non-significant effect on RTs seems to contradict many studies that used the dual-task paradigm in adverse listening settings and found RTs to vary with listening conditions and thus concluded that RT is a sensitive and suitable objective indicator of listening effort (see Gagné, Besser, and Lemke 2017). Since in our two experiments, RTs were not sensitive to the experimental variations of the listening conditions, although we used "classic" RT tasks as a secondary task, one could argue that the availability of spatial cues does not play a significant role in listening effort when listening takes place in a quiet setting and/or when the listening task necessitates more complex cognitive performances. However, in our opinion, the effects of the two spatial listening conditions on the dependent variable number of correct trials, as demonstrated in both experiments, cannot be interpreted otherwise than in terms of listening effort differences (cp. also the interim discussions on Exp. I and Exp. II, respectively). These resulted in

different amounts of spare cognitive capacity available to the listener to perform the secondary task, although listening took place in quiet in both experimental conditions, and only the availability of spatial auditory cues for the talkers' positions varied. With this, the question arises about the extent to which listening effort indicators, proven to be sensitive in adverse listening situations (see Gagné, Besser, and Lemke 2017), are suitable in "less effortful" settings, as when listening takes place in a quiet place, where the target talker is not accompanied by background noise or interferers. For now, we suggest further testing the number of correct trials in a secondary task as a promising but yet to be verified indicator of variations in listening effort in less effortful listening situations when RT data are not a sensitive measure.

We consider our results to resonate well with the ELU model (Rönnberg et al. 2013). However, this model is primarily applied to explain greater listening effort in sub-optimal listening situations (e.g. noisy setting, hard of hearing). Our findings that spatial separation of talkers reduced listening effort suggest that auditory-perceptive information, such as the spatial position of talkers, is relevant in a quiet setting as well, releasing cognitive resource allocation while not burdening the explicit and more effortful processing route (cp. Introduction). Consequently, providing spatial cues drives the listener to follow the more automated implicit processing route, which allows the participants to enhance cognitive performance, such as memory for conversational contents. Thus, our findings provide new evidence that auditory-perceptive information (i.e. the spatial position of talkers) plays a role in higher-level auditory cognition, like remembering what has been said, and might stimulate theory development to enhance our understanding of the mechanisms underlying listening focussing on quiet listening situations as well.

Certainly, listening to a set of sentences and recalling heard information not only relies on memory processes and functions but also involves perceptual functions, attention, verbal-logical reasoning as well as executive functions like updating, inhibiting and/or switching. This is also due to the fact that cognitive processing of coherent text requires the ongoing integration of new incoming information into already processed, subsequently maintained and potentially to-be-updated information so that the listener develops an adequate mental model of the heard story. In the present experiments, the dependent variable in the listening task was correct reproduction of heard conversational contents, which was presented explicitly or could be inferred from the overheard content (e.g. "He has two brothers and no sister" means, that there are three children in this family altogether). This primary task can therefore be described as a memory task, even if the performance of participants is not exclusively or isolated due to memory processes and functions.

Presumably, quiet listening settings are a standard listening situation, as well as the standard cognitive task when listening to speech is to process the heard information cognitively. After so many years of hearing research and many studies focussing on the fundamental mechanisms of hearing and listening using a simple cognitive task with basic stimuli in adverse listening settings, it might be time to move on and broaden our knowledge with a more standard listening situation with more real-life like cognitive tasks.

The findings of hearing research have contributed to the underlying mechanisms of hearing and listening using predominantly simple cognitive tasks in adverse listening situations. Now it may be the time for listening research to focus more on "ideal" listening conditions, such as unimpaired listening in quiet and more complex cognitive listening performances, such as short-term memory of running speech.

Acknowledgements

The authors thank Ann-Sophie Schenk, Charlotte Dickel, Jian Pan, Wiebke Stöver and Wisam Nassar for data collection and for preparing the datasets.

Declaration of interest

No potential conflict of interest was reported by the author(s).

Funding

This research was financially supported by the Head-Genuit-Stiftung (<https://head-genuit-stiftung.de/>).

ORCID

Edina Fintor  <http://orcid.org/0000-0003-0280-251X>

References

- Aspöck, L., R. A. Viveros Munoz, and J. Fels. 2020. "Additional Data for Experiments on Spatial Release from Masking in Different Room Acoustic Conditions." *RWTH Aachen University*. <https://publications.rwth-aachen.de/record/789293/files/Documentation%20of%20additional%20data%20for%20experiments%20on%20spatial%20release%20from%20masking%20in%20different%20room%20acoustic%20conditions.pdf>
- Boersma, P., and D. Weenink. 2020. "Praat: Doing Phonetics by Computer." Computer program Version 6.1.12. *Praat*. <http://www.praat.org/>. Accessed 15 April 2020.
- Bregman, A. S. 1990. *Auditory Scene Analysis*. Cambridge, MA: MIT Press.
- Bronkhorst, A. 2000. "The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions." *Acta Acustica United with Acustica* 86 (1): 117–128.
- Carlyon, R. P., R. Cusack, J. M. Foxton, and I. H. Robertson. 2001. "Effects of Attention and Unilateral Neglect on Auditory Stream Segregation." *Journal of Experimental Psychology. Human Perception and Performance* 27 (1): 115–127. doi:10.1037//0096-1523.27.1.115.
- Cherry, E. C. 1953. "Some Experiments on the Recognition of Speech, with One and with Two Ears." *The Journal of the Acoustical Society of America* 25 (5): 975–979. doi:10.1121/1.1907229.
- Conway, A. R. A., N. Cowan, and M. F. Bunting. 2001. "The Cocktail Party Phenomenon Revisited: The Importance of Working Memory Capacity." *Psychonomic Bulletin & Review* 8 (2): 331–335. doi:10.3758/bf03196169.
- Edwards, B. 2016. "A Model of Auditory-Cognitive Processing and Relevance to Clinical Applicability." *Ear & Hearing* 37 (1): 85S–91S. doi:10.1097/AUD.0000000000000308.
- Gagné, J.-P., J. Besser, and U. Lemke. 2017. "Behavioral Assessment of Listening Effort Using a Dual Task Paradigm: A Review." *Trends in Hearing* 21: 233121651668728. doi:10.1177/2331216516687287.
- Gallun, F. G., A. C. Diedesch, S. D. Kampel, and K. M. Jakien. 2013. "Independent Impacts of Age and Hearing Loss on Spatial Release in a Complex Auditory Environment." *Frontiers in Neuroscience* 7: 252. doi:10.3389/fnins.2013.00252.
- Haftner, E. R., J. Xia, S. Kalluri, R. Poggesi, C. Hansen, and K. Whiteford. 2013. "Attentional Switching When Listeners Respond to Semantic Meaning Expressed by Multiple Talkers." *The Journal of the Acoustical Society of America* 133 (5): 3381. doi:10.1121/1.4805827.
- Hasher, L., and R. T. Zacks. 1979. "Automatic and Effortful Processes in Memory." *Journal of Experimental Psychology* 108 (3): 356–388. doi:10.1037/0096-3445.108.3.356.
- Humes, L. E., J. H. Lee, and M. P. Coughlin. 2006. "Auditory Measures of Selective and Divided Attention in Young and Older Adults Using Single Talker Competition." *The Journal of the Acoustical Society of America* 120 (5): 2926–2937. doi:10.1121/1.2354070.
- Ihfeldt, A., and B. Shinn-Cunningham. 2008. "Disentangling the Effects of Spatial Cues on Selection and Formation of Auditory Objects." *The Journal of the Acoustical Society of America* 124 (4): 2224–2235. doi:10.1121/1.2973185.
- Imhof, M. 2010. "What is Going on in the Mind of a Listener? The Cognitive Psychology of Listening." In *Listening and Human Communication in the 21st Century*, edited by A. D. Wolvin, 97–126. Malden, MA: Wiley-Blackwell.
- Kahneman, D. 1973. *Attention and Effort*. Upper Saddle River, NJ: Prentice-Hall Inc.
- Koch, I., E. Poljac, H. Müller, and A. Kiesel. 2018. "Cognitive Structure, Flexibility, and Plasticity in Human multitasking—An integrative review of dual-task and task-switching research." *Psychological Bulletin* 144 (6): 557–583. doi:10.1037/bul0000144.
- Leys, C., C. Ley, O. Klein, P. Bernard, and L. Licata. 2013. "Detecting Outliers: Do Not Use Standard Deviation around the Mean, Use Absolute Deviation around the Median." *Journal of Experimental Social Psychology* 49 (4): 764–766. doi:10.1016/j.jesp.2013.03.013.
- Masiero, B., and J. Fels. 2011. "Perceptually Robust Headphone Equalization for Binaural Reproduction." Paper presented at the 130th Audio Engineering Society Convention, London, UK, May 13–16.
- McGarrigle, R., K. J. Munro, P. Dawes, A. J. Stewart, D. R. Moore, J. G. Barry, S. Amitay, et al. 2014. "Listening Effort and Fatigue: What Exactly Are we Measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper.'" *International Journal of Audiology* 53 (7): 433–440. doi:10.3109/14992027.2014.890296.
- Oberem, J., V. Lawo, I. Koch, and J. Fels. 2014. "Intentional Switching in Auditory Selective Attention: Exploring Different Binaural Reproduction Methods in an Anechoic Chamber." *Acta Acustica United with Acustica* 100 (6): 1139–1148. doi:10.3813/AAA.918793.
- Picou, E. M., and T. A. Ricketts. 2014. "The Effect of Changing the Secondary Task in Dual-Task Paradigms for Measuring Listening Effort." *Ear and Hearing* 35 (6): 611–622. doi:10.1097/AUD.0000000000000055.
- Rogers, R. D., and S. Monsell. 1995. "Costs of a Predictable Switch between Simple Cognitive Tasks." *Journal of Experimental Psychology: General* 124 (2): 207–231. doi:10.1037/0096-3445.124.2.207.
- Rönnerberg, J., T. Lunner, A. Zekveld, P. Sörqvist, H. Danielsson, B. Lyxell, and Ö. Dahlström. 2013. "The Ease of Language Understanding (ELU) Model: Theoretical, Empirical, and Clinical Advances." *Frontiers in System Neuroscience* 7: 31. doi:10.3389/fnins.2013.00031.
- Rubinstein, J. S., D. E. Meyer, and J. E. Evans. 2001. "Executive Control of Cognitive Processes in Task Switching." *Journal of Experimental Psychology. Human Perception and Performance* 27 (4): 763–797. doi:10.1037/0096-1523.27.4.763.
- Sarampalis, A., S. Kalluri, B. Edwards, and E. Haftner. 2009. "Objective Measures of Listening Effort: Effects of Background Noise and Noise Reduction." *Journal of Speech, Language, and Hearing Research* 52 (5): 1230–1240. doi:10.1044/1092-4388(2009/08-0111).
- Schmitz, A. 1995. "Ein Neues Digitales Kopfhörersystem." *Acta Acustica United with Acustica* 81 (4): 416–420.
- Seeman, S., and R. Sims. 2015. "Comparison of Psychophysiological and Dual-Task Measures of Listening Effort." *Journal of Speech, Language, and Hearing Research* 58 (6): 1781–1792. doi:10.1044/2015_JSLHR-H-14-0180.
- Surprenant, A. M. 1999. "The Effect of Noise on Memory for Spoken Syllables." *International Journal of Psychology* 34 (5–6): 328–333. doi:10.1080/002075999399648.
- Wingfield, A. 2016. "Evolution of Working Memory and Cognitive Resources." *Ear & Hearing* 37 (1): 35S–43S. doi:10.1097/AUD.0000000000000310.
- Xia, J., S. Kalluri, B. Edwards, and E. R. Haftner. 2014. "Cognitive Effort and Listening in Everyday Life." *ENT & Audiology News* 23: 88–89.
- Xia, J., N. Nooraei, S. Kalluri, and B. Edwards. 2015. "Spatial Release of Cognitive Load Measured in a Dual-Task Paradigm in Normal-Hearing and Hearing-Impaired Listeners." *The Journal of the Acoustical Society of America* 137 (4): 1888–1898. doi:10.1121/1.4916599.
- Zekveld, A. A., M. Rudner, S. E. Kramer, J. Lyzenga, and J. Rönnerberg. 2014. "Cognitive Processing Load during Listening is Reduced More by Decreasing Voice Similarity than by Increasing Spatial Separation between Target and Masker Speech." *Frontiers in Neuroscience* 8: 88. doi:10.3389/fnins.2014.00088.